Title

Conformational Polymorphism Gives Rise to Specialized Function in Bacterial Microcompartment Shell Proteins

Author

Thompson, Michael Christopher

Publication Date

2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Conformational Polymorphism Gives Rise to Specialized Function in Bacterial Microcompartment Shell Proteins

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy in Biochemistry and Molecular Biology

by

Michael Christopher Thompson

2014

ABSTRACT OF THE DISSERTATION

Conformational Polymorphism Gives Rise to Specialized Function in Bacterial Microcompartment Shell Proteins

by

Michael Christopher Thompson

Doctor of Philosophy in Biochemistry and Molecular Biology

University of California, Los Angeles, 2014

Professor Todd O. Yeates, Chair

It is well-established that proteins are dynamic molecules, endlessly interconverting back and forth between different conformational substates. A primary goal of structural biology is to understand the functional relevance of this conformational polymorphism. The work described here demonstrates the connection between conformational polymorphism and specialized protein functions in the BMC-domain superfamily of symmetric protein oligomers that form the semi-permeable, polyhedral shells of bacterial microcompartment organelles. Three specific structural and biochemical studies of BMC-domain proteins are presented as examples. The first of these studies focuses on a tandem BMC-domain protein, EutL, in which broken oligomeric symmetry allows for a conformational rearrangement that relates to molecular transport functions. The results presented support a model of allosteric regulation of EutL function. The second study presented highlights a case of a difficult structure determination of a BMC-domain protein, CcmK1 (L11K), in which the arrangement of molecules in the crystal lattice suggests the protein oligomers have lower internal symmetry than previously believed, resulting from conformational polymorphism. The third study aims to characterize a BMC-domain shell protein, GrpU, in which an

unexpected occurrence of broken symmetry allows the formation of a unique iron-sulfur cluster binding site. These three examples collectively illuminate the role of conformational polymorphism in the BMC-domain family, and provide interesting insight into the complex transport properties of the bacterial microcompartment shell, which remains relatively poorly understood. Finally, a crystallographic analysis is presented that extends the present study beyond the BMC-domain proteins by exploring a novel methodology for detecting and characterizing conformational polymorphism, which can be applied broadly in protein crystallography.

This dissertation of Michael Christopher Thompson is approved.

David S. Eisenberg

Robert P. Gunsalus

Todd O. Yeates, Committee Chair

University of California, Los Angeles

2014

To the teachers and mentors who inspired me to go on this adventure…

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1. Introduction

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I want to thank my family. I wouldn't have done any of this without you. I thank my parents, Susan and Quentin Thompson, for giving me a lasting love for learning and motivating me to succeed from an early age. Mom and Dad, thank you for your never-ending support, and for always letting me choose my own path through my education. I also want to thank the rest of my family, especially my grandparents, Toni and Larry Beaver, for also supporting me and for being understanding of the fact that I didn't get to spend much time with them while I was in school, even though I was close geographically. The same goes for my friends, who were often neglected while this work was getting done. I especially want to thank Laura Schelhas, who was my best discovery while at UCLA. Laura, you kept me sane and well-fed through most of graduate school, and things would have been a lot harder without you. I love you. Finally, I would like to thank Otto, Chloe, and Daisy for much needed distractions from hard work.

Next, I want to express my deepest thanks to Professor Todd Yeates, who was the most important person who contributed to my success at UCLA. I am fairly sure that the work presented in this dissertation is not what was planned for me when I joined the Yeates Lab. Todd, I am very grateful for your patience and your willingness to let your students explore their own ideas and interests. Even though you let me drive projects in my own directions, you still taught me so much that I never thought I would be able to grasp. Among other things, my understanding of mathematics and my ability to use computers have both been dramatically improved, and I am lucky to have experienced your insight as a mentor.

I want to thank Dr. Duilio Cascio and Dr. Michael Sawaya, who are an endless source of information and inspiration when it comes to crystallography. Duilio and Mike, you guys were such I huge part of my training at UCLA, that I can't thank you enough. I can't think of two people who are so different, but at the same time so similar. I have learned tons from each of you, and even though you have very different styles and ways of approaching problems, you are both so generous with your time and

knowledge and you make a great team. Also thanks for making synchrotron trips an unforgettable experience.

I'd like to thank all the past and present members of the Yeates Lab for making a pleasant working environment. I would especially like to thank the following people, who have had a significant impact on the outcome of my work: Julien Jorda, Chris Crowley, Nicole Wheatley, Dan McNamara, Yen-Ting Lai, David Leibly, Sunny Chun, Inna Pashkov, Danny Gidaniyan, Neil King, Jason Forse, Yingssu Tsai, Shiho Tanaka, and Jeff Kopstein. Also, I had the opportunity to work with the following undergraduates, who helped get a lot of work done: Hoda Ahmed, Krystal McCarty, Vitaliy Davydov, and Joseph Alzagatiti.

While at UCLA, I had the opportunity to work with a number of excellent scientists and staff members, all of whom helped me in some significant way. Dr. Zhongyu Yang, Dr. Robert Peterson, Dr. Martin Phillips, Dr. Julian Whitelegge, Harold Aschmann, Matthew Graf, Michael Collazo, and Jason Navarro all helped with experiments and/or instrumentation. I benefitted greatly from having your expertise. Dr. Luki Goldschmidt, you are a wizard with computers, a good friend, and one of the most generous people I know. Your help (and patience) saved me countless hours and I learned a lot from you. Also, I'd like to thank Cindy Chau for all of her administrative assistance, which saved me a lot of time that I got to spend doing science. Finally, I'd like to thank the staff of the NE-CAT beamlines at APS, who provided vast technical expertise and world-class facilities for our diffraction experiments, not to mention enjoyable visits to Argonne.

During my time in the Yeates Lab, we were lucky to be visited by Professor Tom Bobik from Iowa State University. It was a pleasure to work with Tom, and I am grateful for his insight, from which my work directly benefited.

I also wish to thank the faculty members who graciously served on various committees in order to help me fulfill my requirements. This includes Prof. David Eisenberg, Prof. Rob Gunsalus, Prof. Ronald Kaback, and Prof. Jorge Torres. I would like to give special thanks to David Eisenberg, with whom I interacted during our weekly meetings and while teaching. David, you are an excellent mentor and

example, and I am lucky to have had the chance to interact closely with you. Additionally, I want to thank Prof. Jim Bowie and Prof. Rob Clubb, both of whom allowed me to rotate in their labs during my first year.

For three of my years at UCLA, I was fortunate to receive funding from the UCLA-NIH Cellular and Molecular Biology Training Program (Ruth L. Kirschstein National Research Service Award GM007185). I am also grateful to Prof. Steve Clarke and Christine Briganti, who were responsible for the program while I participated.

Last but not least, I would like to thank my lab mates in the Alber Lab at UC Berkeley, an excellent group of people without whom I would most likely not have made it to UCLA. I am forever grateful to Prof. Tom Alber, who tragically passed away just a few months before this dissertation was completed. Tom, you taught me so much and helped get me to where I am today, and I continually look to my experiences with you for inspiration – you will be missed. In the Alber Lab I had the opportunity to work with a great group of people who mentored me at an early stage, and taught me a lot of the important skills that helped me succeed at UCLA. I would especially like to acknowledge Laurie Gay, Nat Echols, Jaime Fraser, Christina Baer, Christine Gee, Terry Lang, Ho Leung Ng, Christoph Grundner, and Andy Greenstein.

Previous Publications and Contributions of Co-Authors:

Chapter 2 contains results which will be submitted for publication shortly after the completion of this dissertation. Dr. Duilio Cascio helped with crystallographic structure determination. Dr. Christopher Crowley and Jeff Kopstein helped with protein purification and crystallization. Dr. Julien Jorda performed bioinformatics analysis. Dr. Robert Peterson performed NMR experiments, and Dr. Julian Whitelegge assisted with mass spectrometry.

Chapter 3 is a version of: Thompson MC, Yeates TO. "A Challenging Interpretation of a Hexagonally Layered Protein Structure." Acta Cryst. D70 (2014): 203-208.

Chapter 4 is a version of : Thompson MC, Wheatley NM, Jorda J, Sawaya MR, Gidaniyan SD, Ahmed H, Yang Z, McCarty KN, Yeates TO. "Identification of a Unique Fe-S Cluster Binding Site in a

Glycyl-Radical Type Microcompartment Shell Protein." This manuscript was submitted for publication and is under review during the completion of this dissertation.

Chapter 5 contains results results which will be submitted for publication shortly after the completion of this dissertation. Prof. Todd Yeates developed the theory upon which the described methodology is based.

Professor Todd Yeates directed all of the research presented in this dissertation.

VITA

2007	B.A. in Molecular and Cell Biology, UC Berkeley
2007-2008	Staff Research Associate, Alber Lab, UC Berkeley
2009-2010	Teaching Assistant, Department of Chemistry and Biochemistry, UCLA
2009-2012	UCLA-NIH Cellular and Molecular Biology Training Fellowship
2011	Second Place Poster Award: West Coast Protein Crystallography Workshop XX
2012	Graduate Student Poster Award: 26th Annual Symposium of the Protein Society
2012	Samson H. Cheng Biochemistry Teaching Award, Department of Chemistry and Biochemistry, UCLA
2013	American Crystallographic Association Travel Fellowship, ACA Annual Meeting
2013	First Place (Graduate and Postdoc) Poster Award, UCLA Seaborg Symposium

PUBLICATIONS AND SELECTED PRESENTATIONS

Thompson MC, Chun S, Bobik TA, Yeates TO. 58th Annual Meeting of the Biophysical Society. San Francisco, CA. Poster: "An Allosteric Model for Pore Opening by Substrate Binding in the EutL Microcompartment Shell Protein." (February 2014)

Thompson MC, Yeates TO. "A Challenging Interpretation of a Hexagonally Layered Protein Structure." Acta Cryst. D70 (2014): 203-208. PMID: 24419393

Thompson MC, Jorda J, Ahmed H, McCarty K, Yeates TO. Annual Meeting of the American Crystallographic Association. Honolulu, HI. Oral Presentation: "Structural Characterization of a Divergent Microcompartment Shell Protein Suggests a Novel Type of Cofactor Interaction." (July 2013)

Thompson MC, Sawaya MR, Yeates TO. West Coast Protein Crystallography Workshop XXI. Monterey, CA. Poster: "Analysis of Varied Crystal Pathologies in Hexagonally Packed Microcompartment Shell Proteins." (March 2013)

Thompson MC, Crowley CS, Cascio D, Jorda J, Peterson RD, Whitelegge JP, Kopstein JS Yeates TO. 26th Annual Symposium of the Protein Society. SanDiego, CA. Poster: "The Protein Shell of the Ethanolamine Utilization Microcompartment is a Redox-Sensitive Barrier." (August 2012)

Yeates TO, **Thompson MC**, Bobik TA. "The protein shells of bacterial microcompartment organelles." Curr. Opin. Struct. Bio. 21 (2011): 223-231. PMID: 21315581

Thompson MC, Crowley CS, Cascio DS, McNamara DE, Kopstein JS, Yeates TO.West Coast Protein Crystallography Workshop XX. Monterey, CA. Poster: "Crystallographic Studies of CPE0900: Insights into Assembly and Function of the Bacterial Microcompartment Shell." (March 2011)

Thompson MC, Forse GJ, Ram N, Guo F, Yeates TO. Gordon Research Conference – Diffraction Methods in Structural Biology. Lewiston, ME. Poster: "Synthetic Symmetrization as a Strategy to Crystallize RNA and RNA:Protein Complexes." (July 2010)

CHAPTER 1

Introduction

1.1 Bacterial Microcompartments (MCPs): Specialized Internal Structures in Prokaryotes

A common misconception in the study of biology is that prokaryotic cells are simplistic in terms of their cellular architecture, generally lacking any sophisticated organization of intracellular components. Indeed, most textbooks contain images depicting the prokaryotic cytoplasm as a homogenous mixture of biomolecules, often displayed in opposition with elaborate images of eukaryotic cells that show complex arrangements of membrane-bound organelles and other highly-ordered structures. Historically, this naïve, textbook view of prokaryotic cells is the likely result of their relatively small size when compared to their eukaryotic counterparts, making them significantly more difficult to view using microscopy techniques. In contrast, dramatic technological advancements in imaging technology over the last several decades have led to a revolution in our understanding of prokaryotic physiology[2]. To date, a significant number of complex, subcellular structures with dedicated functions have been discovered in prokaryotic cells, leading to a new paradigm in prokaryotic cell biology[3].

Among the numerous internal structures identified in bacteria are giant protein complexes that facilitate specialized metabolic functions, collectively known as bacterial microcompartments (MCPs). In 1973, Shivley, et al. were the first to characterize the MCPs of the chemoautotroph *Thiobacilis neapolitanus*, demonstrating that the observed "polyhedral inclusions" serve a specific metabolic function[4,5]. Composed entirely of protein subunits, these massive MCP complexes self-assemble from thousands of individual polypeptide chains, which are typically expressed from a single operon containing roughly 10-20 open reading frames[6–9] (Figure 1.1a,b). A specific subset of these polypeptides forms a thin polyhedral shell, approximately 100-150nm in diameter, which encapsulates several sequentially-acting metabolic enzymes and sequesters them from the cytosol (Figure 1.1c). Consequently, the MCP satisfies the functional definition of an organelle: it creates a microenvironment that is distinct from the cytosol, it

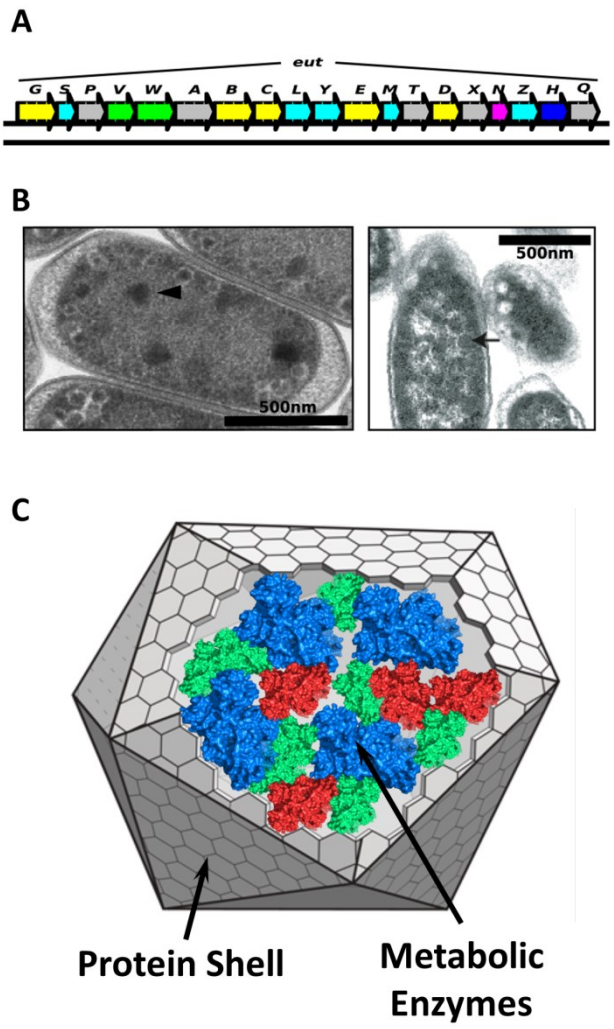


Figure 1.1. Overview of bacterial microcompartments. A) An example of a typical operon encoding an MCP, in this case for ethanolamine catabolism. Genes encoding enzymes are depicted in yellow, BMC-domain shell protein genes are colored cyan, a non-BMC-domain shell protein gene is colored magenta, genes corresponding to encapsulated enzymes are colored yellow, genes for cytosolic components are colored gray, genes for regulatory components are colored green, and the gene for a transmembrane ethanolamine transporter is colored blue. B) Transmission electron micrograph of cells expressing MCPs. The MCP structures appear as dark inclusion bodies (Image courtesy of Jorge Escalante-Semerena). C) A schematic depicting the general structure of an MCP, in which a proteinaceous shell surrounds a series of sequentially-acting metabolic enzymes.

enhances metabolic flux by colocalization of enzymes and substrates, and it is capable of retaining metabolic intermediates that have the potential to cause harm to the cell when exposed to the cytoplasm. Since their initial discovery, at least 7 homologous MCP systems have been identified, each associated with a distinct metabolic pathway, and approximately 17% of sequenced bacterial genomes have been found to contain signatures of MCP operons[10].

1.2 Current Experimental Characterization of MCPs has focused on Three Important Systems

Currently, the majority of our knowledge about MCPs has been derived from experimental characterization and computational analysis of three model systems[8,9] (Figure 1.2).

The first of these MCP systems is the carboxysome, found in cyanobacteria and some chemoautotrophs, which encapsulates the enzymes carbonic anhydrase and ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) that are responsible for the initial steps of carbon fixation as part of the Calvin-Benson-Bassham cycle[6,11,12]. Oceanic cyanobacteria account for over 50% of global carbon fixation[13], making the carboxysome highly relevant to our understanding of ecology. Within the carboxysome, carbonic anhydrase dehydrates bicarbonate and releases carbon dioxide (CO_2) in the vicinity of RuBisCO, an enzyme with notoriously poor catalytic efficiency and substrate specificity[14]. RuBisCO then catalyzes the formation of 3-phosphoglycerate, using CO_2 and ribulose-1,5-bisphosphate as substrates. Encapsulation of these metabolic reactions within the MCP serves two important functions. First, it retains the CO_2 intermediate, which is gaseous at physiological temperature and pressure, making it prone to diffusion out of the cell. Second, colocalization of carbonic anhydrase and RuBisCO serves to enhance the local CO_2 concentration in the presence of RuBisCO, thereby enhancing productive metabolic flux through this relatively inefficient enzyme.

The second type of MCP system is involved in the catabolism of propanediol in a number of saccharolytic bacteria, including both human and plant pathogens. These saccharolytic pathogens degrade cell-surface carbohydrates from host tissues, releasing 1,2-propanediol. Subsequently, 1,2-propanediol is converted to propionate, which enters the methylcitrate cycle, serving as a source of metabolic energy.

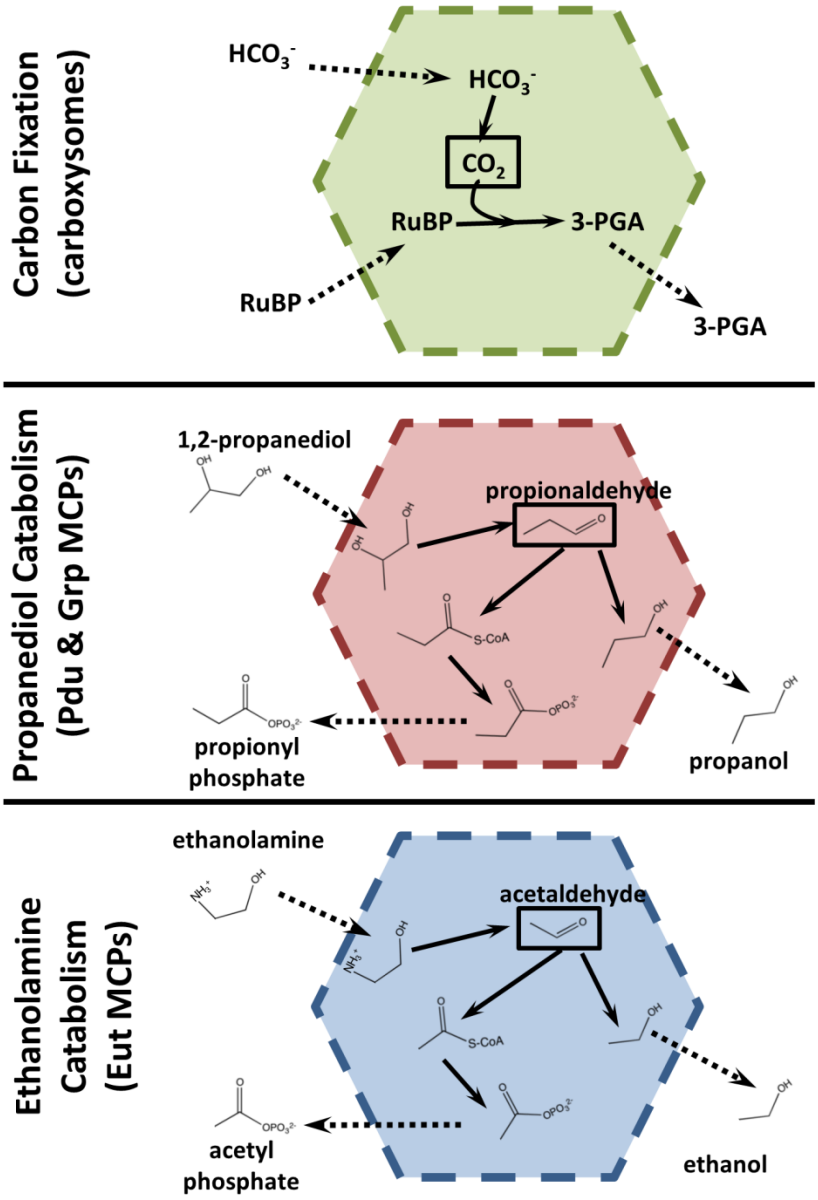


Figure 1.2. Relevant examples of MCPs. The three panels each depict a different biochemical pathway that is encapsulated within an MCP. In each panel, the reactive and/or toxic intermediate compound that the MCP shell must retain is surrounded with a box. A) The carboxysome is an MCP involved in carbon fixation. B) Two types of MCPs, Pdu and Grp, are involved in propanediol catabolism. C) The Eut MCP encapsulates a pathway for ethanolamine catabolism.

The initial steps of this catabolic process involve the intermediate propionaldehyde, which is toxic to cells and also mutagenic due to its reactivity with biological macromolecules. In order to retain the propionaldehyde produced by this process, the enzymes that produce and consume this intermediate are confined together within an MCP[15,16]. During the first reaction that occurs within these MCPs, 1,2-propanediol is dehydrated to propionaldehyde by a diol dehydratase enzyme. Subsequently, two distinct dehydrogenase enzymes catalyze the formation of either propionyl-CoA or propanol, using the aldehyde as a substrate. Finally, the propionyl-CoA is converted to propionyl-phosphate by an acyltransferase enzyme. Two distinct types of propanediol catabolic MCP have been identified to date, which differ slightly with respect to the first enzymatic step[10]. The Pdu type of MCP (short for Propanediol Utilization), contains a B_{12} coenzyme-dependent diol dehydratase, while the Grp type of MCP (short for Glycyl-Radical Propanediol) contains a diol dehydratase that functions by a glycyl-radical mechanism.

The third type of MCP system is one that is responsible for the catabolism of ethanolamine, and therefore has been given the name Eut (short for Ethanolamine Utilization)[17,18]. The Eut MCP system is commonly found in enteropathogenic bacteria that inhabit the human gut and are associated with human disease[19,20]. These bacteria include pathogenic Enterobacteriaceae, such as *Salmonella* and *Escherichia coli*, as well as some phylogenetically-distant pathogens including species from the genus *Clostridium*. The selective pressure for encapsulation of ethanolamine catabolism within an MCP results from the formation of acetaldehyde as an intermediate. As with the propionaldehyde intermediate retained by propanediol catabolic MCPs, acetaldehyde is chemically reactive and mutagenic, and additionally, unlike propionaldehyde, acetaldehyde has an extremely high vapor pressure at physiological temperature and pressure. The high vapor pressure of acetaldehyde, combined with its small size and hydrophobicity, allow it to evaporate from the cell, leading to loss of nutrients[21]. The enzymatic reactions that take place inside the Eut MCP are highly similar to the reactions that occur within the propanediol catabolic MCPs. First, ethanolamine is degraded to ammonia and acetaldehyde by the ethanolamine-ammonia lyase enzyme. Next, acetaldehyde molecules encounter one of two distinct dehydrogenase enzymes, and are either reduced to form ethanol, or converted to acetyl-CoA by oxidative coupling. A final enzymatic

reaction (acyltransferase) of acetyl-CoA with orthophosphate creates acetyl-phosphate and releases coenzyme-A. The products of the reactions that occur within the Eut MCP can subsequently be converted to other metabolites that are used in the tricarboxylic acid cycle or various biosynthetic pathways.

1.3 The MCP Shell: A Complex Function Emerges from Assembly of BMC-domain Proteins

Like the plasma membranes of their eukaryotic counterparts, MCP shells must operate as selectively-permeable diffusion barriers between the cytoplasm and the lumen in order for MCP structures to act as functional organelles. MCP shells must restrict the efflux of small, hydrophobic metabolic intermediates, while simultaneously allowing the passage of more polar, but similarly-sized, substrates and products[7,15,21,22]. Additionally, in some MCP types, encapsulated enzymes require access to cofactor molecules that are significantly larger than the molecular species that need to be retained within the lumen. The Eut MCP serves as an excellent example of the functional complexity required of the MCP shell; it must permit the diffusion of ethanolamine, ethanol, and acetyl-phosphate, while at the same time preventing the escape of acetaldehyde, a molecule with similar size and chemical properties. Furthermore, the encapsulated ethanolamine-ammonia lyase enzyme sometimes requires the exchange of bulky cofactors with the cytoplasm following mechanism-based inactivation, leading to a requirement for transport of molecular species that are much larger than the small aldehyde that must be contained within the MCP.

The ability of the MCP shell to function as an intricate diffusion barrier is a direct result of its structure and composition[23,24]. The primary building blocks of MCP shells are known as BMC-domain proteins (BMC for Bacterial Microcompartment)[25]. The BMC-domain is a small, approximately 100 amino acids, protein domain that adopts a compact fold containing two β-α-β motifs connected by a β-hairpin loop (a ferredoxin fold). Although relatively simple, the BMC-domain has a fantastic capacity for assembly into higher-order structures (Figure 1.3). BMC-domains first oligomerize into hexamers with cyclic 6-fold symmetry, and roughly hexagonal shape. These hexameric tiles further assemble, in an edgewise fashion, to form tightly packed, two-dimensional sheets. Integration of pentameric vertex

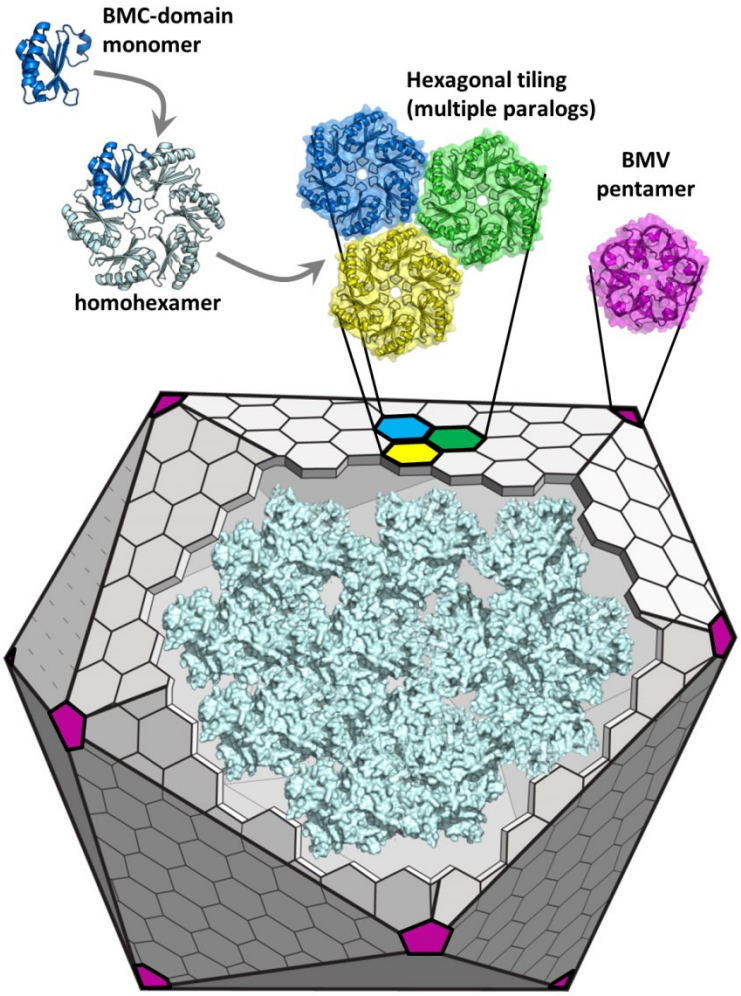


Figure 1.3. Assembly and structure of the MCP shell. The major components of the MCP shell are the BMC-domain proteins, which oligomerize into hexameric tiles that further assemble to form the flat facets of a polyhedral shell. Multiple BMC-domain paralogs combine with pentameric vertex proteins to form an essentially solid shell surrounding the interior enzymes. Small pores perforate the centers of some BMC-domain hexamers.

proteins (BMV proteins) allows these molecular sheets to form a closed polyhedral shell, which is perforated only by small pores found at the centers of some shell protein hexamers[22,26]. The specific chemical properties of these pores (i.e. shape, size, electrostatic potential, etc.) govern their permeability to various small molecules.

An additional level of complexity with respect to the structure and function of the MCP shell exists, based on the fact that multiple BMC protein paralogs are simultaneously present within a single MCP shell[16,17,27,28] (Figure 1.3). Based on X-ray crystal structures of paralogous BMC proteins[25,27,29], and phenotypic analysis of MCPs with varied BMC protein mutations[30–32], it has become clear that each paralog serves a unique role within the context of the shell. Among the key differences between homologous BMC-domain hexamers are differences in the physical and chemical characteristics of their central pores, which are likely to endow them with distinct transport functions. The putative transport functions of various types of BMC protein paralogs have been reviewed extensively[8,9,23,24]. Examples include the CcmK sub-family of proteins from the β-carboxysome, whose small, positively-charged pores are assumed to transport bicarbonate molecules[25,33], the EutL protein from the Eut MCP, whose central pore is capable of transitioning between two conformational substates[27,34], and the PduT protein from the Pdu MCP, which binds an iron-sulfur cluster at its pore[29,31,35] (Figure 1.4). While BMC-domain homologs display a wide array of central pores which are likely to have highly varied functions, they all appear to facilitate transport processes that are already thermodynamically favored. In other words, BMC-domain shell proteins do not perform active transport, but instead serve as a selectivity filter for diffusion.

1.4 Symmetry Influences Structure-Function Relationships in BMC-domain Shell Proteins

The requirement for assembly into a two-dimensional lattice places an obvious constraint on the structure of BMC-domain proteins and oligomers. These proteins assemble into highly-ordered, semi-crystalline structures, with well-defined lattice spacings, wherein the individual tiles must be very tightly packed with virtually no space between[33,36]. This assembly requirement puts very strict constraints on the size and shape of the BMC-domain hexamers, because their perimeters must be compatible with the

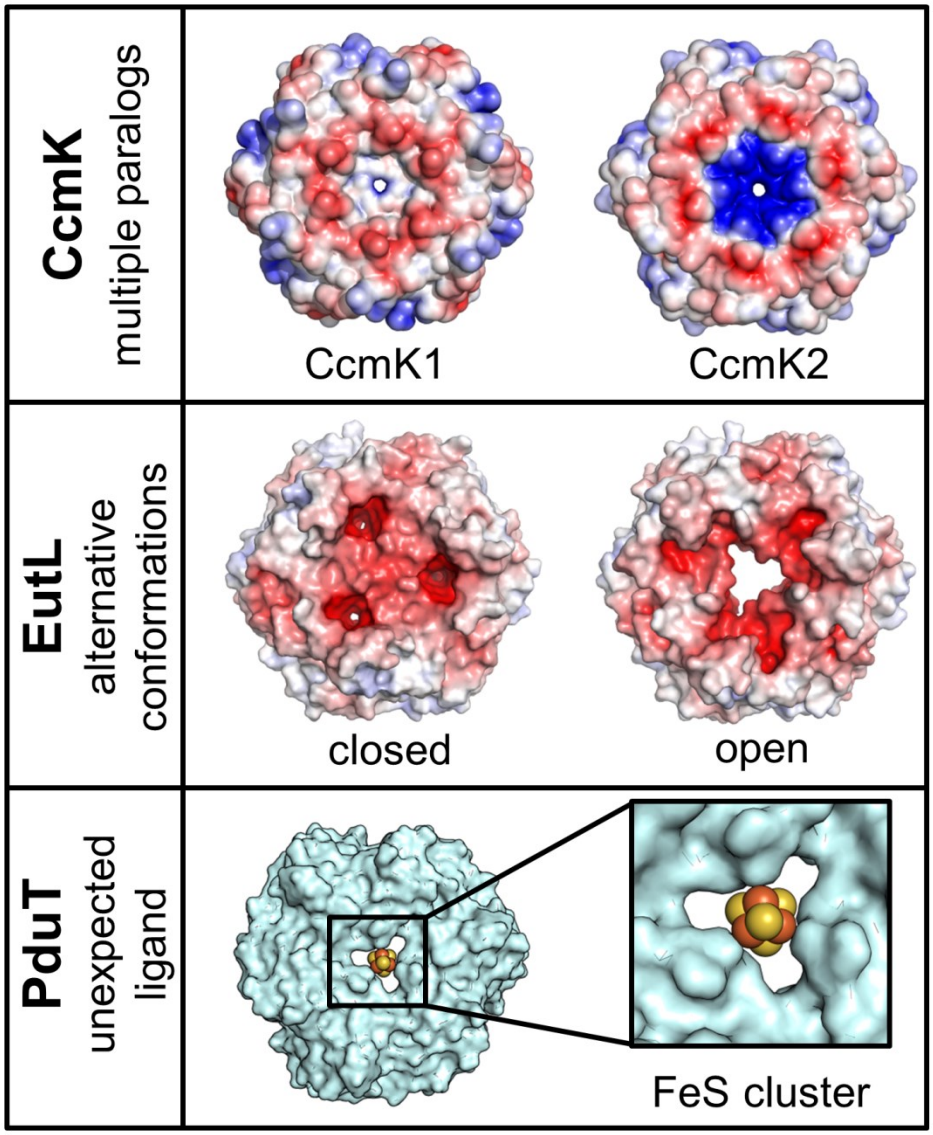


Figure 1.4. BMC-domain hexamers have diverse functions. BMC-domain proteins have diverse functions that combine to give the shell complex molecular transport properties. Variations on BMC domain hexamers (and pseudohexamers) include pores with different electrostatic properties (CcmK paralogs, top panel), large, gated pores (EutL, middle panel), and iron-sulfur clusters (PduT, bottom panel).

lattice-packing interactions that support layer formation. These requisite interactions were recently highlighted by Sinha, et al., who pointed out that a specific set of hydrogen-bonding interactions appear to be required on all six sides of a shell protein hexamer in order for that hexamer to be incorporated into the MCP shell[37]. The structural requirement for symmetric interactions around the perimeter of a hexamer also limits the functionality of the central pore region when 6-fold symmetry is strictly obeyed. Indeed, all previous structures of BMC-domain homohexamers show narrow, chemically uniform central pores (CcmK, CsoS1, PduA, EutM) or no pores at all (PduU, EutS). While small pores through the MCP shell appear to be important for substrate and product flux, the chemistries that occur within most MCPs also require access to molecules that are clearly too large to pass through these small, simple perforations.

In order to create BMC-domain hexamers with more heterogeneous pore regions, nature employs a strategy based on pseudosymmetry (Figure 1.5). Early on, it was recognized through analysis of BMC-domain sequences, that within the BMC-domain family of shell proteins there is a subset that are "tandem" BMC-domain proteins[38]. As the name suggests, these polypeptides consist of two BMC-domains, which have been genetically fused to one another. As suspected, biochemical and structural evidence confirmed that these tandem BMC-domain proteins oligomerize to form pseudohexameric trimers, which fulfil the symmetry requirement for assembly in terms of their size, shape, and tiling interactions, but are more structurally diverse in their central pore regions as a result of differences between the two fused BMC-domains[29,33,34,39,39,40].

Studies of these tandem BMC-domain shell proteins have revealed specialized functions, demonstrating an important role for symmetry breaking in enhancing the versatility of the BMC-domain superfamily. Some examples of tandem BMC-domain proteins with specialized functions include EutL[27,34] and CsoS1D, which have been observed in multiple conformations that differ in their porosity, suggesting gated transport of large molecules, and also PduT, which is known to bind an iron-sulfur cluster at its central pore, providing evidence of involvement in some type of redox chemistry[29,31,35] (Figure 1.4). Within the MCP shell, each of these tandem BMC-domain pseudohexamers contributes its own specialized function, collectively imparting the shell with complex transport properties.

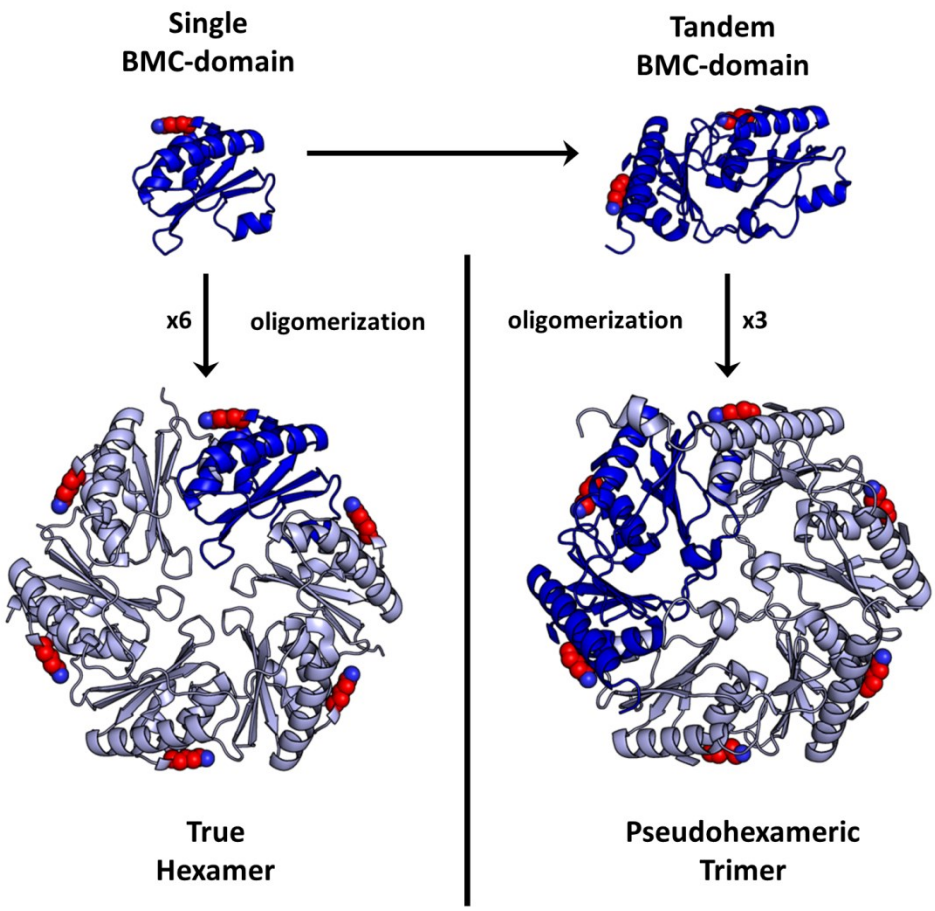


Figure 1.5. Symmetry, pseudosymmetry, and BMC-domain oligomerization. BMC-domain proteins typically form 6-fold symmetric hexamers, as depicted on the left. In the case of the tandem BMC-domain proteins, two BMC-domains become genetically fused, leading to a pseudohexameric trimer, as depicted on the right. Lysine residues that are critical for oligomerization are shown as red spheres, demonstrating that even when strict 6-fold symmetry is not present in the tandem BMC-domain oligomers, key symmetric interactions are preserved.

The relationship between broken oligomeric symmetry and function in the BMC-domain family of proteins is not unique. In fact, this relationship is somewhat common in proteins, and has been studied by biochemists, biophysicists, and structural biologists for decades[41]. Many examples of protein complexes that rely on various forms of broken symmetry for their function exist, dating as far back as the earliest X-ray crystal structures of proteins. In the early 1960's, the work done by Max Perutz and John Kendrew, which produced the first X-ray crystal structures of protein molecules, winning the 1962 Nobel Prize in Chemistry, led to the understanding that oxygen binding and release by hemoglobin relies on the pseudosymmetry inherent to the $\alpha_2\beta_2$ tetramer[42–44]. Shortly after, Paul Boyer proposed the "binding-change" mechanism (also called the "flip-flop" mechanism) for ATP synthase catalysis, which invoked coupling between symmetry breaking and enzymatic function in a large molecular machine[45]. Later, this mechanism was structurally verified by John Walker's research group[46], leading to another Nobel Prize shared by Boyer and Walker in 1997. Symmetry breaking is critical for the catalytic mechanism of HIV protease[47], an established target of structure-based drug design efforts, as well as for the regulatory activity of tropomyosin during muscle contraction[48], for the allosteric mechanism of bacterial L-lactate dehydrogenase[49], and for the ligand binding activity of the human growth hormone receptor[50]. The examples provided are not exhaustive, but are just a few prominent examples out of many available in the literature. Clearly, the importance of broken symmetry with respect to function cannot be underestimated in the description of oligomeric proteins.

Within oligomeric proteins that exhibit broken symmetry, there is a necessity for structural polymorphism within the oligomer. In a pseudosymmetric oligomer, such as a tandem BMC-domain pseudohexamer, or a hemoglobin $\alpha_2\beta_2$ tetramer, the structural polymorphism results from having different subunits with different sequences. Previous work on the BMC-domain proteins has highlighted a role for pseudosymmetry, as described above. Interestingly, in some tandem BMC-domain proteins, the structural polymorphism that results in pseudosymmetry also produces conformational polymorphism in the form of multiple pore conformations[27,34,40]. The conformational transitions observed in these proteins appear to behave according to the strict 3-fold oligomeric symmetry of the pseudohexamers. In addition to

pseudosymmetry, a different type of broken symmetry also exists, termed quasisymmetry, in which polypeptide chains with identical sequences adopt polymorphic conformations, typically induced by interactions with a ligand or with other copies of the protein subunit.

1.5 Open Questions in the Study of Molecular Transport through the MCP Shell

While we have gained tremendous insight into the mechanisms of molecular transport through the MCP shell, there are still many unanswered questions. Our primary understanding of these mechanisms has been derived from X-ray crystallographic studies of individual BMC-domain shell proteins. These structural studies revealed the central pores of the hexamers to be the conduits of molecular flux, and suggested that their physical and chemical properties lead to the selective permeability of the shell[25,27,29]. Studies of the tandem BMC-domain pseudohexamers demonstrated specialized functions in these proteins, including the opening of large pores as a result of conformational arrangements[27,34,40]. While structural studies have been highly informative, experimental confirmation and mechanistic descriptions of transport hypotheses derived from structural analysis is still somewhat scant. For example, the conformational changes observed in the EutL and CsoS1D shell proteins have clear functional implications; however, the mechanisms by which these motions are regulated have remained elusive, obscuring our understanding of the roles of these larger, gated pores in molecular transport. Additionally, we are still lacking a comprehensive understanding of the extent to which symmetry breaking yields specialized function in this highly-symmetric family of protein oligomers. Finally, the wealth of genetic information available from high-throughput sequencing is resulting in an ever-growing list of BMC-domain shell proteins with interesting sequence idiosyncrasies, implying potentially unique and divergent functions.

Gaining a full understanding molecular transport through the MCP shell will enable the use of MCPs for interesting synthetic biology and biotechnology applications. Recently, MCPs have received attention for their potential versatility in synthetic biology endeavours, providing genetically-encoded building blocks for nanoscale reaction vessels[51–54]. While our understanding of MCP assembly has

enabled targeting of specific protein contents to the MCP lumen[55–57], our understanding of molecular transport through assembled MCP shell is lacking. This gap in our knowledge of MCP transport functions has, thus far, prevented the creation of MCPs with designed transport properties. Unfortunately, targeting a designed metabolic pathway to the lumen of an MCP is useless unless those enzymes can access their substrates, and in order to effectively engineer MCPs with novel functions we must understand how to engineer MCP shells with appropriate permeabilities. Understanding how nature has designed molecular transport in the BMC-domain family of MCP shell proteins is a critical first step towards engineering novel MCP functions.

1.6 Summary of the Dissertation

The research presented as part of this dissertation is organized into four chapters (Chapters 2-5). Individually, each chapter described below represents a distinct research project. Collectively, Chapters 2-4 serve to illustrate the intimate connection between conformational polymorphism and specialized function in the BMC-domain superfamily of MCP shell proteins. Chapter 5 departs from direct description of biological phenomena, providing the details of a method for detection of structural heterogeneity in protein crystals using a novel mathematical analysis of X-ray diffraction data sets.

Chapter 2 describes a line of research directed toward understanding how an established conformational change is regulated in the EutL shell protein. EutL is a tandem-BMC-domain protein, as described above, which is found within the shell of the Eut MCP. In the case of EutL, the role of broken symmetry resulting from domain duplication became immediately obvious, owing to a series of four crystal structures determined by Tanaka, et al. and by Sagermann and Takenoya, et al.[27,34,58] These crystal structures show EutL in two distinct conformations, indicating that the broken symmetry creates the potential for interconversion between multiple conformations. Of these four structures, two reveal the trimer to be in a “closed” conformation, where the central region of the symmetric oligomer is tightly packed, leaving no pore through the center. The other two structures reveal the protein in an “open” conformation, where a structural rearrangement of the central loops creates a relatively large triangular

pore perforating the center of the oligomer. While the aforementioned structures demonstrated that EutL can occupy two distinct conformations, they did not provide a good explanation for how the interconversion between the two states is regulated. The work presented in Chapter 2 describes efforts to understand the roles of allostery and redox regulation in controlling pore conformation.

Chapter 3 describes the structure determination of a CcmK1 shell protein mutant (L11K) from the β-carboxysome of the cyanobacterium *Synechocystis* PCC6803, which presented challenges including twinning, non-crystallographic symmetry, and packing of hexameric units in a special arrangement. Following initial difficulties in space group assignment, the correct space group was clarified after model refinement revealed additional symmetry. This study provides an instructive example for crystallographers, in which broken symmetry requires a new choice of unit cell origin in order to identify the highest symmetry space group. Of biological relevance, additional observations related to the packing arrangement of molecules in this crystal suggest that these hexameric shell proteins might have lower internal symmetry than previously believed, implying some degree of structural flexibility. Furthermore, analysis of the electrostatic surface properties of this shell protein mutant yields some insight into the role of electrostatics in the transport of small molecules through the MCP shell.

Chapter 4 describes studies of an unusual BMC-domain hexamer from the Grp MCP. This shell protein, GrpU, whose amino acid sequence is particularly divergent from other members of the BMC-domain superfamily of proteins that effectively defines all MCPs[10]. Expression, purification, and subsequent characterization of the protein showed, unexpectedly, that it binds an iron-sulfur cluster. X-ray crystal structures of two GrpU orthologs provided the first structural insight into the homohexameric BMC-domain shell proteins of the Grp system. The X-ray structures of GrpU, both obtained in the apo form, combined with absorbance spectroscopy and computational modeling, show that conformational heterogeneity of a conserved amino acid motif leads to broken hexameric symmetry at the central pore of the cyclic assembly, resulting in a structurally polymorphic iron-sulfur cluster binding site that appears to be unique among metalloproteins studied to date.

Chapter 5 diverges from the direction of the previous three Chapters that aim to characterize specific biological systems, instead describing efforts to develop a method of crystallographic data analysis that can identify structural heterogeneity in crystalline protein samples. Chapters 2-4 describe BMC-domain hexamers as being among the many examples of the importance of conformational heterogeneity in protein function. Obviously, there is a general need to characterize conformational heterogeneity in protein samples. The work presented in Chapter 5 aims to address this need by using principal component analysis to examine multiple X-ray diffraction data sets collected from the same macroscopic crystal. We demonstrated that this method can detect structural heterogeneity at the level of the crystalline unit cell (i.e. unit cell nonisomorphism), and also can be potentially useful for creating electron density maps intended to highlight molecular features that display conformational heterogeneity.

Finally, Chapter 6 summarizes the work presented in the dissertation, discussing implications for the study of MCPs as well as for our understanding of protein function in general. Emphasis will be made on relating the individual research projects presented to the underlying idea that conformational polymorphism, which relates directly to broken oligomeric symmetry, is critical for specialized function in the BMC-domain proteins, ultimately imparting the MCP shells with complex transport properties.

CHAPTER 2

An Allosteric Model for Control of Pore Opening by Substrate Binding in the EutL Microcompartment Shell Protein

2.1 Introduction

The intestinal microbiota constitutes a fascinating ecological community. These microorganisms, primarily bacteria, which colonize the human digestive tract, are roughly an order of magnitude more numerous than human cells within a single human body[59–61], and they perform a number of functions that benefit their host. The various species of bacteria aid in carbohydrate digestion and produce vitamins that can be absorbed by the host[60–62]. Additionally, the ecology of the intestinal microbiota has been directly linked to human health, effecting many important aspects of physiology including immunological activity[63,64], tumor growth[60], obesity[65,66], and even cognitive function[67]. Consequently, imbalance within this ecological community can lead to a variety of human diseases. Typically, the intestinal microbiome maintains a healthy distribution of species as a result of competitive exclusion; no single species can become too prolific due to constant competition for resources within the environment[60,63].

In order to gain a competitive advantage over other microbes in the gut, certain bacterial species have evolved the ability to utilize ethanolamine as a metabolic resource[18,19,68–70]. Genomic analysis revealed that over 100 sequenced bacterial genomes contain the so-called "eut operon," which contains genes required for utilization of this small molecule[68]. Ethanolamine is produced through degradation of phosphatidylethanolamine, a phospholipid that is present in high quantities within cell membranes. Due to constant turnover of the intestinal epithelium, as well as the composition of the diet, ethanolamine is plentiful in the human intestine[19,71,72]. The ethanolamine molecule enters bacterial cells through a transmembrane ethanolamine transporter,[19] and is degraded by a well-characterized catabolic pathway[17,18,69]. Initially, ethanolamine is deaminated by ethanolamine-ammonia lyase (EutBC), providing

ammonia as a nitrogen source, and acetylaldehyde as a carbon source. Subsequently, acetaldehyde is ligated to coenzymeA, forming acetyl-CoA, which can enter a variety of other metabolic pathways. For example, acetyl-CoA can be oxidized in the TCA cycle to produce NADH, or it can be a substrate for anabolic processes such as fatty-acid biosynthesis.

The presence of the eut operon in the genome and metabolic utilization of ethanolamine have both been linked to bacterial enteropathogenesis[19,68,73,74]. Ethanolamine utilization allows pathogenic species to proliferate in the host digestive tract by giving them a nutritional advantage over other species, and also through potential modulation of the host innate immune response[19]. A probabilistic analysis of phylogenetic patterns by Li, et al. identified a strong correlation between the occurrence of eut genes and the "food poisoning" phenotype[73]. Experimentally, it has been demonstrated that global virulence regulators induce expression of eut genes in several species[75,76], and it has also been shown that eut genes are down-regulated during commensal colonization of the host by *Enterococcus faecalis*[77]. Furthermore, Winter and Thiennimitr, et al. concluded that inflammation of the host intestine due to infection is critical for anaerobic ethanolamine catabolism, because it provides the necessary respiratory electron acceptor, tetrathionate[78–80].

In most enteropathogenic facultative anaerobes that metabolize ethanolamine, the eut operon contains genes encoding BMC-domain proteins[17,19,68,73]. The presence of these genes is diagnostic of encapsulation of the Eut metabolic enzymes within a bacterial microcompartment (MCP) structure, and many organisms are known to make Eut MCPs. The function of these elaborate MCP structures is to retain acetaldehyde, a volatile and toxic metabolic intermediate[20,21] (Figure 2.1). Organisms known to form "eut MCPs" include *Clostridium*, *Salmonella*, *E. coli*, and *Klebsiella*, which can have a variety of negative effects on human health[60], ranging from food poisoning to increased risk of colon cancer, and it has been hypothesized that encapsulation of ethanolamine metabolism within a MCP is critical for pathogenesis in some bacterial species[19,68,73]. Encapsulation of ethanolamine metabolism within an MCP allows the bacteria to turn over large amounts of this metabolic resource without elevating the intracellular acetaldehyde concentration, giving them a nutritional advantage over other microbes[19].

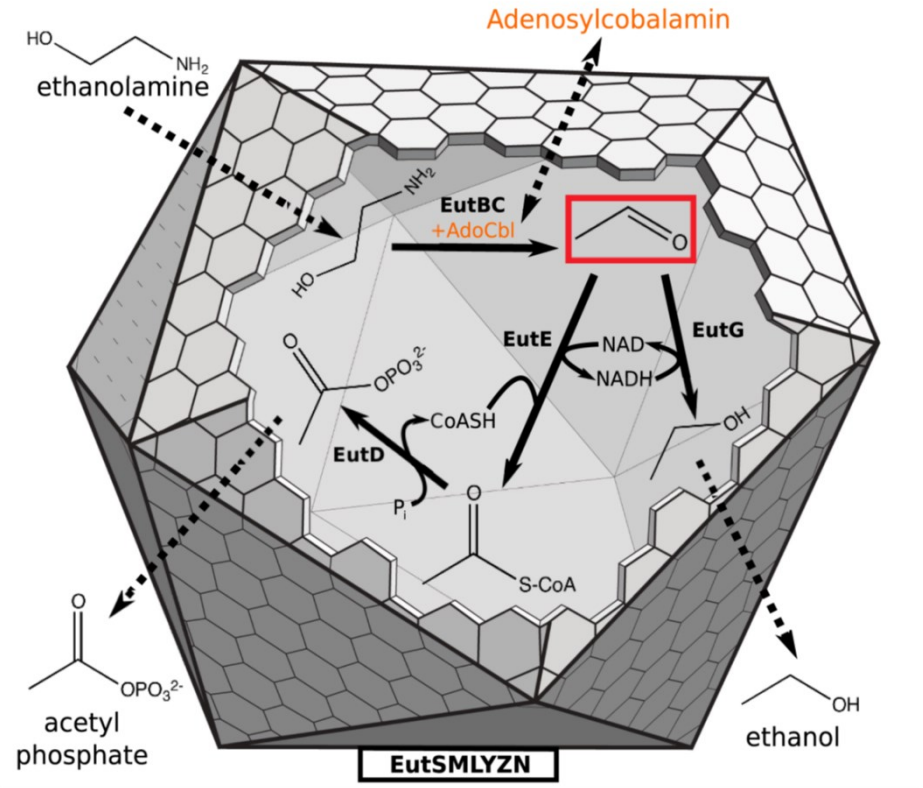


Figure 2.1. The Eut microcompartment. A schematic of the Eut microcompartment illustrates a layer of hexagonal shell protein tiles surrounding a series of enzymes that catalyze the depicted reactions. Solid arrows indicate chemical reactions, and dotted arrows represent molecular transport events. The shell must be permeable to substrates, products, and at least one bulky cofactor (shown in orange), while simultaneously retaining the acetaldehyde intermediate shown in the red box.

The shell of the Eut MCP has to solve a challenging problem. The reason for encapsulating ethanolamine catabolism is to retain a very small, volatile intermediate, acetaldehyde[20,21]. At the same time, the biochemical system within the Eut MCP must also exchange at least one large cofactor with the cytoplasm. The nicotinamide cofactors, as well as coenzymeA, can be recycled within the MCP[81,82], but the ethanolamine-ammonia lyase enzyme requires an adenosylcobalamin cofactor for catalysis[18,69] (Figure 2.1). In the enzyme active site, the adenosylcobalamin coenzyme (vitamin B_{12}) is subject to mechanism-based inactivation in the absence of substrate[83], and therefore needs to be replaced. The Eut operon contains ATP-dependent enzymatic machinery for regenerating adenosylcobalamin[84,85], but it is unclear whether this set of enzymes is located inside or outside of the MCP structure. If the cofactor regeneration system is outside the MCP shell, the shell must support the transport of cobalamin compounds to replenish the enzyme with fresh cofactor. If the cofactor regeneration system is inside the MCP, then ATP must be able to enter the MCP. Either of these molecules, cobalamin or ATP, are both much bigger than acetaldehyde, so the MCP shell must have a mechanism for allowing the passage of larger molecules without promoting the leakage of the small intermediate.

Previous X-ray crystallographic studies of the EutL shell protein from *Escherichia coli* have suggested an important transport role for this tandem BMC-domain pseudohexamer[27,34,58]. A pair of X-ray crystal structures of *E. coli* EutL determined by Tanaka, et al. revealed two conformations of the protein[27] (Figure 2.2). In the "closed" conformation, the central region of the trimer is occluded by ordered loop segments. In the "open" conformation a wide triangular-shaped pore is present at the center of the trimer. This pore – approximately 10Å across – is much wider than the pores in typical BMC homohexamers, and appears large enough for cofactor transport[27]. The idea that EutL might provide a gated pore for transport of large molecules explains how the MCP shell can provide interior enzymes with cofactors, while still retaining small intermediates like acetaldehyde. While previous crystallographic work has established that EutL can occupy two conformations, the regulation of the transition between the two states remains unclear. Understanding the mechanism by which these gated pores are regulated would substantially enhance our knowledge of how bacterial organelles maintain selective permeability.

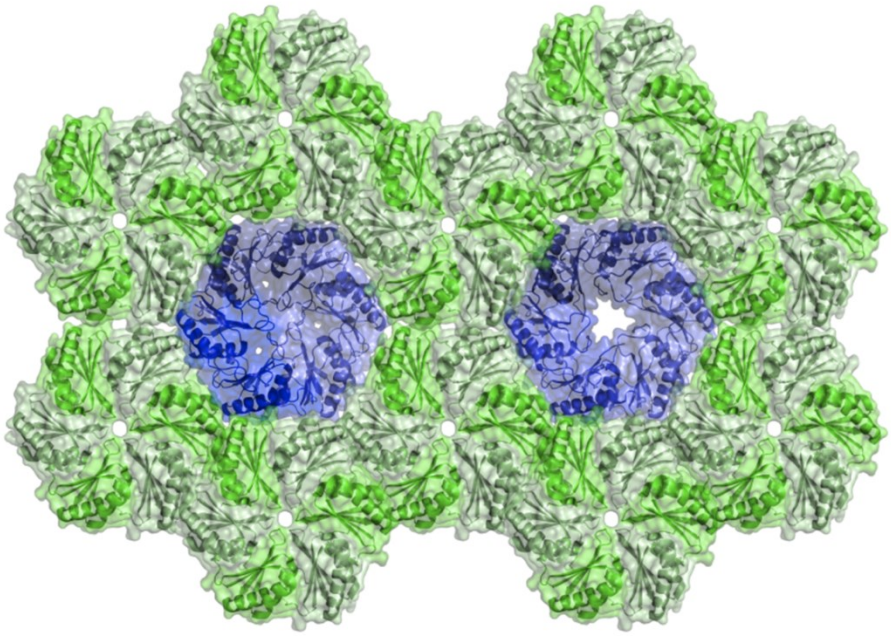


Figure 2.2. Open and closed conformations of EutL. X-ray crystal structures of E. coli EutL in both "closed pore" (3I82, left), and "open pore" (3I87, right) conformations. B) Both the open and closed EutL trimers (blue), along with EutM hexamers (green), are depicted as part of a layer of BMC shell proteins. Note the significant change in porosity caused by the interconversion of the open and closed conformations of EutL.

The goal of the work described in this chapter was to better understand the structural features responsible for conformational changes that occur in tandem BMC shell proteins like EutL. The present study focuses on understanding the atomic interactions that govern the transition between the two states. We address whether this conformational change is triggered by a specific signal, and propose, based on structural and biophysical data, that the substrate ethanolamine acts as a negative allosteric regulator of EutL pore opening.

2.2 Results – Initial crystallographic studies of EutL from *Clostridium perfringens*

Studies of EutL have provided important clues for understanding molecular transport across BMC protein shells. Previous structures of *E. coli* EutL supported the idea that the broken symmetry resulting from domain duplication allows for conformational change at the center of the trimer, which permits the opening of a large central pore[27,34]. While the importance of this conformational change in facilitating molecular transport seems clear, the details of how it takes place have yet to be fully elucidated. The aim of the study presented in this section was to shed light on this question through examination of an uncharacterized EutL homolog from *Clostridium perfringens*.

2.2.1 Structure determination of untreated *C. perfringens* EutL

We determined X-ray crystal structures of EutL from *Clostridium perfringens* (referred to hereafter as cpEutL) in order to elucidate structural features of the protein that contribute to its function. Using the hanging drop vapor-diffusion method, we grew crystals belonging to two different crystal forms (Figure 2.3a,b) representing space groups $P4_32_12$ (tetragonal) and $C2$ (monoclinic). These crystals diffracted to resolutions of 2.0 Å and 1.7 Å respectively. Further details of the data collection and atomic refinement for these structures are provided in the Methods section and in Table 2.1. Continuous electron density could be traced for residues 1-216 out of 225 in every polypeptide chain of all three structures, and additional residues could be modeled in a majority of the polypeptide chains. The two structures are very similar, having a coordinate RMSD of only 0.42 Å on all atoms. The coordinates and diffraction data

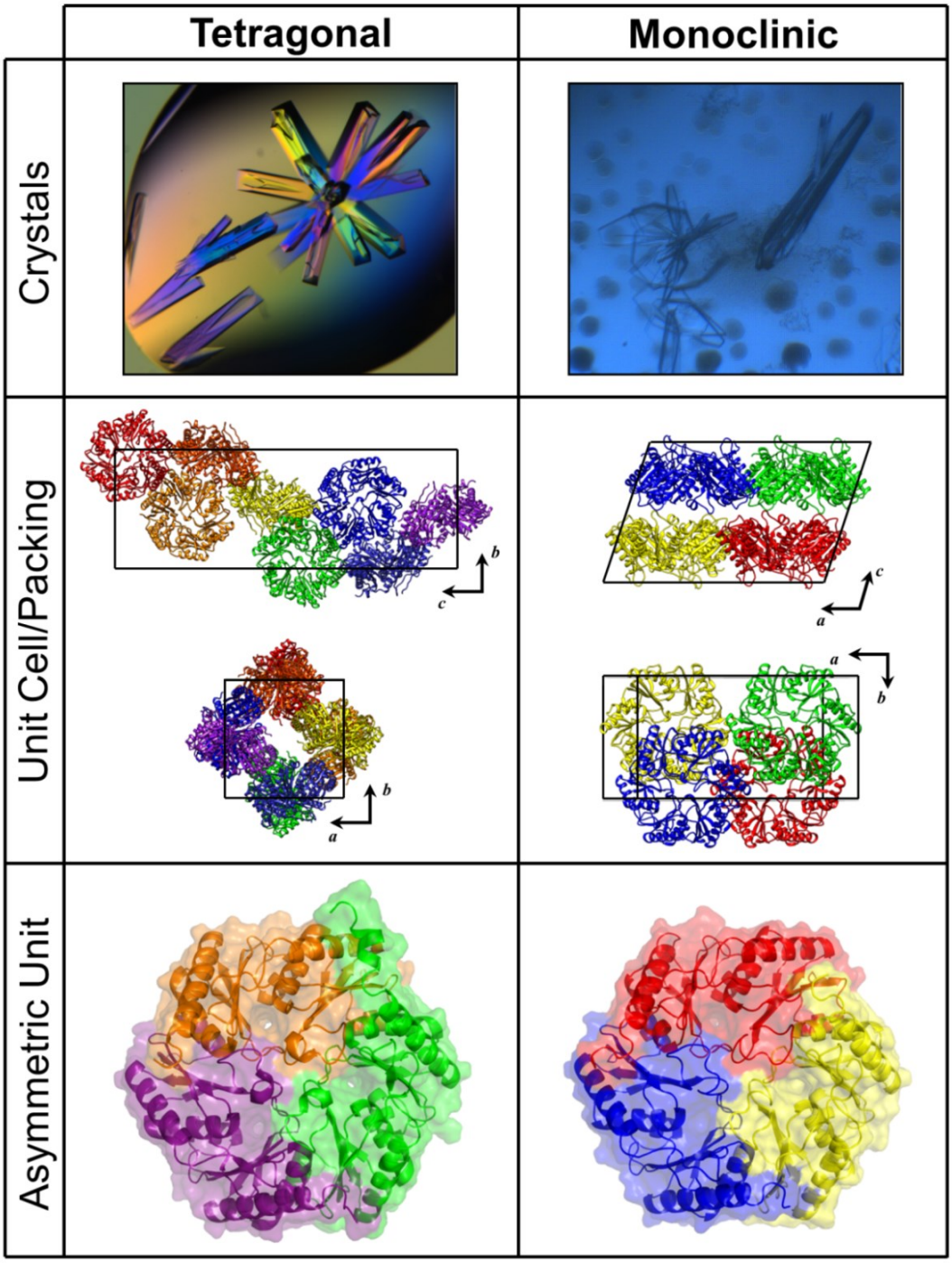


Figure 2.3. Two crystal forms of cpEutL. Comparison of the tetragonal (left) and monoclinic (right) crystal forms of cpEutL: crystal habit (top panel), unit cell and crystal packing (middle panel), and asymmetric unit (bottom panel).

Parameter	Tetragonal	Monoclinic	Oxidized	Reduced	Ethanolamine	Room Temp	Hydroxocobalamin
Wavelength	0.9795Å	0.9792Å	0.9792Å	0.9792Å	0.9795Å	0.9792Å	0.9795Å
Resolution Range	61.34-2.00Å (2.07-2.00Å)	33.4-1.7Å (1.74-1.70Å)	19.9-2.1Å (2.15-2.1Å)	19.8-1.8Å (1.85-1.8)	83.04-1.07Å (1.74-1.70Å)	72.87-1.90Å (1.95-1.90Å)	83.27-2.10Å (2.15-2.10Å)
Unit Cell	a=b=86.7Å, c=252Å α=β=γ=90°	a=118.44Å, b=66.03Å, c=79.38Å α=γ=90°, β=108.27°	a=b=87.8Å, c=252Å α=β=γ=90°	a=b=88.0Å, c=252Å α=β=γ=90°	a=b=87.96Å, c=251.91Å α=β=γ=90°	a=b=88.89Å, c=254.53Å α=β=γ=90°	a=b=88.23Å, c=252.01Å α=β=γ=90°
Space Group	$P4_32_12$	$C2$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$
Unique Reflections	65257	63737	57810	91132	109437	79392	110962
Multiplicity	4.0 (4.0)	3.7 (3.7)	4.5 (4.7)	4.7 (4.4)	12.5 (8.4)	4.5 (4.6)	6.7 (6.8)
Completeness	98.6% (99.4%)	99.5% (99.5%)	98.8% (96.7%)	98.7% (95.0%)	100% (100%)	97.6% (97.2%)	100% (100%)
$\langle I/\sigma I \rangle$	20.1 (6.4)	11.02 (1.62)	12.6 (3.8)	18.4 (3.9)	17.10 (1.79)	10.07 (1.33)	13.54 (2.01)
Wilson B-factor	25.7Å^2	16.86	21.1Å^2	19.8Å^2	25.3Å^2	29.9Å^2	47.06Å^2
$CC_{1/2}$	0.998 (0.952)	0.998 (0.587)	0.997 (0.925)	0.999 (0.894)	1.000 (0.718)	0.999 (0.570)	1.000 (0.944)
CC*	0.999 (0.988)	0.999 (0.860)	0.999 (0.980)	1.000 (0.971)	1.000 (0.914)	1.000 (0.852)	1.000 (0.985)
CC_{work}	0.957 (0.939)	0.968 (0.786)	0.964 (0.915)	0.969 (0.918)	0.970 (0.863)	0.981 (0.799)	0.939 (0.903)
CC_{free}	0.949 (0.914)	0.956 (0.696)	0.954 (0.888)	0.960 (0.927)	0.966 (0.833)	0.956 (0.794)	0.910 (0.853)
R_{work}	0.170	0.1634	0.164	0.160	0.1470	0.1379	0.2012
R_{free}	0.193	0.1951	0.191	0.177	0.1623	0.1591	0.2420
No. Atoms	9983	5380	10383	10655	5658	5401	4877
Protein Residues	659	653	663	663	663	661	650
Solvent Molecules	245	504	421	619	524	221	37
Average B-factor	29.4Å^2	24.1Å^2	23.9Å^2	24.3Å^2	30.9Å^2	35.8Å^2	56.3Å^2
$RMSD_{bonds}$	0.007Å	0.009Å	0.013Å	0.006Å	0.010Å	0.009Å	0.015Å
$RMSD_{angles}$	1.1°	1.15°	1.4°	1.0°	1.246°	1.137°	2.055°
Rama. Plot: Favored Allowed Outliers	98% 2% 0%	98.35% 1.65% 0%	98% 2% 0%	98% 2% 0%	98.85% 1.15% 0%	99.57% 0.43% 0%	97.25% 2.75% 0%
MolProbity Clashscore	3.6	1.34	4.0	3.8	0.49	0.39	1.77
PDB ID	4EDI	4TLH	4FE0 (withdrawn)	4FDZ	4TME	4TM6	N/A (preliminary)

Table 2.1. Data collection and refinement statistics for seven cpEutL structures.

Values in parenthesis reflect the highest resolution shell. R-factors are calculated with riding hydrogens present in the model.

[a] R_{work} and R_{free} are given by the following equation, computed for the working and test sets of reflections respectively: $R = \frac{\sum_{hkl} \| F_o | - | F_c \|}{\sum_{hkl} | F_o |}$

for the two structures have been deposited in the Protein Data Bank under PDB IDs 4EDI (tetragonal), and 4TLH (monoclinic).

2.2.2 Structural overview

The structures of cpEutL recapitulate the typical topology and oligomerization reported previously for tandem BMC-domain protein[27,29,34,35,38–40] (Figure 2.3). Within a single monomer, each of the two genetically fused BMC-domains adopts a permuted α/β BMC-domain fold. Domain 1 contains β-strands 1-4 (forming an antiparallel β-sheet) and α-helices A-C. Domain 2 contains β-strands 5-8 (also an antiparallel β-sheet) and α-helices D-F. A single trimer is present in the asymmetric unit of both crystal forms, which corresponds to the expected biological assembly of the protein (Figure 2.3c). The three polypeptide chains of the trimers assemble in the C3-symmetric manner typical of tandem BMC-domain proteins, giving the trimer the characteristic hexagonal shape required for its incorporation into the two-dimensional lattice of the MCP shell. In the monoclinic crystal, lattice formation creates layers of two-dimensional sheets, consistent with assembly of the flat facets that form the MCP shell. The structures of untreated cpEutL superimpose upon the closed conformation of their E. coli homolog (PDB ID: 3I82) with an RMSD of less than 1.0 Å on C-alpha atoms.

2.2.3 Loop structures in the closed conformation

The crystal structures of cpEutL reveal the pseudohexameric trimer in a closed-pore conformation that is very similar to the closed conformation seen for E. coli EutL[27,34,58]. The closed conformation of cpEutL, like the homologous E. coli protein, is characterized by a tight packing of three symmetry-related loops, one from each polypeptide chain in the trimer, about the three-fold molecular symmetry axis (Figure 2.4). These loops consist of residues 68-82, which connect β-strands 3 and 4 in domain 1. The packing of these loops at the center of the trimer is stabilized by the edge-to-face interaction of three aromatic rings, and by hydrogen bonding between asparagine residues. Each monomer contributes a tyrosine residue (Tyr69) and an asparagine residue (Asn74) to this symmetric interaction (Figure 2.4). In

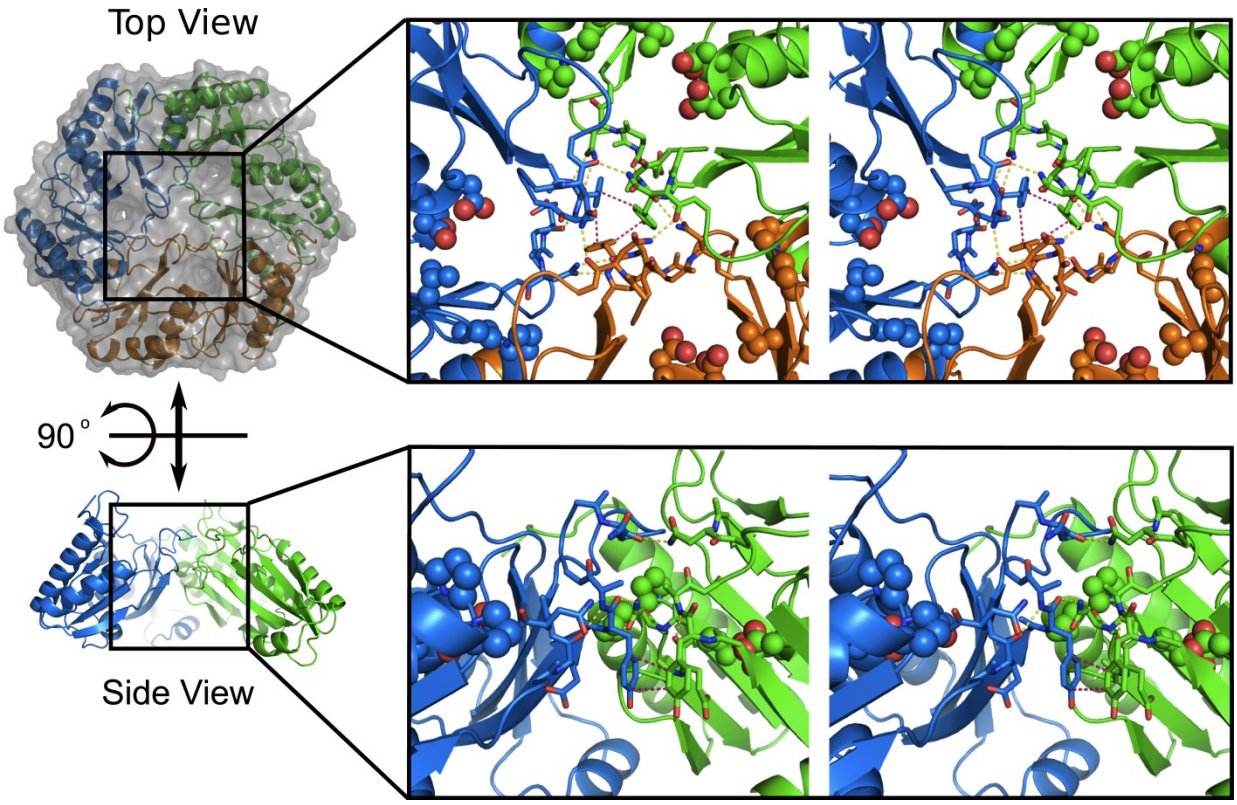


Figure 2.4. Overall structure of cpEutL and details of loop interactions. A ribbon diagram depicts the crystallographic model of untreated cpEutL. The trimeric structure is very similar to the closed conformation of E. coli EutL. Expanded views (stereo) show the details of the atomic interactions that stabilize the closed-pore conformation. Hydrogen bonds are shown as dotted yellow lines, and interactions between aromatic rings of tyrosine side chains are shown as dotted pink lines. Residues that form these interactions are shown as sticks (Y69, A70, N74, E181, N183). The orange subunit has been removed for clarity in the lower panel.

addition to the β3-β4 loop, the analogous loop region in domain 2, consisting of residues 176-185, which connects β-strands 7 and 8 in domain 2, also protrudes toward the center of the trimer. This β7-β8 loop makes contacts, both direct and water-mediated, with the β3-β4 loop within a single monomer, and between neighboring monomers in the trimer (Figure 2.4). These interactions between the β7-β8 loops and the β3-β4 loops may be important for stabilizing the closed conformation. While the conformations of both the β3-β4 and β7-β8 loops in cpEutL are consistent with the closed structure of EutL from *E. coli* (PDBIDs 3I82 and 3GFH), both of those loop regions are in different and/or disordered conformations in the open structures of the *E. coli* homolog (PDB IDs 3I87 and 3MPV).

2.2.4 Loose atomic packing in the monomers

In line with previous reports[27,58], we observe small holes in areas of low packing density, one within each of the monomers, in the closed-pore conformation of the trimer (Figure 2.4). These holes, present also in previous structures of EutL homologs in the closed conformation (PDB IDs 3IO0, 3I82, 3GFH, 4FAY), are located at the N-terminal end of helix A between domains 1 and 2 within an individual monomer. The openings extend from one face of the trimer to the other, making three narrow, hourglass-shaped channels through the assembly (Figure 2.4). The openings are lined with conserved acidic and polar residues on one side (D44, D45, T182), and hydrophobic residues on the other (F112, F184). The loose atomic packing in this region appears to provide space to accommodate conformational rearrangements and the formation of alternative interactions in the open-pore conformation of the trimer (Figure 2.5).

We measured the minimum radius of the opening using the HOLE2 algorithm[89], and found that the narrowest constriction along a traversal from one side of the trimer to the other has a radius of only 1.2 Å (Figure 2.6). This measurement is in close agreement with the analysis of E. coli EutL performed by Sagermann, et al., who calculated a radius of 1.1 Å for the narrowest point[58]. The narrow constriction points of the ligand-binding channels consist of residues D44, D45, V151, T182, and F184. For comparison, the van der Waals radius for a carbon atom is approximately 1.7 Å[90–93]. Whether these

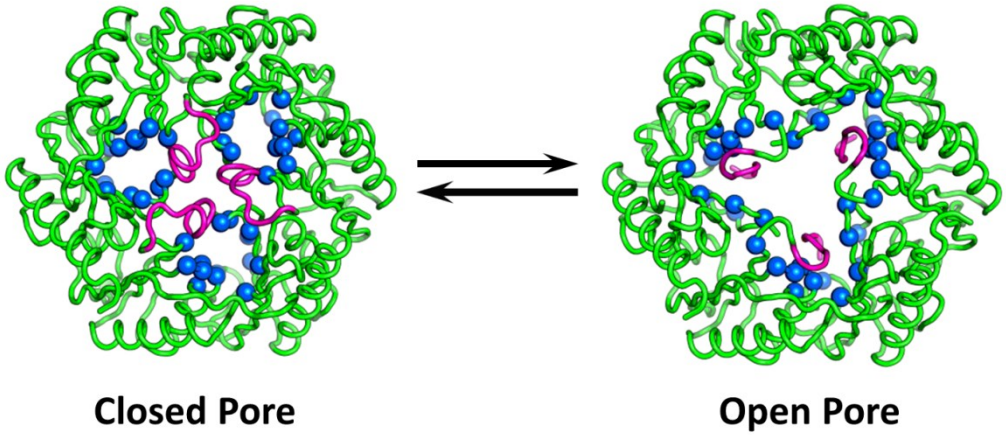


Figure 2.5. Conformational rearrangement of the EutL pore. Three narrow channels through the EutL trimer (one per monomer) are depicted in blue. Residues lining these channels are shown with spheres surrounding their Cα atoms. In order for the EutL pore to open, the β3-β4 loops (magenta) must retract from the center of the oligomer into these narrow channels.

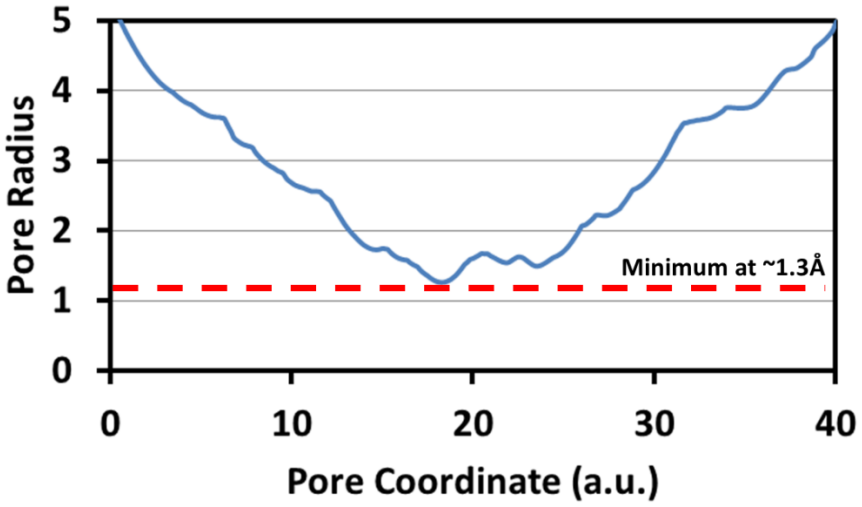


Figure 2.6. Dimensions of the channels in EutL. The plot shows pore radius as a function of arbitrary pore coordinate for one of the narrow channels that perforate the EutL trimer. The narrowest point of the channel has a radius of only 1.2Å.

openings could support molecular transport (perhaps upon further rearrangement) is an open question. In our view, the narrowness of the openings and the presence of larger pores in other Eut shell proteins belie an alternate explanation.

2.2.5 Crystallographic evidence for disulfide bonding in cpEutL

Analysis of electron density maps calculated from cpEutL diffraction data revealed, unexpectedly, a disulfide bond in a tandem BMC-domain protein. The refined structure of cpEutL reveals an interesting triad of cysteine residues – Cys127, Cys196, and Cys200 – poised within disulfide-bonding proximity. This triad, well conserved among EutL orthologs, consists of a helical CXXXC motif and a third Cys residue contributed by a neighboring β-strand, which creates an isosceles triangle with the β-carbons of the three cysteine side chains at the vertices. In our structure of cpEutL, Cys127 of chain A appears in two alternate rotameric configurations, with one of those rotamers able to form a disulfide bond with Cys200. Electron density corresponding to such a disulfide bond appears (at a contour level of 1.6 σ) in a *2mFo-DFc* map (Figure 2.7a). Initial refinement was performed using a model that did not contain a disulfide bond between those two residues. We modeled the disulfide-bonded rotamer of Cys127 only after positive density appeared in a difference (*mFo-DFc*) map. Subsequent refinement was carried out with Cys127 in two partially-occupied rotamers; the occupancies in the final refined model were 0.46 and 0.54 for the bonded and non-bonded conformations respectively. To minimize phase bias, we also calculated a difference map in which the two cysteine residues of the disulfide bond were excluded from the structure factor calculation. This map also showed positive density (4.9σ) for both rotamers of Cys 127 as well as for the Cys 127-Cys200 disulfide bond (Figure 2.7b).

Because disulfide bonds are sometimes broken by intense synchrotron X-ray radiation[90–93], we also studied the effect of X-ray radiation on the electron density features corresponding to the disulfide bond. First, a difference map was recalculated from diffraction data collected during the first 20° of the X-ray experiment, which gave a data set that was approximately 82% complete. That density for the disulfide was stronger in this map than in the original map calculated from the full data set (2.2σ versus

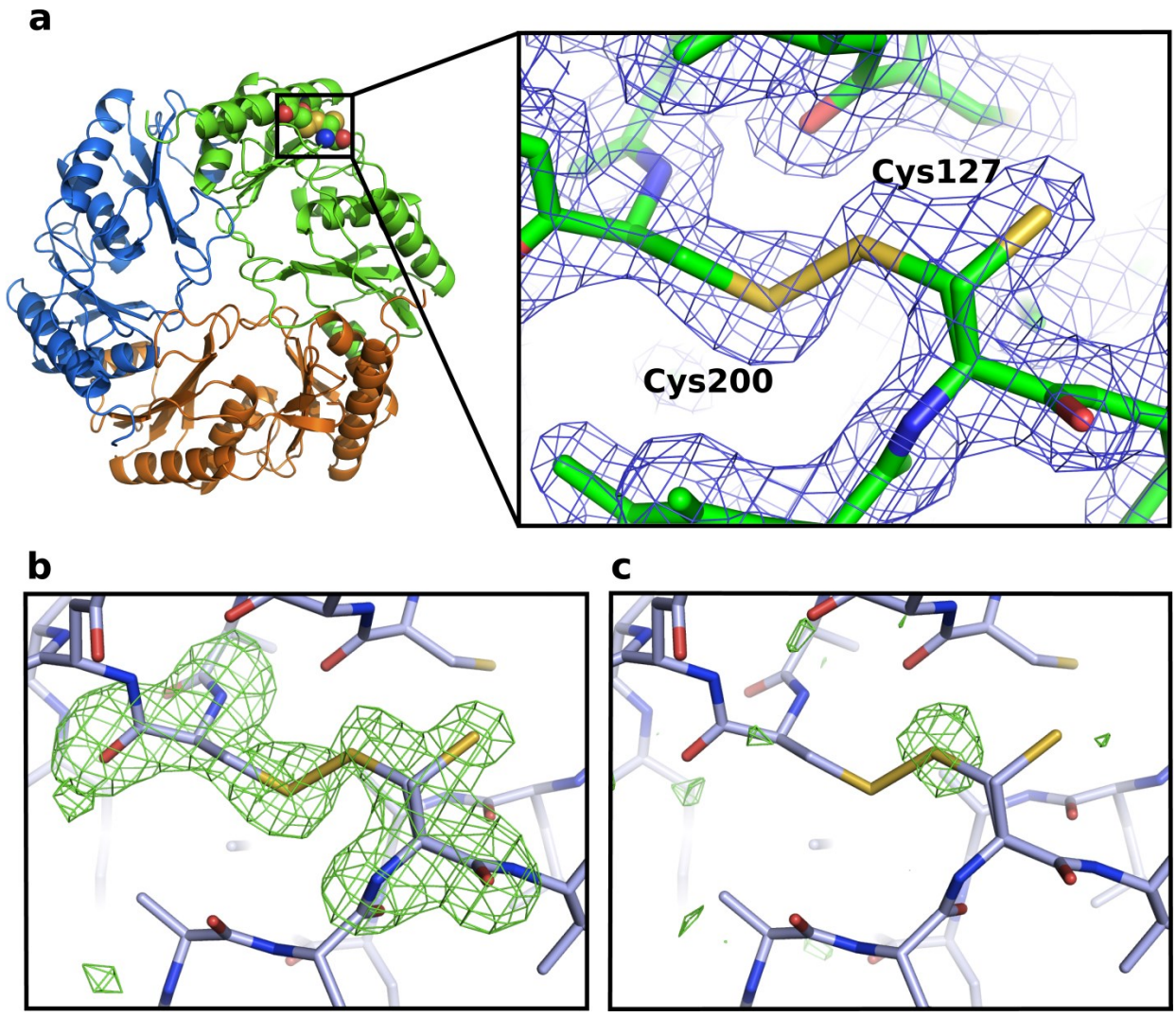


Figure 2.7. Electron density features reveal a disulfide bond in cpEutL. A) A 2mFo-DFc electron density map contoured at 1.1σ shows contiguous electron density between the sulfur atoms of Cys127 and Cys200, indicating the presence of a partially occupied disulfide bond. B) A difference (mFo-DFc) electron density map, in which the atoms of Cys127 and Cys200 have been excluded from the structure factor calculation to eliminate phase bias, shows strong positive density for both conformations of Cys 127 and for the disulfide bond. C) A Fearly-Flate, electron density map, comparing data from the first 20° of X-ray data collection with the last 20°, indicates that the Cys127-Cys200 disulfide bond is strongly populated prior to X-ray exposure, and is partially destroyed during the experiment.

1.6σ), suggesting that the disulfide bond was partially broken during the course of data collection. To further demonstrate the radiation-sensitivity of the disulfide bond, we calculated a F_o-F_o isomorphous difference map comparing data from the first 20° of the X-ray experiment with data from the last 20° of the experiment (F_{early}-F_{late}). This map revealed a strong peak of positive density surrounding the sulfur atom of the bonded rotamer of Cys127 (Figure 2.7c). These X-ray calculations indicate that the disulfide bond was well populated in the crystallized protein, with partial loss occurring as a result of radiation damage.

2.2.6 Analysis of unassigned electron density

During atomic refinement, unassigned electron density became visible, which was incompatible with the closed-pore conformation of the protein that had been built. These prominent electron density features could be explained by modeling a small population of the protein molecules in the crystal as though they instead occupied a putative open conformation. Much of the unexplained density appeared running through the centers of the small openings described above (Figure 2.8a,b). Similar density features were present in both *2mFo-DFc* and *mFo-DFc* (difference) electron density maps. We attempted to model this density with water molecules, as well as several components of the mother-liquor including HEPES buffer and ethylene glycol, but none of those modeling efforts produced a reasonable fit to the density as judged by geometric considerations and the appearance of negative density after subsequent refinement.

Motivated by geometric inconsistencies between the unassigned density and our crystallographic model, we noticed that the unassigned density resides where the protein backbone of the β3-β4 loop would be expected if the EutL protein were in an open conformation (based on analogy to the open structure of E.coli EutL) (Figure 2.8b). Using our crystal structure of cpEutL and the open structure of E. coli EutL (PDB ID 3I87), we created a homology model of the open conformation of cpEutL. We used this approximate model, along with our crystallographic model of the closed conformation, for a

simplistic "occupancy refinement" calculation, which revealed conformational heterogeneity within the cpEutL crystals.

Briefly, we calculated composite structure factors as a linear sum of occupancy-weighted structure factors derived from closed and open models of cpEutL:

$$F_{calc} = \alpha F_{c,closed} + (1-\alpha) F_{c,open}$$

A traditional R-value between calculated and observed structure factor amplitudes was used to judge the agreement with observed diffraction data. Whether the R-value was based on reflections in the 'working set' (R_{work}) or the 'free set' (R_{free}), the minimum residual error was obtained when $\alpha \cong 0.93$ (Figure 2.8c). This result is consistent with the conclusion that the cpEutL crystals are structurally heterogeneous, with most of the molecules in the closed conformation, and about 7% of the molecules in an open conformation resembling the open conformation of E. coli EutL.

2.3 Results – Experimental validation and functional characterization of disulfide bonding in EutL

Our X-ray crystal structure of cpEutL revealed the first disulfide bond in a tandem BMC-domain protein. The presence of the disulfide bond in the crystallographic electron density, and the conservation of the cysteine triad that contains the bond, suggest a functional importance. Because the function of the EutL shell protein is related to a conformational change, we hypothesized a coupling between the redox state of the disulfide bond and the conformational equilibrium of the pore. This section describes work directed toward establishing the hypothesized connection.

2.3.1 X-ray crystal structures of the oxidized and reduced states of EutL

The same crystallization conditions that yielded tetragonal crystals of the untreated protein also gave crystals of a protein sample that had been chemically reduced by treatment with 5mM tris-(2-carboxyethyl)phosphine (TCEP), as well as a sample that had been oxidized by treatment with 10mM hydrogen peroxide (H_2O_2). Crystals of the TCEP- and H2O2-treated samples diffracted to resolutions of 1.8 Å and 2.1 Å respectively. Further details of the data collection and atomic refinement for these

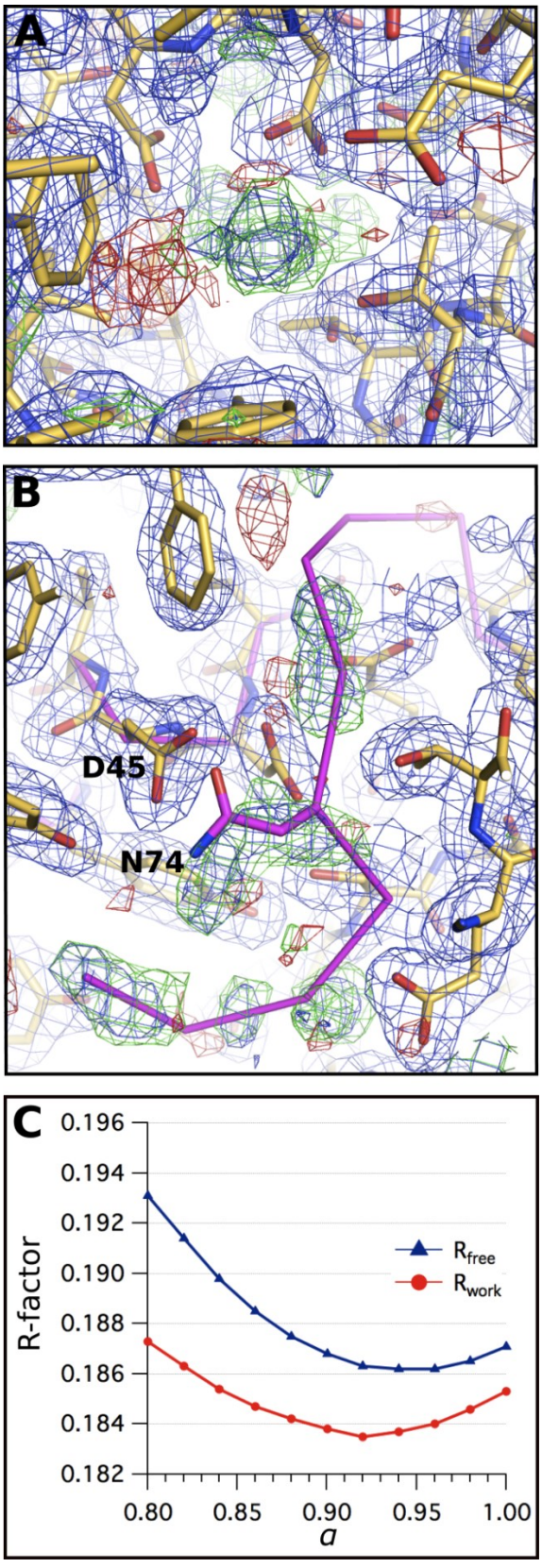


Figure 2.8. Attempt to model conformational heterogeneity in cpEutL. A) Extra, unassigned electron density is present in regions of low packing density. View from the top surface of the cpEutL trimer (yellow). The 2mFo-DFc map contoured at 1.2σ is shown in blue, and the mFo-DFc map contoured at ±3.0σ is displayed in green and red, representing positive and negative features respectively. B) The unassigned density is displayed at an angle rotated approximately 90° from the view shown in A. A homology model of the open conformation is shown as a Cα trace (magenta), which passes directly through the unassigned density. The homology modeling places Asn74 in the putative open conformation in a position in which it could make hydrogen bonds with the highly conserved Asp45 residue, a feature that is consistent with positive density in the mFo-DFc map. C) The average residual error for structure factor calculations (Rwork,Rfree) is minimized by modeling the closed conformation at fractional occupancy (α), and the putative open conformation at complementary occupancy (1-α). The occupancy calculations were performed using a 4.5Å resolution cutoff.

structures are also available in the Methods section and in Table 2.1. Final atomic models derived from the oxidized and reduced samples were strikingly similar to one another, and also to the initial cpEutL structure determined from an untreated sample. These structures superimpose with an RMSD of only 0.2 Å (Cα atoms). It is noteworthy that the electron density features surrounding the Cys127-Cys200 disulfide bond are nearly identical in the maps derived from the oxidized crystals and from the initial untreated crystals (Figure 2.7), indicating that cpEutL is prone to spontaneous oxidation when exposed to an aerobic atmosphere. Also, the disulfide bond is completely absent in the electron density map derived from the reduced crystals (Figure 2.9). The coordinates of the reduced structure were deposited in the Protein Data Bank under PDB ID 4FDZ, and the oxidized structure was deposited, but later withdrawn due to its high degree of similarity with the untreated structure, which was deposited previously (4EDI).

To further illustrate the potential for cpEutL to form disulfide bonds under oxidizing conditions, we calculated $F_{oxidized}$-$F_{reduced}$ difference maps from the corresponding isomorphous datasets, using model phases. Those difference maps showed decisively strong positive density features (≥5.4σ) situated between the sulfur atoms of Cys127 and Cys200, indicating that this disulfide bond is present in the oxidized protein, but not in the reduced sample (Figure 2.10). Interestingly, similarly strong positive density can be seen between the sulfur atoms of Cys127 and Cys196 (Figure 2.10). This positive density is accompanied by negative density around the Cys196 residue (since it had been modeled in a non-disulfide bonded conformation), consistent with the interpretation that Cys196 would have to occupy a different rotameric configuration to form a disulfide bond with Cys127. The observed electron density features imply that the oxidized crystals might contain a mixture of Cys127-Cys200 and Cys127-Cys196 disulfide bonds.

2.3.2 Two-dimensional SDS-PAGE

Tryptic digestion of the cpEutL protein followed by two-dimensional SDS-PAGE demonstrates that tryptic fragments are held together by disulfide bonds formed in the native state. Starting with an oxidized cpEutL sample, we used iodoacetamide to alkylate any cysteine residues not involved in

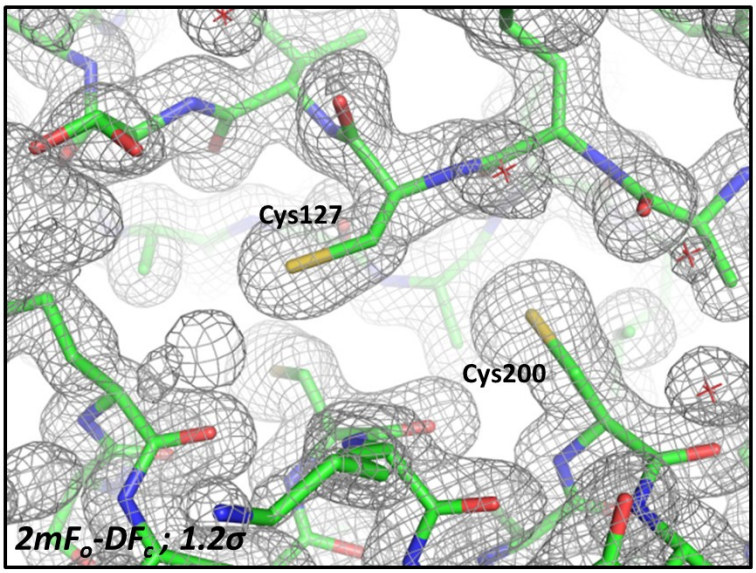


Figure 2.9. Chemical reduction abolishes the disulfide bond in cpEutL electron density. Representation of 2mFo-DFc electron density surrounding the Cys127/Cys200 residue pair in maps corresponding to the reduced cpEutL structure. There is no trace of any density corresponding to a disulfide bond between the thiol groups.

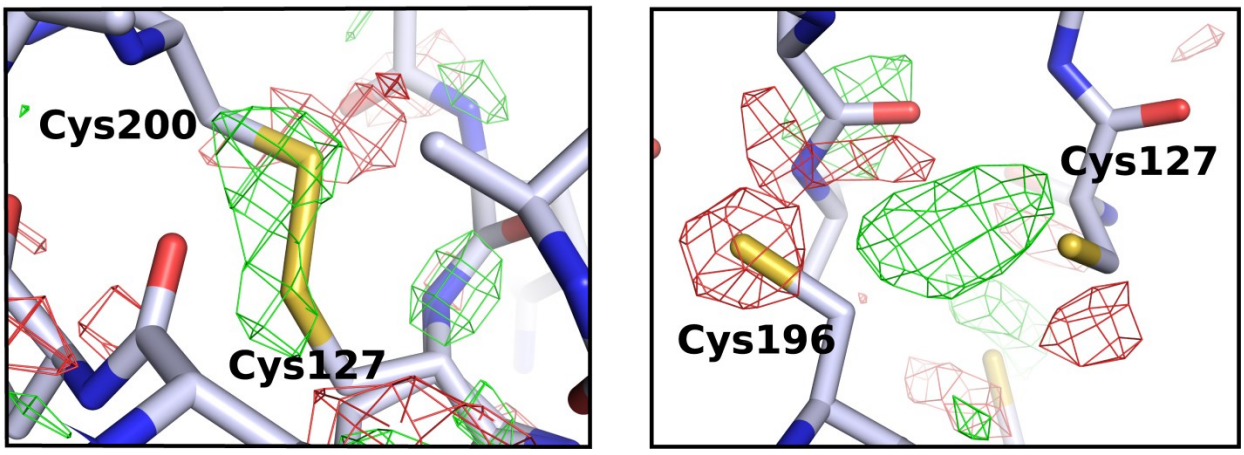


Figure 2.10. Crystallographic comparison of oxidized and reduced cpEutL. Two images of $F_{oxidized}$-$F_{reduced}$ isomorphous difference maps show positive peaks in regions between sulfur atoms, demonstrating that disulfide bonds exist beyond those visible in the normal $2mF_o$-DF_c maps. The maps are contoured at 3.0σ.

disulfide bonds and then digested the protein sample using trypsin. We then used the digested sample in a two-dimensional SDS-PAGE experiment in which the first dimension was run under non-reducing conditions and the second dimension was run under reducing conditions. Staining the gel with coomassie blue revealed several distinct bands corresponding to tryptic fragments of cpEutL. All of the major bands fell on the gel diagonal with the exception of one, which appeared below the diagonal (Figure 2.11a). The presence of this band below the gel diagonal demonstrates that one or more tryptic fragments migrate as lower molecular weight species when chemically reduced, indicating that those fragments are involved in disulfide bonds in the oxidized state.

2.3.3 Mass spectrometry

Using size-exclusion chromatography coupled to electrospray ionization mass spectrometry (SEC-MS), we observed multiple species of cpEutL with different numbers of intramolecular disulfide bonds. In the mass spectrum of the non-reduced sample, peaks corresponding to cpEutL form a bimodal distribution, with maxima at 23,701Da and 23,705Da, indicating two species whose molecular masses differ by 4Da (Figure 2.11b). When this protein sample was chemically reduced by the addition of 50mM DTT, the distribution shifted in favor of the higher molecular weight species, demonstrating its increased abundance (Figure 2.11b). Because these two species have a difference of 4Da, and their relative abundances are shifted to favor the more massive species upon reduction, we conclude that the higher molecular weight species corresponds to a polypeptide with two fewer disulfide bonds, and therefore four additional H atoms. We note that cpEutL contains six cysteine residues, and therefore has the potential to form multiple disulfide bonds. The mass spectrometry experiments support the possibility of complex disulfide bonding schemes, consistent with the crystallographic evidence showing multiple disulfide bonding arrangements for cysteines 127, 196, and 200. At least one additional cysteine residue, whose identity is yet unknown, would have to participate in a second disulfide bond in the oxidized form.

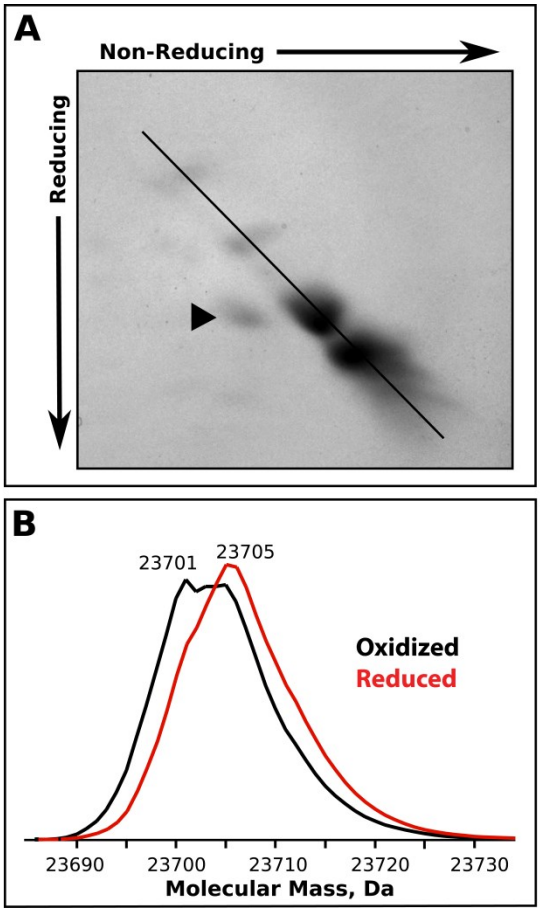


Figure 2.11. Biochemical evidence for intramolecular disulfide bonding in cpEutL. A) A two dimensional SDS-PAGE experiment using a trypsin-digested cpEutL sample and run under non-reducing (first dimension) and reducing (second dimension) conditions shows spots below the gel diagonal, demonstrating that cpEutL contains disulfide bonds formed in the native state. The triangle denotes the predominant band below the diagonal. B) Mass spectra (SEC-MS) of non-reduced, native cpEutL (black) and chemically reduced, but otherwise identical, cpEutL sample (red) reveals two species with molecular masses of 23701Da and 23705Da. The comparison of mass spectra from otherwise identical samples comparing oxidized and chemically reduced cpEutL samples shows that the higher molecular weight species is favored under reductive conditions, indicating a 4 Da mass difference resulting from reduction of two disulfide bonds between the two states. The mass spectrometry signal is plotted on a linear scale with arbitrary units.

2.3.4 NMR spectroscopy

NMR spectra of cpEutL, collected under oxidizing and reducing conditions, illustrate that a structural change accompanies a change in the redox environment. We collected ^{1}H-^{15}N TROSY-HSQC spectra of both oxidized and reduced cpEutL in order to determine if a change in the redox environment might induce structural changes in the protein. Comparison of two-dimensional ^{1}H-^{15}N TROSY-HSQC spectra collected under oxidizing and reducing conditions revealed that, while the majority of the spectral features remained unchanged, there are notable differences (Figure 2.12). The presence of differences in position and/or intensity for a small subset of the NMR peaks is consistent with a structural rearrangement in limited parts of the protein. To rule out the possibility that the change in NMR signal was due to a change in oligomerization or partial aggregation of the protein (e.g. due to spurious intermolecular disulfide bonding), we performed SDS-PAGE and native PAGE on both NMR samples. Results of both native and SDS-PAGE appear essentially identical under oxidizing and reducing conditions.

2.3.5 Bioinformatic analysis of disulfide bonding

Motivated by our observation of disulfide bonding in cpEutL, we sought to determine if disulfide bonds might be common among the tandem BMC domain proteins. At the time of our analysis, the Pfam database contained 3317 unique entries within the BMC domain protein family (PF00936), 969 of which were tandem BMC domain proteins. We filtered these sequences based on a 95% identity criterion, which produced a set of 840 single and 268 tandem domain sequences.

Analysis of potential disulfide bonds in this set of BMC proteins was complicated by the multiple distinct topological forms these proteins adopt, arising from circular permutations and alternate arrangements (e.g. clockwise vs. counterclockwise) of tandem domains in the context of a pseudohexameric trimer. As this would have prevented (linear) sequence alignment to EutL in many cases, we adapted a method from our earlier work[94,95], using sequence-structure mapping to identify prospective disulfide bonds (Figure 2.13a). This approach was enabled by the availability of structural

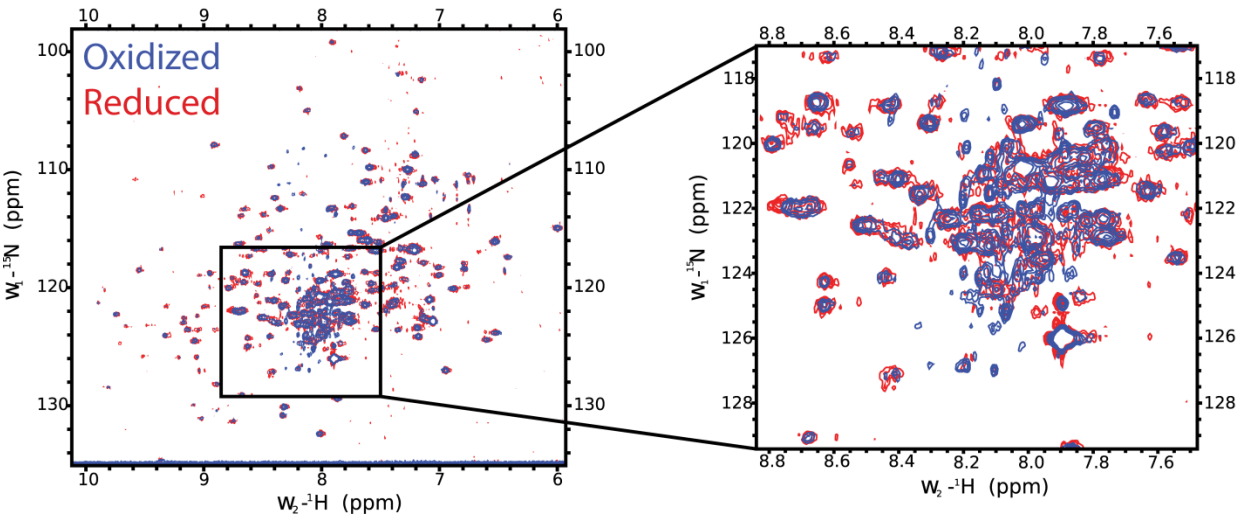


Figure 2.12. NMR spectroscopy supports a redox-modulated structural change. Two ^{1}H-^{15}N TROSY-HSQC NMR spectra of cpEutL, collected under oxidizing (blue) and reducing (red) conditions, have notable differences in chemical shift and intensity for a subset of peaks, consistent with a structural rearrangement due to changing electrochemical environment.

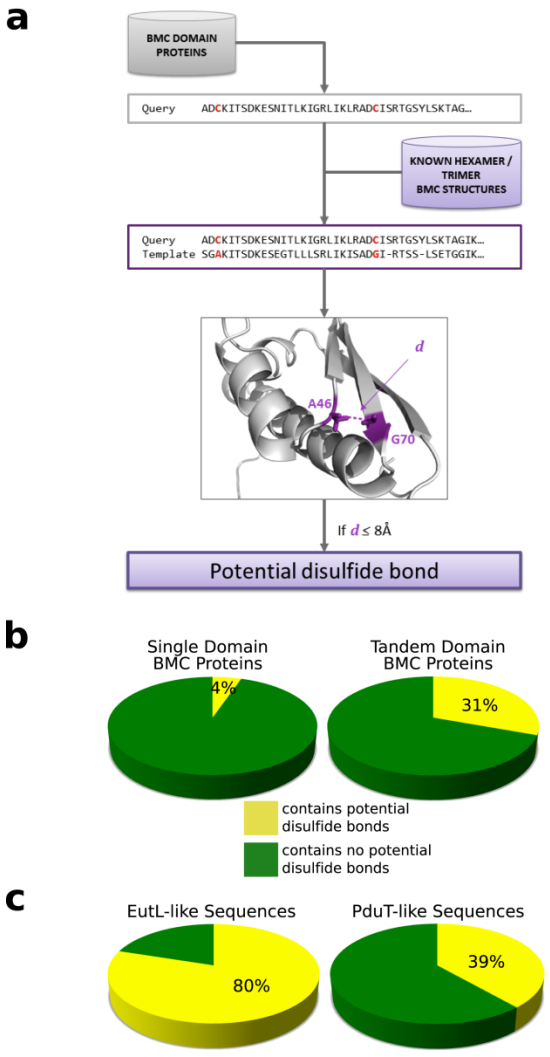


Figure 2.13. Disulfide bonding is widespread among some tandem BMC domain proteins. A) A flow-chart describes the computational methodology used to find potential disulfide bonds in BMC shell proteins. B) The results reveal a much greater abundance of disulfide bonds in tandem BMC domain proteins overall compared to their single domain counterparts. C) Further investigation of the sequences shown to contain putative disulfide bonds reveals that disulfide bonds are common to specific types of tandem domain shell proteins, notably those most similar to EutL and PduT.

templates for BMC proteins in several different topological forms. For each unique Pfam entry (query sequence), we identified the closest BMC shell protein homolog whose structure had been experimentally determined (template structure). We mapped each query sequence onto the best hexameric or trimeric template structure and measured the Cα-Cα distances for all pairwise combinations of cysteine residues, including those that might exist between different monomers in the context of an intact oligomer. This equalized the opportunities for disulfide bonds to be identified in the two different kinds of assemblies, irrespective of the size of the individual protein subunits. If the distance was less than 8 Å for any pair of mapped cysteine residues, that sequence was identified as containing a potential disulfide bond.

We performed this analysis separately for the single and tandem BMC domain proteins. We found that only 5% of single BMC domain proteins satisfied this approximate test for a potential disulfide bond, whereas 31% of the tandem BMC domain proteins were positive under the same test (Figure 2.13b). The simple homology modeling-based criterion used here to identify candidate disulfide bonds is only an approximate indicator for the existence of a true disulfide bond in any one protein sequence, but the overall results are nonetheless decisive when applied over large sets of sequences. Here the results show clearly that disulfide bonding is relatively common in (trimeric) tandem domain BMC proteins but not in (hexameric) single domain BMC proteins.

We examined the subset of sequences predicted to have potential S-S bonds. Overall, potential disulfide bonds appeared to be common in those tandem domain shell proteins that matched most closely to the EutL and PduT templates – specific shell proteins from two different MCPs. Approximately 80% (44/55) of the sequences within the group of EutL-like proteins were identified as having potential disulfide bonds, whereas 39% (22/57) of those within the PduT-like group were likewise identified (Figure 2.13c). In contrast, disulfide bonds were predicted to be relatively rare in tandem domain shell proteins whose sequences more closely matched other templates, including: PduB (Pdu MCP), CcmO (β-carboxysome) and CsoS1D (α-carboxysome).

2.4 Results – EutL binds to ethanolamine at a site that is incompatible with molecular transport

The closed conformation of EutL is a structural arrangement that leaves three small openings perforating the pseudohexameric homotrimer, one in each subunit. While the possible movement of small molecules across these small openings has been discussed[34,38,58], the opening and closing of the large central pore remains the most striking structural feature of EutL[27,34]. We characterized the interaction of the substrate molecule, ethanolamine, with these sites and evaluated the possibility of molecular transport through these channels. Our results suggest that EutL is unlikely to transport ethanolamine, but instead ethanolamine binding to EutL results in negative allosteric regulation of pore opening.

2.4.1 Calorimetric investigation of ligand binding to cpEutL

As a first step toward defining the function of the EutL shell protein, we used isothermal titration calorimetry (ITC) to determine if cpEutL binds to any of the small molecules involved in the ethanolamine metabolic pathway. We tested the binding of four small molecules, including ethanolamine, ethanol, acetate, and acetyl-phosphate. (Acetaldehyde was not included in the analysis because it reacts with our buffered protein solution, producing a large amount of heat that obscures any potential ligand-binding signal present in the experiment.) Interestingly, all of the interactions of these ligands with cpEutL appear to release heat; however, our data for the ethanolamine titration are notably different from the data for the other titrations (Figure 2.14). Specifically, plots of ΔH vs. molar ratio of ligand to protein produce virtually flat lines for all of the ligands except ethanolamine, which produces a logarithmic curve indicative of a specific protein-ligand interaction.

Our calorimetric data from the ethanolamine titration allowed us to create a binding model that describes two ethanolamine binding sites per monomer with very different affinities and specificities for their ligand (Figure 2.15). We fit the binding data to a sequential, two-site binding model to derive thermodynamic parameters for the interaction (Table 2.2). The quality of the thermodynamic binding model described by the parameters in Table 2.2 was evaluated using a chi-square analysis, which revealed an excellent fit of the model to the data. The model describes a high-affinity site, where binding is only

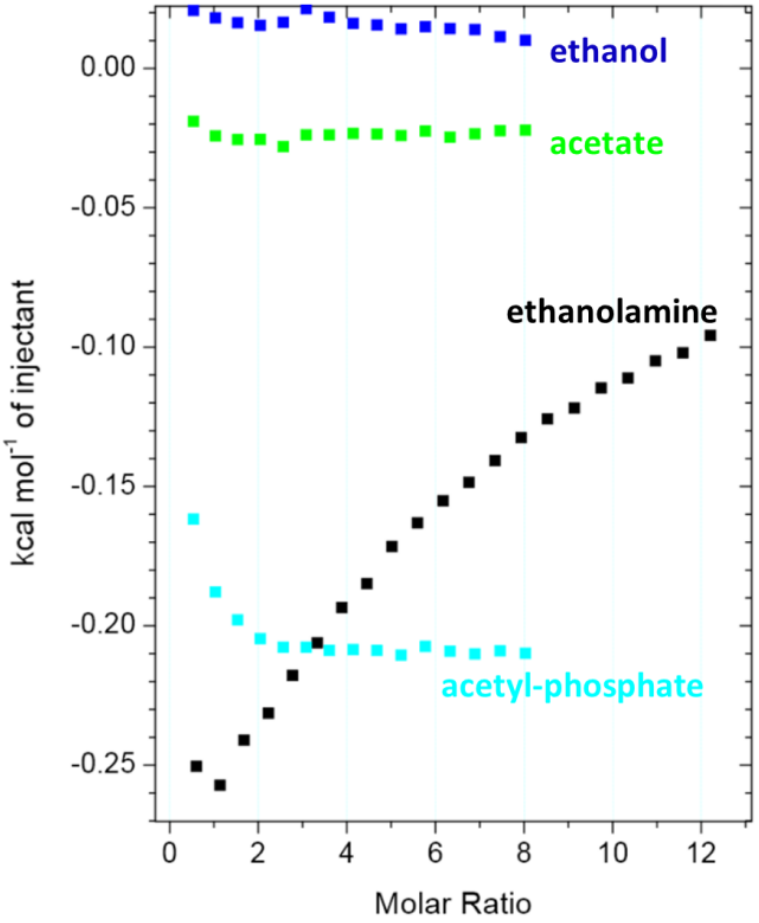


Figure 2.14. EutL binds specifically and selectively to ethanolamine. A comparison of integrated ITC data for titrations of several small molecules into a cpEutL solution reveals that ethanolamine is the only titrant whose interaction with cpEutL releases heat in a manner consistent with a specific binding event. Additionally, cpEutL seems to be quite selective for ethanolamine over similar small ligands associated with the Eut MCP.

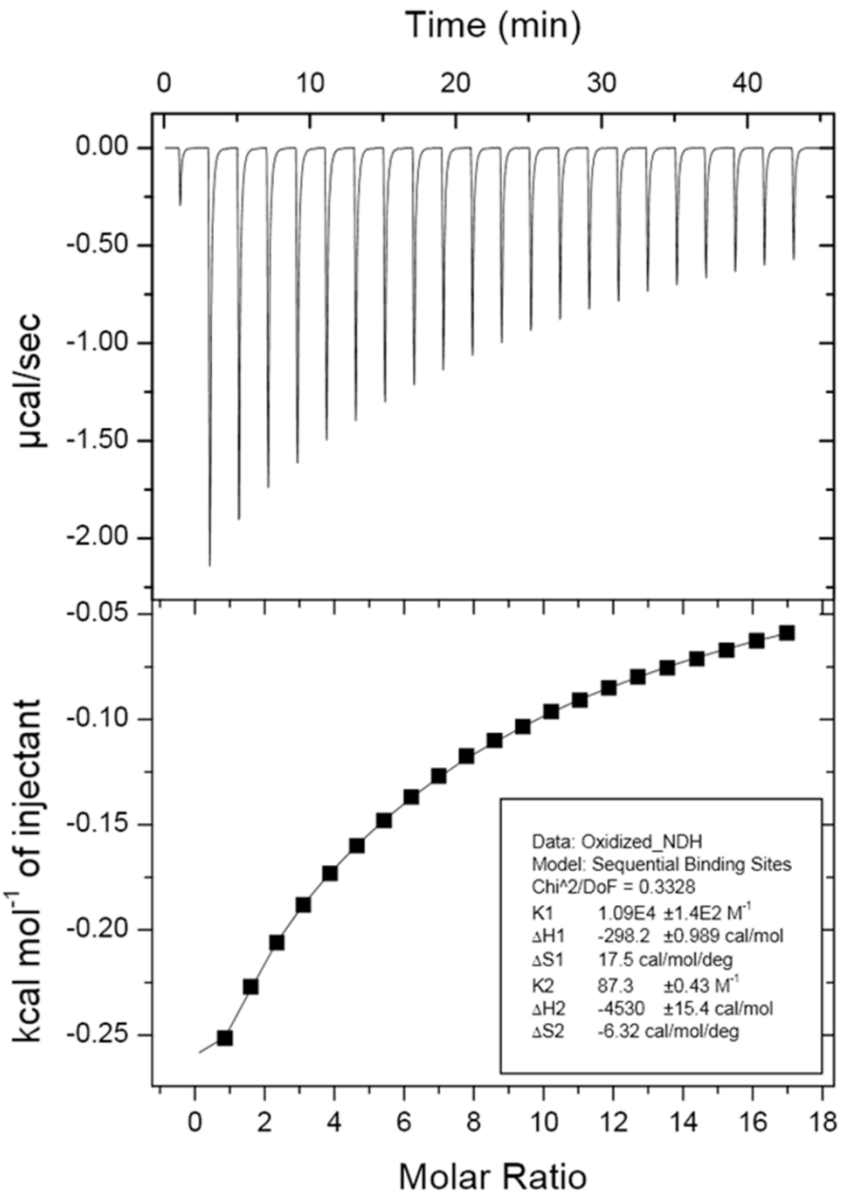


Figure 2.15. Isothermal titration calorimetry data. Images depicting the raw (top panel) and integrated (bottom panel) ITC data measured for a titration of ethanolamine into cpEutL. The data shown were used to derive thermodynamic parameters for the ethanolamine-cpEutL interaction using a sequential, two-site binding model.

	Oxidized		Reduced	
	Site 1	**Site 2**	**Site1**	**Site2**
K_d (μM)	91.7 ± 1.1	11,500 ± 100	321 ± 3	97,000 ± 700
ΔH (cal mol^{-1})	-298.2 ± 0.989	-4530 ± 15.4	-247.9 ± 1.26	-25,550 ± 174
ΔS (cal mol^{-1} K^{-1})	17.5	-6.32	15.2	-81.1
χ^2/DoF	0.3328		0.2742	

Table 2.2. Thermodynamic parameters for cpEutL binding to ethanolamine. Parameters calculated from isothermal titration calorimetry experiments, carried out at 298K. The data were fit to a two-site, sequential binding model.

slightly favored by enthalpy but is accompanied by a significant increase in entropy, and a low-affinity site, where binding is highly exothermic, but comes at an entropic cost.

We also performed ITC experiments using protein that had been chemically reduced with tris(2-carboxyethyl)phosphine. We found that reduction reduced the affinity of EutL for ethanolamine nearly fourfold at the high-affinity binding site and nearly eightfold at the low-affinity binding site (Table 2.2). Additionally, the thermodynamic parameters extracted from the data differ significantly between the oxidized and reduced states.

2.4.2 X-ray crystal structure of ethanolamine-bound cpEutL

We determined the X-ray crystal structure of cpEutL with ethanolamine bound, in order to elucidate the structural determinants of ligand binding. Using the hanging drop vapor-diffusion method, we grew cpEutL crystals belonging to space group $P4_32_12$, as described in section 2.2. Crystals were soaked in mother liquor supplemented with ethanolamine, cryoprotected, and frozen at 100K. We then collected diffraction data to a resolution of 1.7Å. Data collection and refinement statistics for the ethanolamine-bound structure are provided in Table 2.1.

In the ethanolamine-bound cpEutL structure, two ligand molecules occupy each of the three narrow channels that perforate the trimeric assembly. The presence of two ethanolamine molecules per monomer is consistent with the binding model that was fit to the ITC data. The first ethanolamine molecule is bound on the cytosolic side of the constriction point of the channel, while the second ethanolamine lies within the channel on the luminal side of the constriction (Figure 2.16a). This first ethanolamine molecule packs against Phe112 and Phe184, positioned so that its protonated amine group forms ionic interactions with acidic residues Asp44 and Glu82 and its hydroxyl group also makes a hydrogen bond to Glu82 (Figure 2.16b). The second ethanolamine molecule interacts less intimately with the protein, packing against Phe176, while forming an ion pair with Asp45, a hydrogen bond with Thr180, and several water-mediated contacts to other residues (Figure 2.16b). The ethanolamine binding channel is highly conserved, and its acidic nature is illustrated by calculation of electrostatic surface

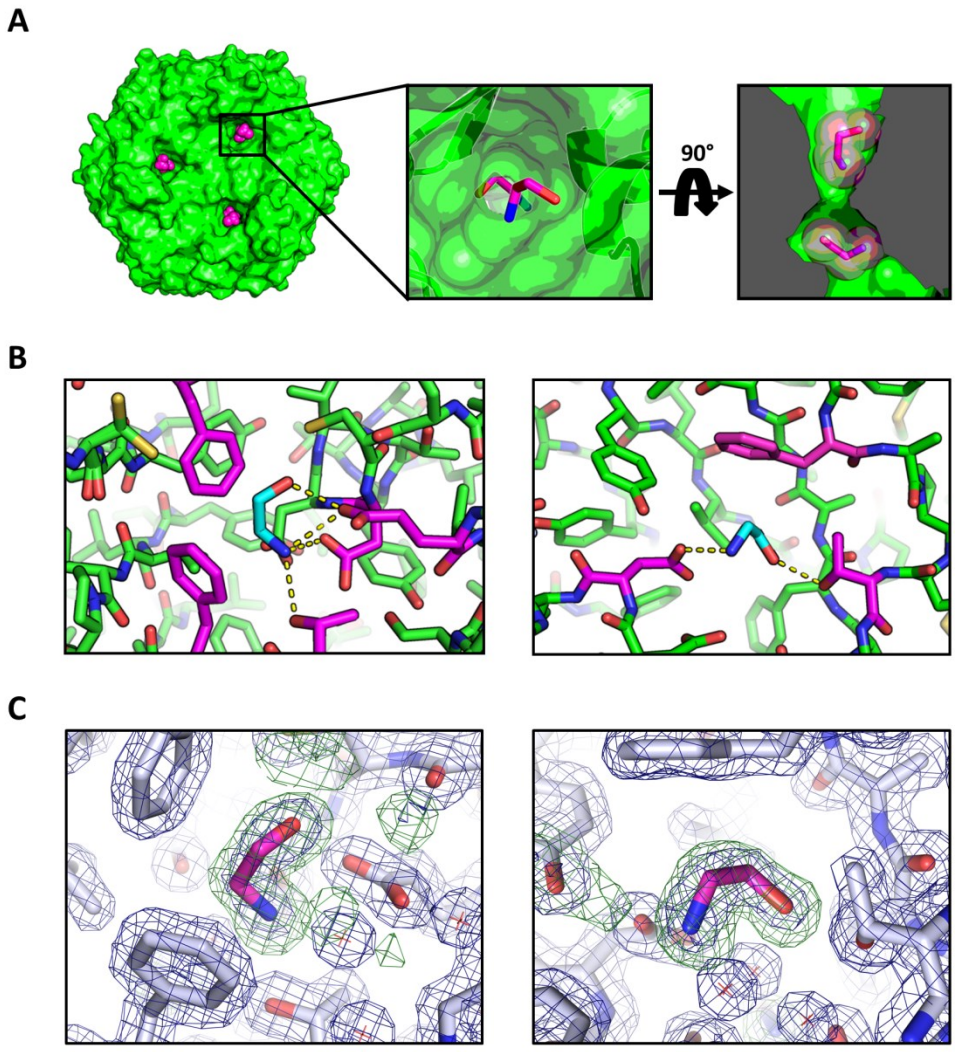


Figure 2.16. Ethanolamine binds to the narrow channels that perforate cpEutL. A) X-ray crystal structures reveal that ethanolamine molecules (magenta in panel A) bind to EutL in the narrow channels that perforate the trimmers. Each monomer contains a single channel with two bound ethanolamine molecules. Within each channel, one ethanolamine molecule is bound on either side of the narrow, hourglass-shaped constriction point. B) Conserved amino acids (magenta in panel B) form hydrophobic and polar interactions with ethanolamine molecules (cyan in panel B) that occupy each of the two binding sites. C) Electron density (2mFo-DFc, blue; mFo-DFc omit, green) supports the placement of the ethanolamine ligands (magenta in panel C) in the crystallographic model.

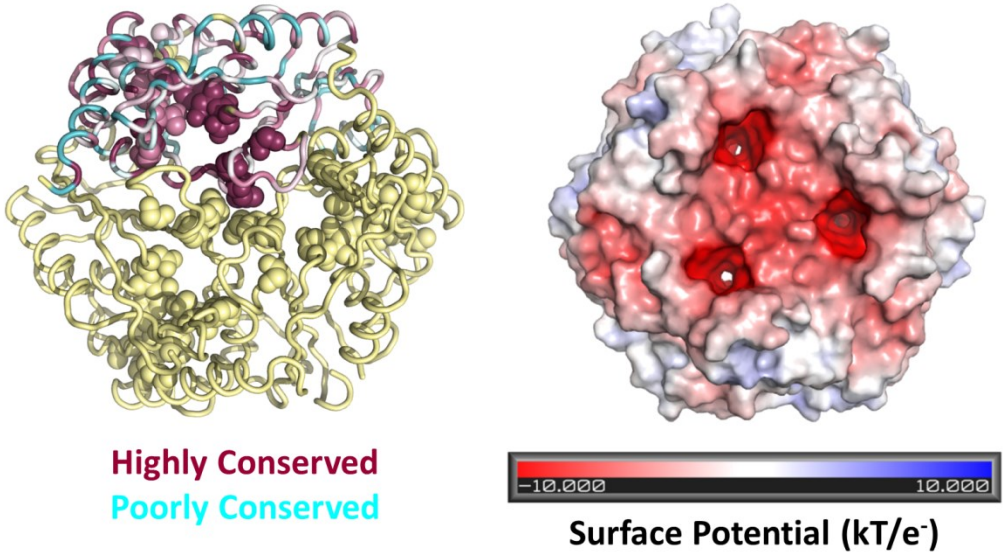


Figure 2.17. Molecular properties of cpEutL. Two images of the cpEutL moleceutule depicting sequence conservation and electrostatic surface potential. The image on the left shows a cpEutL trimer, with one chain colored by sequence conservation, with dark magenta indicating highly conserved positions and dark teal indicating highly variable positions. Some functionally important residues, including those surrounding the ethanolamine binding sites, are shown as spheres. The image on the right shows the surface of cpEutL colored according to electrostatic surface potential, which was calculated by solving the Poisson-Boltzmann equation. The ethanolamine binding sites carry a strong negative charge.

potential (Figure 2.17). The negative charge of the acidic residues within the channel complements the positive charge on the primary amine group of the protonated ethanolamine molecule.

Placement of the bound ethanolamine molecules into the model was justified by strong features in both *2mFo-DFc* and mF_o-DF_c electron density maps. After modeling the bound ethanolamine and performing several cycles of atomic refinement, we generated an omit map using calculated structure factors and phases that did not include contributions from the ligand atoms. The resulting map showed very strong density corresponding to the omitted ligand molecules (average peak height of 13.7σ in mFo-DFc map), verifying their inclusion in the model (Figure 2.16c).

2.4.3 Crystallographic examination of binding site dynamics

In order to explore the possibility that protein dynamics allow for expansion of the ligand-binding cavities in EutL allowing ethanolamine molecules to traverse the MCP shell, we performed several specialized crystallographic analyses. First, we determined the X-ray structure of cpEutL at room temperature, rather than under cryogenic conditions. Next we used both cryogenic and room-temperature X-ray diffraction data to perform time-averaged refinement of cpEutL structural ensembles.

Motivated by the idea that cryocooling can remodel loosely packed regions of protein structures, we determined the structure of cpEutL using X-ray diffraction data collected under ambient temperature conditions (approximately 298K). For the room temperature diffraction studies, we used the tetragonal crystal form of cpEutL. Interestingly, by comparing the X-ray data from the room temperature and cryogenic studies, we found that cryocooling shrinks the tetragonal unit cell by 1.1% along the c-axis, and by 2.5% along the a- and b-axes. This unit cell shrinkage is in line with previous reports[96,97]. Also, while the final structures are nearly identical, as were the refinement protocols used to obtain them, the room temperature structure gave considerably better refinement statistics, indicating the room temperature model fits the respective diffraction data better than the cryogenic model (Table 2.1). The overall similarity of the room temperature and cryogenic structures, especially at the ethanolamine binding sites,

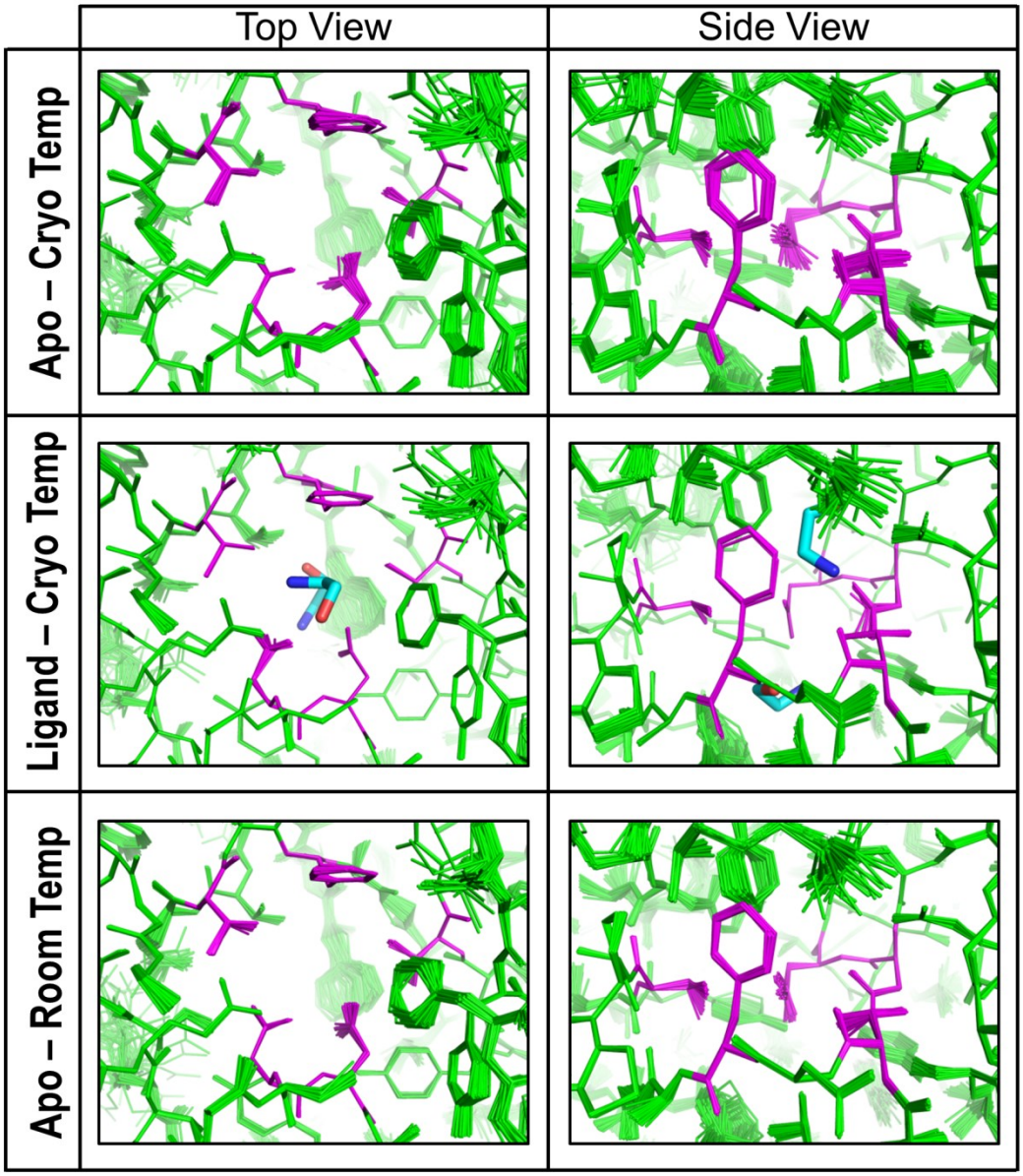


Figure 2.18. Time-averaged refinement of cpEutL structural ensembles. The images depict the ethanolamine binding channels in cpEutL as modeled using time-averaged ensemble refinement against cryogenic cpEutL diffraction data collected from both apo and ethanolamine-bound (top and middle panels respectively) crystals, and also against cpEutL diffraction data collected from apo crystals at room temperature (bottom panel). The residues that form the constriction point of the hourglass-shaped channel are shown in magenta. The ensembles show that the channels do not undergo any dynamic expansion that can account for substrate transport.

proves that the geometry of these sites is not affected by cryocooling in a way that would complicate interpretation of ligand binding.

Next, we looked for evidence of binding site dynamics using time-averaged refinement of cpEutL structural ensembles. This method uses molecular dynamics simulation to create a time-averaged ensemble of structures that provides the best fit to the observed X-ray data[98]. Starting from our cryogenic X-ray crystal structures, we created ensemble models for both the apo and ethanolamine-bound states. The ensemble model of the apo state consisted of 50 individual structures, and fit the diffraction data very well, with R/R_{free}=0.1383/0.1680. Similarly, the ensemble model of the ethanolamine-bound state consisted of 63 individual structures, with R/R_{free}=0.1350/0.1588. We also created an ensemble model using our room temperature X-ray data for cpEutL, which consisted of 50 individual structures, with R/R_{free}=0.1293/0.1571. Based on standard crystallographic R-factors, the ensemble models fit the diffraction data significantly better than the single-conformer models. In the ensemble models, the residues that form the narrow constriction points of the hourglass-shaped ligand binding channels show little structural fluctuation (Figure 2.18).

2.5 Results – Crystallographic studies of vitamin B_{12} binding to EutL

While it appears that the EutL pore opens to allow the passage of a large cofactor, the identity of this putative molecule is unknown. To develop a complete model of EutL function, we need to discover why the pore opens, or in other words, we need to identify the transport substrate of the open pore.

2.5.1 Crystallization of EutL bound to hydroxocobalamin

We were able to obtain crystals of EutL bound to hydroxocobalamin by cocrystalization and by crystal soaking. When EutL was cocrystalized with the ligand, broad screening of crystallization conditions gave many initial hits, virtually all of which demonstrated a flat, plate-like habit. These crystals had a distinct pink hue, characteristic of the hydroxocobalamin compound (Figure 2.19a). Exposure to X-rays showed that the plate-like crystals universally suffered from lattice translocation

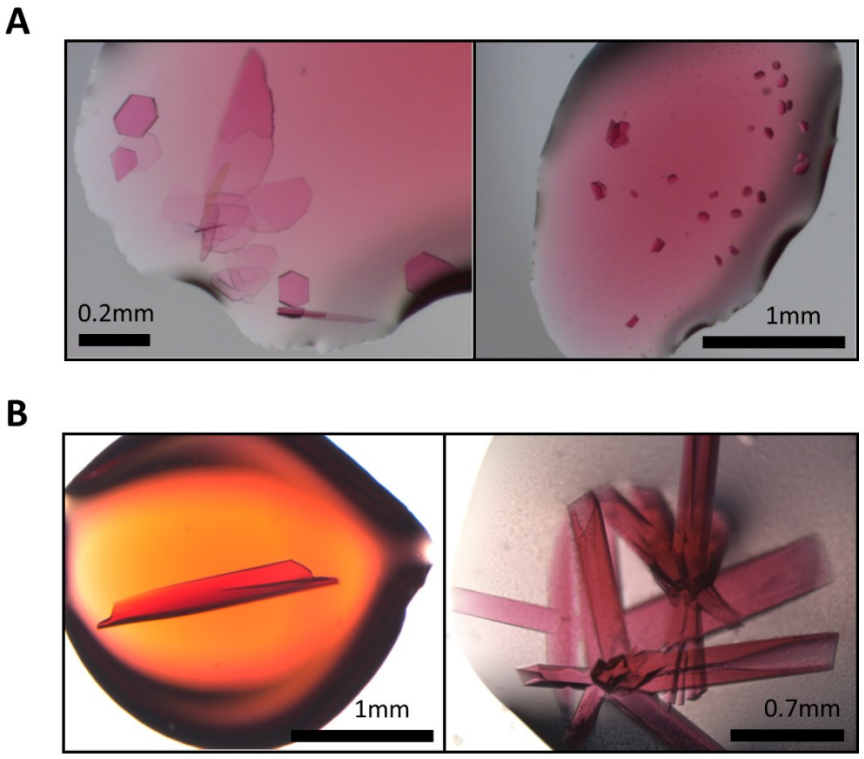


Figure 2.19. Crystals of cpEutL bound to vitamin B12. A) Examples of crystals that were obtained by co-crystallization of cpEutL with hydroxocobalamin. These crystas were hexagonal in nature, but suffered from severe lattice translocation disorders. B) Examples of crystals that were obtained by soaking tetragonal cpEutL crystals in solutions containing hydroxocobalamin. The image on the left shows a crystal soaked in mother liquor with 10mM hydroxocobalamin, while the image on the right shows crystals that were incubated in mother liquor that was supplemented with several small crystals of solid hydroxocobalamin.

disorders, which prevented structure determination. Fortunately, we could also obtain crystals of EutL bound to hydroxocobalamin by soaking our tetragonal crystal form into the compound. The soaking procedure resulted in dark, pink or red crystals (Figure 2.19b), which remained pink even after washing and back-soaking the crystals overnight in mother liquor without the ligand.

2.5.2 Preliminary co-crystal structure of EutL and hydroxocobalamin

The tetragonal cpEutL crystals soaked in hydroxocobalamin diffracted to a resolution of 2.1Å. Data collection and current refinement statistics for the hydroxocobalamin-bound structure are provided in Table 2.1. Electron density maps show clear, positive difference density features (16.9σ in mFo-DFc maps) that match the shape of the corrin ring of the hydroxocobalamin molecule (Figure 2.20a). Placement of hydroxocobalamin into the density positions the imidazole ring of His32 as an axial ligand to the cobalt atom of the hydroxocobalamin (Figure 2.20b). In this binding orientation, the ligand makes little contact with the protein aside from the His-Co interaction (Figure 2.20c), presumably giving the ligand some degree of rotational freedom, which may explain the fact that only the corrin ring is adequately resolved in the electron density.

2.6 Discussion

Studies of EutL have been critical to our understanding of molecular transport across the protein shells of MCP structures. Previous structures of *E. coli* EutL provided early evidence that the broken symmetry resulting from domain duplication allows for conformational change at the center of the trimer, which permits the opening of a large central pore[27,34]. While the importance of this conformational change in facilitating molecular transport seems clear, the details of how it takes place have yet to be fully elucidated. The crystallization conditions that yielded closed versus open structures are uninformative with respect to understanding the biologically relevant mechanism of the conformational transition. Takenoya, et al. have shown that the open conformation can be induced by very high concentrations (1M)

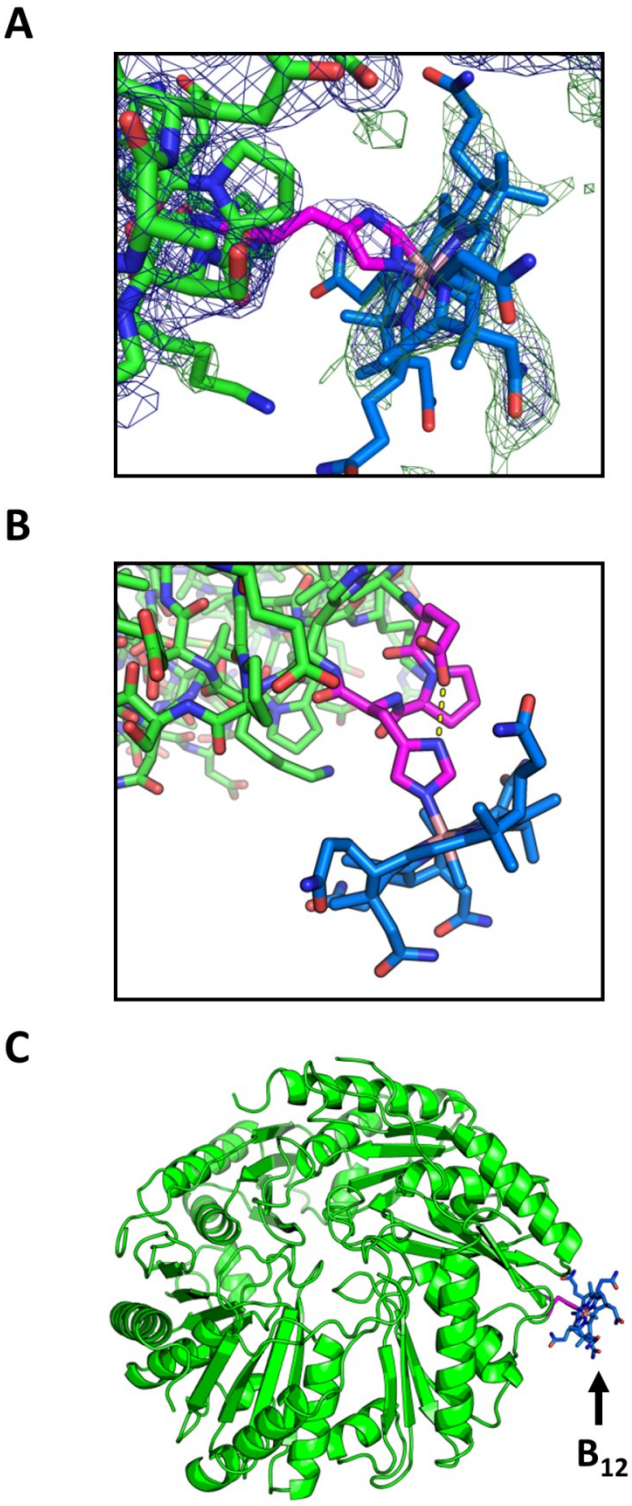


Figure 2.20. Preliminary crystallographic analysis of vitamin B_{12} binding to cpEutL. A) A large planar feature in preliminary electron density maps ($2mF_o$-DF_c, blue; mF_o-DF_c, green) is consistent with the corrin ring of hydroxocobalamin, and its position supports ligation of the central Co atom to a histidine residue (H32) in cpEutL. B) The binding of hydroxocobalamin to cpEutL occurs by coordination of H32 to the Co atom of the ligand. This ligation motif (magenta), wherein a histidine coordinates vitamin B_{12} while being stabilized by a nearby acidic residue, is known in other cobalamin-binding proteins. C) The vitamin B_{12} binding site is located at the edge of the cpEutL trimer on the luminal face of the oligomer, and in a position that does not interfere with the interactions required for formation of the MCP shell.

of zinc ions, though they are candid regarding their doubts about whether a high zinc concentration is likely to be physiologically relevant for microcompartment function[34]. The experiments presented here highlight features of EutL that appear to be important in allowing the structural rearrangement to take place, and suggest a biologically plausible mechanism for regulating the pore opening.

2.6.1 Key structural seatures of EutL facilitate conformational change

Our crystal structures of EutL form *Clostridium perfringens* (cpEutL) are all strikingly similar to structures of the E.coli homologs studied previously, occupying conformations that are almost identical to what is seen for the closed form of the *E. coli* protein. Critical interactions that stabilize the closed conformation are consistent between cpEutL and E. coli EutL. This includes important hydrogen-bond interactions within the β3-β4 and β7-β8 loop regions, and interactions between the aromatic rings at the center of the trimer (Figure 2.4). Analysis of sequence conservation using the ConSurf algorithm[99,100] shows that residues involved in these interactions (Y69, N74, N183) are strongly conserved among EutL orthologs.

The closed cpEutL structure also reveals features that appear important for enabling an alternate open conformation, based on comparison with the E. coli structure. Each of the cpEutL monomers contains a narrow, elongated opening, which would accommodate movement of the β3-β4 loop during opening of the central pore (Figure 2.4). This opening is lined with highly conserved acidic residues (Figure 2.17). The corresponding aspartate residues in E. coli EutL make hydrogen bonds with the β3-β4 loop in the open conformation. Finally, the presence of highly conserved glycine residues (G81, G186) at the ends of the β3-β4 and β7-β8 loops might be explained by a need for flexible movement of those segments.

Although we observe primarily one major conformation of cpEutL in our crystal structures, in some electron density maps there is strong evidence for an alternate “open” conformation. We demonstrated that modeling a small fraction of cpEutL trimers in the presumptive open conformation

could marginally reduce crystallographic R-factors (Figure 2.8c), providing possible evidence that cpEutL is capable of undergoing the same conformational change as the E. coli protein.

2.6.2 X-ray crystallography reveals disulfide bonding in EutL

Our crystallographic electron density shows, for the first time, that EutL is able to form disulfide bonds. In our untreated (tetragonal) and oxidized crystal structures of cpEutL, there is strong residual electron density corresponding to a partially-occupied Cys127-Cys200 disulfide bond (Figure 2.7). Positive electron density in $F_{oxidized}$-$F_{reduced}$ difference maps provides additional support for that bond as well as for a partially occupied, alternate disulfide bond between residues 127 and 196 (Figure 2.10). While the electron density features are not as clear as expected for fully formed disulfide bonds, we note that disulfide bonds are generally susceptible to radiation damage by X-rays used for structure determination[90–93]. The Cys127-Cys200 disulfide might be especially prone to photoreduction because its dihedral angle, at approximately 124°, is outside the ideal range of 90°±12° (the full range of S-S dihedrals in the PDB is 27°-153°)[101]. Calculation of difference electron density maps comparing wedges of X-ray data from early and late in the course of data collection (F_{early}-F_{late}) show positive features that confirm the Cys127-Cys200 disulfide bond is broken by X-ray irradiation (Figure 2.7).

2.6.3 Attempts to Confirm the Functional Relevance of Disulfide Bonding in EutL

To confirm the presence of disulfide bonds observed in the crystal structures, we used a combination of mass spectrometry (SEC-MS) and two-dimensional gel electrophoresis. Mass spectrometry experiments with non-reduced and reduced cpEutL show that the protein exists as multiple species, which likely correspond to polypeptides with different numbers of intramolecular disulfide bonds (Figure 2.11b). A tryptic digestion and two-dimensional SDS-PAGE experiment further confirmed the presence of disulfide bonds (Figure 2.11a).

Motivated by the idea that disulfide bonds generally have a structural or functional significance[102], we sought to determine if their presence in cpEutL might be coupled to the functionally

relevant conformational change that opens the central pore. NMR experiments (^{1}H-^{15}N TROSY-HSQC) showed notable differences between the oxidized and reduced spectra, indicating structural differences between the two states (Figure 2.12). In the absence of specific NMR peak assignments, at the present time we cannot confirm how closely the underlying conformational changes detected for cpEutL correspond to those observed previously for E. coli EutL. In light of current structural data for EutL, it is possible that the structural change detected involves the pore region, although further experimentation will be required to confirm this hypothesis.

Bioinformatic evidence also points to a functional significance of disulfide bonding in EutL. Sequence alignment of EutL orthologs reveals that the cysteine triad containing Cys127, Cys196, and Cys200 is highly conserved. Additionally, a computational survey of disulfide bonding in BMC-domain shell proteins revealed that EutL orthologs and PduT orthologs are enriched for potential disulfide bonds (Figure 2.13). BMC-domain proteins are a highly divergent family, including topological variations[29,103]. It is notable that the two types of tandem domain shell proteins in which disulfide bonds appear to be common have different topological arrangements and likely play different roles within functionally distinct MCPs[29,35]. The presence of disulfide bonds within specific, divergent subtypes of MCP shell proteins suggests that they could be functionally relevant, though a more complete understanding of their pattern of distribution and their possible roles awaits further study.

2.6.4 EutL binds ethanolamine, the substrate of the Eut MCP

Recent work by Pang, et al. suggested that a EutL homolog, PduB, interacts with the substrate molecule of the MCP to which it belongs[39], prompting us to investigate the interaction of EutL with ethanolamine. Using isothermal titration calorimetry, we demonstrated that ethanolamine binds specifically to EutL, while other small molecules associated with the Eut MCP do not. Titrations of ethanol, acetate, and acetyl-phosphate into EutL release heat in a manner consistent with nonspecific interactions, whereas titration of ethanolamine into EutL produces a binding curve with exponential character, indicative of a specific binding event (Figure 2.14). The data for the titration of ethanolamine

into EutL could be fit to a binding model that includes two binding sites per EutL monomer, with significantly different dissociation constants (Figure 2.15). The dissociation constants calculated for the two binding sites are 91.7μM and 11.4mM. The concentration of ethanolamine in the human gut is in the range of 0.5-1mM[79], which means that only one of these two sites is likely to be physiologically relevant. The dissociation constant for physiological binding is consistent with EutL being bound to ethanolamine when bacteria inhabit the gut of their human host.

In addition to determining the thermodynamics of ethanolamine binding to cpEutL, we used X-ray crystallography to visualize, at relatively high resolution, the ligand molecules bound to the protein. The ethanolamine molecules bind in the narrow channels that perforate each monomer within the closed conformation of the cpEutL trimer. Two ligand molecules bind in each channel, one on either side of the narrow constriction point. The ethanolamine molecule bound on the cytosolic side of the channel constriction point makes numerous direct interactions with amino acid side chains that line the channel, including ion pairing and hydrogen-bonding of the protonated amine with conserved acidic residues (D44 and E82), and van der Waals contacts with conserved aromatic residues (F112, F184). In contrast, the ethanolamine molecule bound on the luminal side of the channel appears to interact much less intimately with the protein molecule, forming only a few direct interactions (a hydrogen bond to T180, an ionic interaction with D45, and van der Waals interactions with Phe 176) and several indirect, water-mediated interactions.

The presence of two ethanolamine molecules bound to each monomer is consistent with the binding model derived from ITC experiments, however it is difficult to deduce from the crystal structure which binding site is the physiologically relevant, high-affinity site and which is the low-affinity site. Based on the observed binding interactions, it is tempting to speculate that the ethanolamine molecule bound to the cytosolic side of the channel represents the higher-affinity site, because it forms more direct interactions with the conserved residues of the binding site, and as a result, appears to be in a more stable orientation. Although the ITC experiments demonstrate that the higher-affinity binding site appears to have a relatively small enthalpic contribution and a larger entropic contribution, this apparently rigid

binding orientation could be rationalized by considering the potential expulsion of water molecules from the channel that would accompany ethanolamine binding. Unfortunately, this hypothesis is currently difficult to assess, because electron density maps for the apo form of the protein are somewhat noisy in regions within the ligand binding channels, making it difficult to accurately model the ordered water molecules that occupy these sites. Indeed, further experimentation will be required to confidently assign the measured thermodynamic parameters to the observed binding sites.

Interestingly, we observed that changing the electrochemical environment of the cpEutL protein affected its affinity for ethanolamine. Upon reduction, both binding sites decrease their affinity for ligand. Because the ethanolamine binding channels are only present in the closed form of EutL, the observed change of the binding constant serves as the best evidence yet that there might be a coupling between electrochemistry and the conformational equilibrium of the EutL pore, possibly related to the observed disulfide bond. The high-affinity binding site, which is physiologically relevant, shows a nearly fourfold increase in the dissociation constant, from 91.7μM to 321μM, following chemical reduction. This decrease in binding could be explained by a shift in the conformational equilibrium toward the open conformation, which cannot bind the ligand. Because the concentration of ethanolamine in the human gut is 0.5-1mM, it is not clear whether this fourfold increase in the dissociation constant is functionally significant. It is worth noting, however, that there is a link between ethanolamine metabolism by enteric bacteria and oxidative stress induced by the host immune response, which could provide a physiological basis for electrochemical regulation of MCP function[78,79].

2.6.5 Ethanolamine binding channels do not conduct substrate transport

Previously, there has been discussion about the possibility of substrate transport through the narrow channels that form the ethanolamine binding sites in EutL and other homologous tandem BMC-domain proteins[34,38,39,58]. Based on our analyses, we believe these channels are too narrow to support the passage of substrates. Our geometric measurements reveal that the narrow constriction point of the hourglass-shaped channel has a radius of only 1.2Å (Figure 2.6). For comparison, the van der Waals

radius of a carbon atom is 1.7Å. Considering that ethanolamine has two sp^3-hybridized carbon atoms, we think it is highly unlikely that this substrate molecule can diffuse through such a narrow pore. Additionally, the shell of the Eut MCP also contains a homohexameric shell protein, EutM, whose central pore has a radius of roughly 2-3Å[27]. The energetic barrier for ethanolamine crossing the EutM pore is likely to be so much smaller than for ethanolamine passing through the narrow EutL channel, that any potential flux of ethanolamine through EutL becomes irrelevant in the presence of EutM.

While the static X-ray crystal structures of EutL showed narrow ethanolamine binding channels, we recognized the possibility that protein dynamics might result in expansion of the ligand-binding channels. In order to investigate this idea, we used a recently developed method known as time-averaged ensemble refinement, which allowed us to model anisotropic and anharmonic molecular fluctuations consistent with our measured X-ray diffraction data[98]. The structural ensemble corresponding to the apo form of the protein reveals very little dynamic motion among the residues corresponding to the channel constriction point (D44, D45, V151, T182, F184). We also generated a structural ensemble using X-ray data from ethanolamine-soaked crystals, which is virtually identical to the ensemble calculated for the apo protein, arguing against the hypothesis that ligand-binding induces protein dynamics that open the channel. The results of the ensemble refinements, taken together, imply that protein dynamics cannot account for the significant opening of the ethanolamine binding channel that would be required for substrate transport (Figure 2.18).

Moving one step further, we also wanted to explore the possibility that the geometry and dynamics of the ligand binding channels are affected by cryocooling. It is known that cryocooling protein crystals to temperatures below the glass transition can collapse voids in protein structures that result from imperfect atomic packing, and also that it can remodel conformational distributions of the protein molecules in the crystal[96,97,104–106]. Because the ethanolamine binding sites in cpEutL are narrow and hollow, we reasoned they might be subject to some of these negative effects of cryocooling. We collected X-ray diffraction data at room temperature and used that data for refinement of an atomic model, which turned out to be essentially identical to the cryogenic structure with respect to the geometry of the

ethanolamine binding sites, indicating the binding sites are not distorted by cryocooling even though the unit cell shrinks substantially. Also, time-averaged ensemble refinement against the room temperature data produced an ensemble that was highly similar to the ensemble models generated from cryogenic data, confirming that temperature does not have a significant effect on the dynamics observed in the crystal (Figure 2.18).

2.6.6 Ethanolamine is a negative allosteric regulator of EutL pore opening

Rather than acting as channels for substrate transport, we propose instead that the ethanolamine binding sites are allosteric sites for negative regulation of pore opening. Crystal structures of E. coli EutL determined by Tanaka, et al.[27] demonstrated that the three channels through the EutL pseudohexamer are crucial for the conformational change of the pore, because the β3-β4 loops retract into this empty space as the central pore opens (Figure 2.5). By filling this empty space with atoms, the presence of ethanolamine molecules in these channels prevents the movement of the loop segments that must rearrange to open the pore. The steric clash that prevents pore opening in the presence of bound ethanolamine is clearly illustrated by superimposing a model of the open conformation of EutL on a model of the ethanolamine-bound structure (Figure 2.21). The incompatibility of pore opening and ethanolamine binding strongly suggests that this small molecule serves as a negative allosteric regulator of the conformational rearrangement. Additionally, although there is some ambiguity about which ethanolamine binding site is physiologically relevant, we note that both molecules bound in the crystal structure produce a steric clash with the open conformation, consistent with the allosteric model. This model of allosteric regulation (Figure 2.22) does not provide an explanation for what causes the pore to open, but it does provide indication of how it can become locked closed.

2.6.7 Vitamin B_{12} is a potential transport substrate of EutL

A major shortcoming in our knowledge about EutL is that we are unaware of the identity of the molecule that passes through its pore. The chemistry of the Eut MCP suggests that is it either ATP or

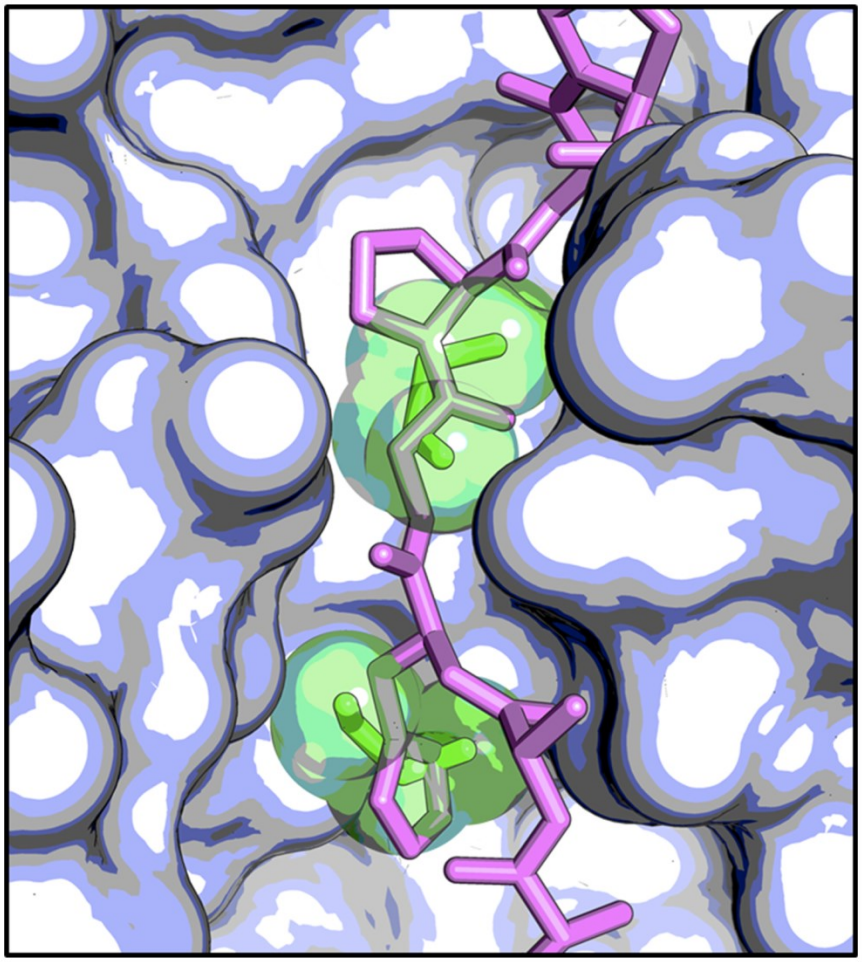


Figure 2.21. A steric clash forms the basis for allosteric regulation. The image depicts the ethanolamine binding channel as a molecular surface, with the ethanolamine molecules shown as green sticks surrounded by transparent spheres. The pink polypeptide shows the open conformation of E. coli EutL superimposed on the cpEutL:ethanolamine complex. The overlay reveals that ethanolamine binding prevents a rearrangement from the closed conformation to the open conformation by blocking the space into which the β3-β4 loop moves during the transition.

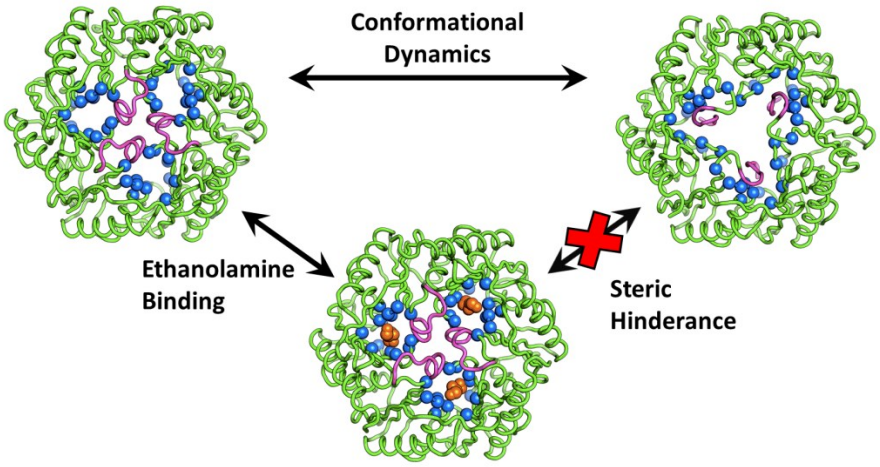


Figure 2.22. Negative allosteric regulation of EutL pore opening by ethanolamine. Ethanolamine binds to the closed conformation of EutL resulting in steric hiderance that prevents the transition to the open conformation.

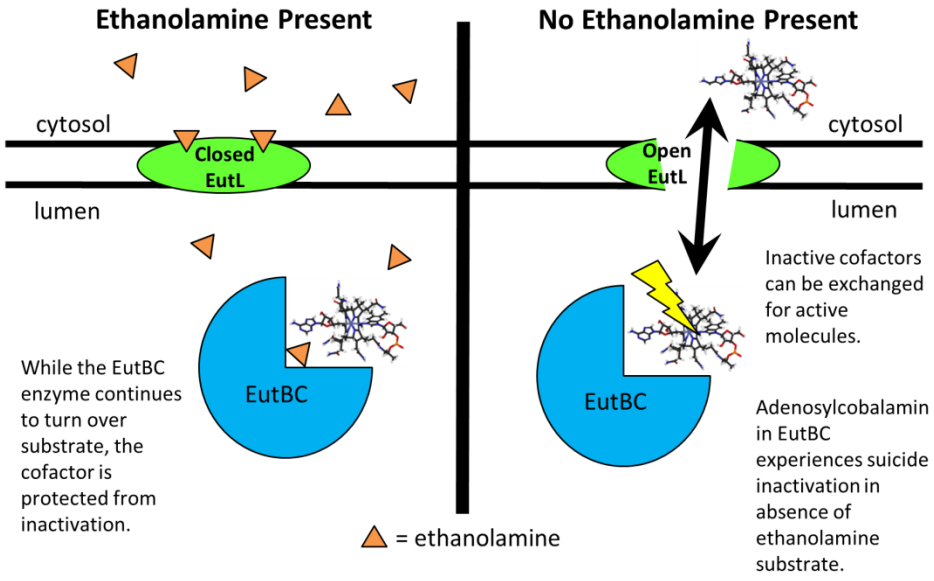


Figure 2.23. A new model for EutL function. The proposed allosteric mechanism is consistent with the enzymology of the Eut MCP. In the absence of the ethanolamine substrate, EutBC becomes inactivated and large cofactors (possibly cobalamins) must cross the shell in order to regenerate enzymatic activity.

some form of cobalamin, but uncertainties about the subcellular localization of some Eut enzymes leaves ambiguities about which molecular species has to cross the shell. A preliminary co-crystal structure of cpEutL bound to hydroxocobalamin, suggests that the cobalamin compounds might bind to EutL. In our structure, hydroxocobalamin is ligated to His32 through a His-Co coordination bond, positioned so that it protrudes toward the liminal face of the EutL pseudohexamer. The cobalamin ligation motif observed in our structure, involving a histidine and an acidic residue (Glu30 in this case), has been seen before in other cobalamin-bound protein structures[107]. Typically, the imidazole group of a histidine side chain displaces the axial dimethylbenzimidazole ligand to the Co atom, resulting in the coordination of the cobalamin to the protein.

While many features of the EutL-hydroxocobalamin co-crystal structure are consistent with some expected features of protein-cobalamin interactions, there are also some puzzling aspects of this observation. Notably, the apparent ligation motif (sequential amino acids EPH) that seems to be important for cobalamin binding is not conserved in EutL homologs, which is not what would normally be expected for a functionally relevant amino acid sequence. Also, the position of the hydroxocobalamin molecule in the structure is somewhat odd; the only significant contacts between the ligand and the protein are the His-Co coordination bond. The ligand molecule is protruding from the luminal face of the pseudohexamer (Figure 2.20b). If cobalamin compounds were indeed the transport substrates of EutL, one might expect to see the ligand bound in the pore. It may be possible that the observed binding site is an intermediate binding position, used in the transfer of cobalamins to or from ethanolamine-ammonia lyase enzymes, however this idea is purely speculative. Indeed, a great deal more work remains to be done to provide a complete validation and characterization of the crystallographically determined cobalamin binding site in EutL.

2.7 Conclusions - A new model for EutL function

The work presented in Chapter 2 provides a detailed analysis of the structural features within EutL that are critical for the conformational change that defines its function. Specifically, we have identified a conserved disulfide bond in EutL, and have provided initial evidence that this bond might be related to the function of EutL, possibly by shifting the conformational equilibrium in response to electrochemical environment. Additionally, our analysis focused on three small channels that perforate the EutL trimer. These channels, which are essential for the conformational rearrangements that open the central pore, are also the binding sites for ethanolamine molecules. These channels appear too narrow and rigid to support substrate transport, but instead appear to be allosteric sites, where bound ethanolamine molecules produce a steric clash with the open conformation of the oligomer resulting in negative regulation of pore opening. Finally, preliminary structural results suggest that the large cofactor that passes through the open EutL pore may be a cobalamin compound.

Our results, which suggest negative allosteric regulation of EutL pore conformation by ethanolamine binding, also provide a new hypothetical model for EutL function within the context of the Eut MCP. In the absence of substrate, the ethanolamine-ammonia lyase enzyme (EutBC) is susceptible to inactivation due to damage of the adenosylcobalamin cofactor[108]. In order to reactivate the enzyme, the damaged cofactor needs to be replaced. While the details of this process are unclear, it is evident that in order for it to take place either ATP or cobalamin compounds must pass through the MCP shell to become available to the interior enzymes of the MCP[84,85,108]. Because of its capacity to open a large central pore, EutL appears to be the conduit for movement of these larger molecules across the shell. If the large cofactors are required to pass through the shell when EutBC is inactivated due to the absence of substrate, then EutL only needs to open at low ethanolamine concentrations. The finding that ethanolamine serves as a negative allosteric regulator of EutL pore opening is consistent with this idea about the cofactor requirements of the Eut MCP (Figure 2.23).

2.8 Materials & Methods

2.8.1 Cloning, expression, and protein purification

For preparation of our cpEutL construct, we amplified the eutL gene from *Clostridium perfringens* chromosomal DNA (ATCC) and ligated the amplicon into the multiple cloning site of the pET22b expression vector (Novagen). We used the NdeI and XhoI restriction sites, thus incorporating the non-cleavable Leu-Glu-His6 affinity tag at the C-terminus of the 217-residue native sequence. We also prepared, in a similar manner, a construct containing an N-terminal His_6-ENLYFQG sequence, which acts as a cleavable affinity tag. Dideoxy chain termination sequencing[109] confirmed the recombinant DNA sequences.

We expressed recombinant protein using transformed *Escherichia coli* BL21 (DE3) Rosetta cells (Novagen). During the exponential phase of cell growth in selective Luria-Bertani (LB) Broth, we added 1mM isopropyl-β-D-thiogalactopyranoside to induce protein expression for four hours at 30°C. We collected the cells by centrifugation for 15min at 5,000x*g*.

Protein purification was identical for all cpEutL samples used for crystallization. Cells were lysed by high-pressure emulsification (EmulsiFlex C3 - Avestin). The lysis buffer contained 20mM Tris buffer and 300mM sodium chloride at pH 8.0 with a protease inhibitor additive (Sigma Aldrich), 10mM $MgCl_2$, 1mg/mL lysozyme, and 100units/mL of both DNase and RNase. We clarified the cell lysate by centrifugation at 30,000x*g* for 30min, and then we used a HisTrap nickel affinity column (GE Healthcare) to purify the proteins from clarified lysates, eluting the bound protein with lysis buffer containing 300mM imidazole. Finally, we dialyzed the eluent against a buffer containing 20mM Tris and 100mM sodium chloride at pH 8.0. This single purification step resulted in protein of sufficient purity for crystallization.

When working with the N-terminally tagged cpEutL, the following steps, in addition to the initial metal affinity purification, were taken to remove the affinity tag and produce highly pure protein samples. First, the peptide tag was removed by overnight cleavage with TEV protease (1mg/mL) while dialyzing the reaction mixture against buffer containing 20mM Tris pH 8.0, 100mM NaCl, and 1mM dithiothreitiol

(DTT). Following overnight protease treatment, the protein sample was once again passed over a nickel affinity column, except this time, we collected the untagged protein from the column flow through. The protein was then subjected to two additional purification steps, first using a HiTrapQ anion exchange column and then a Superdex200 gel filtration column (both from GE Healthcare). The final purification step left the protein in buffer containing 20mM Tris and 100mM sodium chloride at pH 8.0.

2.8.2 Crystallization

Prior to crystallization experiments, we concentrated the protein samples to approximately 20 mg/mL in 20mM Tris pH 8.0 and 100mM sodium chloride using Amicon Ultra concentrators (Millipore). We measured the protein concentration with the bicinchoninic acid assay[110]. Unless otherwise noted, crystallization experiments were performed using the C-terminally His-tagged cpEutL construct.

We carried out initial crystallization screening of the untreated protein by the vapor-diffusion method in 96-well hanging-drop format using a Mosquito robot and several commercial screening kits. We identified initial hits of the tetragonal crystal form in condition B4 of the JCSG+ screening kit (Qiagen) and initial hits of the monoclinic crystal form in condition E8 of the Wizard screen (Emerald Biosciences).

We optimized the initial crystals in 24-well hanging-drop format (VDX plates – Hampton Research). The highest quality crystals grew from 0.1M HEPES Buffer pH 7.0, 5% PEG-8000, and 8% ethylene glycol, or from 0.1M sodium/potassium phosphate buffer pH 6.6, 0.25M sodium chloride, and 10% (w/v) PEG-8000.

All of our modified structures were obtained using the tetragonal crystal form of cpEutL. We grew the reduced and oxidized crystals in the same fashion as the original tetragonal crystals, except in both cases we pre-treated the protein with either 5mM tris(2-carboxyethyl)phosphine (TCEP) or 10mM H_2O_2. We obtained crystals of the ethanolamine-bound form of cpEutL by first growing the tetragonal crystals and subsequently soaking them overnight in mother liquor containing 20mM ethanolamine. Crystals of B_{12}-bound cpEutL were likewise obtained by overnight soaking in mother liquor with 10mM

hydroxocobalamin. Alternatively, B_{12}-bound protein crystals could be obtained by using a fine needle to add a few small crystals of solid hydroxocobalamin ligand to crystallization drops with pre-formed crystals.

In addition to the B_{12}-bound crystals described above, we also grew some B_{12}-bound crystals that were not used for structure determination because they suffered from lattice translocation disorders. These crystals were obtained by cocrystallization of cpEutL, from which the N-terminal affinity tag was removed by protease treatment, with varying concentrations of hydroxocobalamin.

2.8.3 X-ray data collection and processing

Prior to X-ray data collection, we harvested the crystals and cryoprotected them using 50% mother liquor with 50% glycerol for the original (untreated, tetragonal and monoclinic) crystals, or 50% mother liquor with 2M trimethylamine-N-oxide (all other crystals, oxidized, reduced, ethanolamine, B_{12}). For room temperature data collection, crystals were not subjected to cryoprotectant, but were instead mounted inside of sealed polyester capillaries (Mitegen) along with a small plug of mother liquor to prevent crystal dehydration.

We collected single-crystal X-ray diffraction data at the Advanced Photon Source on beamline 24-ID-C. When we collected data from the original tetragonal and monoclinic crystals, this beamline was equipped with an ADSC Quantum315 CCD detector. Thanks to beamline upgrades, all other data sets were collected on a Pilatus 6M-F detector. We maintained the crystals at cryogenic temperature (100K) throughout the course of the data collection, except during our room temperature studies which were carried out at approximately 278K. For the original tetragonal crystals, we indexed and integrated the reflection data using *DENZO*, and performed scaling with *SCALEPACK*[111]. For all other crystals, we performed indexing, integration, and scaling of the X-ray data using *XDS* and *XSCALE*, and then converted intensities to structure factors with *XDSCONV*[112]. Diffraction data were routinely analyzed with *Xtriage* to check for crystal pathologies[113,114]. Further information regarding data collection and processing is presented in Table 2.1.

2.8.4 Structure determination

Using the program *PHASER* within the *CCP4* suite[115,116], we implemented the method of molecular replacement to calculate phases for diffraction data from the untreated crystals. Initially, we prepared a molecular replacement search model based on the structures of *E. coli* EutL. This model consisted of the core regions of the protein, and excluded any parts of the polypeptide chain that occupy different positions in the open and closed structures of the *E. coli* protein. Solutions were found for both tetragonal and monoclinic crystals, corresponding to spacegroup $P4_32_12$ and $C2$ respectively. Following structure determination from the original cpEutL crystals, the tetragonal model (PDB ID 4EDI) was used for phase calculation by molecular replacement to solve subsequent crystal structures.

For all crystal structures described in this work, a similar model-building and refinement strategy was used. First, using electron density maps calculated with phases from molecular replacement, we rebuilt the missing or incorrect parts of the structures. In some cases, we performed initial atomic refinement with simulated annealing to remove residual model-bias. We then performed iterative steps of manual model rebuilding and atomic refinement of the model to convergence with TLS parameters, a riding hydrogen model, and automatic weight optimization. All model-building was performed using *COOT* (v0.6.2)[117], and refinement steps were performed with *phenix.refine* within the *PHENIX* suite[113,118]. In cases where ligand molecules were modeled, the ligand coordinates were prepared with *phenix.elbow*[119]. Further information regarding model building and refinement is presented in Table 2.1.

2.8.5 Analysis of sequence conservation

We used the *ConSurf* algorithm[99,100] to analyze the evolutionary conservation of critical amino acids in EutL orthologs.

2.8.6 Measurement of channel dimensions

To analyze the dimensions of the molecular channels that traverse the cpEutL trimers, we used the *HOLE2* software[89]. We selected two points, one on either side of the channel present in chain A of our

untreated structure, and used them to define a vector that passed through this channel. This vector was provided as input to *HOLE2* in order to guide determination of the molecular channel.

2.8.7 Modeling static disorder

In order to model the static disorder in the cpEutL crystals resulting from the presence of a small population of the open conformation, we first prepared a model of the putative open conformation of cpEutL. We began by generating a homology model of cpEutL using the *I-TASSER* webserver[120] with the open conformation of ecEutL (PDB ID: 3I87) as a template. Parts of the model not expected to undergo conformational change based on homology with *E. coli* EutL were adjusted to match the closed (untreated) structure of cpEutL. The model of the open conformation was not subjected to atomic refinement against x-ray data. B-factors for all atoms in the open model were set to the Wilson-B value for the untreated x-ray dataset (25.7Å2). The experimentally determined closed structure was edited by removing water molecules and converting atomic B-factors to isotropic values using phenix.pdbtools (v1.7.3-928)[113].

For our experimentally determined closed structure and the modeled open structure, we used *phenix.reciprocal_space_arrays* (v1.7.3-928)[113] to calculate structure factors ($F_{c,closed}$ and $F_{c,open}$) from the models. These calculated structure factors were scaled to the experimentally measured structure factor magnitudes ($|F_{obs}|$) from the untreated x-ray dataset by *phenix.reciprocal_space_arrays*. We then calculated hypothetical structure factors wherein the fractional occupancy of the closed conformation was given by α, and the fractional occupancy of the open conformation was given by the complementary value (1-α). Hypothetical structure factors were then rescaled with the observed structure factor amplitudes from the untreated x-ray dataset, and used for crystallographic *R*-value calculations (Rwork/Rfree). Performing *R*-value calculations for a series of relative occupancy values allowed us to determine the fractional occupancy of the putative open conformation by minimizing the average residual error for structure factor calculations, despite our inability to clearly observe this conformation in the electron density.

2.8.8 Two-dimensional SDS-PAGE

We performed two-dimensional SDS-PAGE experiments using a cpEutL protein sample prepared in the same manner as the samples used for crystallization. After initial purification by metal affinity chromatography and dialysis into crystallization buffer, the protein sample was treated with 10mM oxidized glutathione for 30min and then passed over a Superdex200 10/300GL gel filtration column (GE Healthcare) to remove any polypeptides which became connected by spurious intermolecular disulfide bonding. Immediately following gel filtration, we added 0.2% (w/v) SDS and 100mM iodoacetamide, and incubated for 30min at 37°C to alkylate all cysteine thiols not involved in disulfide bonds. After alkylation, we removed unreacted iodoacetamide by a two-step dialysis procedure against a 1000-fold excess of 20mM Tris Buffer, 100mM NaCl, and 0.2% (w/v) SDS at pH 8.0. For the tryptic digestion of the protein, we used Proteomics Grade Trypsin (Sigma Aldrich). Digestion reactions contained 50μL each of protein sample and reaction buffer (40mM ammonium bicarbonate, 9% acetonitrile), plus 1μg of trypsin. We incubated the digestion reaction at 37°C for 16 hours, and quenched the reaction by adding non-reducing SDS-PAGE loading buffer.

We performed two-dimensional SDS-PAGE experiments according to the methods described by Boutz, et al.[121], using 16.5% polyacrylamide mini gels (Bio-Rad) and a Tris-Tricine buffer system[122]. In this experiment, the first dimension was run under non-reducing conditions, and the second dimension was run after reduction with 50mM DTT. We used Coomassie Brilliant Blue G-250 to stain the gel. In addition to the two-dimensional experiment, we also ran one-dimensional SDS-PAGE using samples containing trypsin without substrate, in order to verify that the trypsin cannot be detected by the staining procedure, thus confirming that the bands we see correspond to cpEutL fragments.

2.8.9 Mass spectrometry

For our mass spectrometry experiments, we prepared a cpEutL protein sample in the same manner as our protein used for crystallization. Two samples were prepared and one was treated with 50mM DTT (reduced), while the other was left untreated (non-reduced). We prepared a 1:10 dilution of

these protein samples in formic acid, and performed size-exclusion chromatography by HPLC in 90% formic acid using a column that was set up in-line with the mass spectrometer. Proteins were eluted directly into a Perkin–Elmer Sciex API III+ triple quadrupole instrument for electrospray ionization mass spectrometry (ESI-MS) analysis. Whitelegge, et al., have reported further details of the method used for mass spectrometry[123].

2.8.10 NMR spectroscopy

We prepared ^{15}N-labeled cpEutL protein following the same procedure described for the unlabeled protein, except expression was carried out in M9 minimal media containing $^{15}NH_4Cl$, rather than LB broth, and the expression time was extended from four hours to six hours. We purified the protein by metal affinity chromatography as described. Following initial purification, we treated the protein using 10mM oxidized glutathione for 30min at room temperature, and then loaded the sample onto a HiLoad Superdex200 16/60 gel filtration column (GE Healthcare) equilibrated in 20mM Tris, pH 7.0, and 100mM NaCl to remove both glutathione and any protein molecules linked by intermolecular disulfide bonding. We concentrated the eluted protein to 0.6mM using an Amicon Ultra concentrator (Millipore), and added 5% D_2O to obtain the oxidized NMR sample.

We recorded the NMR spectrum at 298K on a Bruker Avance 800MHz spectrometer equipped with a cryogenically cooled triple-resonance TCI (^{1}H, ^{13}C, ^{15}N) probe. We recorded the two-dimensional ^{1}H-^{15}N TROSY-HSQC spectrum with 64 transients and 3,586 (^{1}H) x 156 (^{15}N) complex points. After recording the spectrum of the oxidized sample, we retrieved the sample and added 10mM DTT, incubated for 30min at room temperature, and collected the reduced spectrum following the same data collection protocol used for the oxidized spectrum. We processed the spectra using TopSpin (Bruker), and overlaid the spectra using the program Sparky.

2.8.11 Computational analysis of disulfide bonding

For the quantification of potential disulfide bonds in BMC domain proteins, we applied a variation of the sequence-structure mapping approach first introduced in earlier work[94] and then recently extended[95]. First, we collected all of the available sequences from protein family PF00936 in the Pfam database[124], which is comprised of the BMC domain proteins. At the time of our analysis, this protein family contained 3317 unique entries within the BMC domain protein family (PF00936), 969 of which were tandem BMC domain proteins. We filtered these sequences based on a 95% identity criterion. Each query sequence derived from Pfam was matched to its closest BMC domain homolog from the PDB[125] based on pairwise sequence alignments. We identified a potential disulfide bond whenever a pair of cysteine residues from the query sequence mapped to positions in which the C-alpha atoms of the corresponding residues in the homologous hexamer or trimer structure was measured to be less than or equal to 8Å. Performing the analysis in the context of a complete hexamer or pseudo-hexameric trimer provided the two kinds of assemblies the same opportunities for both inter- and intra-domain disulfides to be identified.

2.8.12 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) experiments were performed using a protein sample from which the N-terminal affinity tag had been removed by protease cleavage. In order to very accurately measure the protein concentration of our ITC sample by UV absorption (λ=280nm), we used a molar extinction coefficient derived from amino acid analysis. The protein sample was concentrated to 0.6mM in buffer containing 20mM Tris at pH 7.2, and 100mM NaCl. The titrant was prepared by dissolving ethanolamine hydrochloride in the identical buffer to a final concentration of 50mM. We used the hydrochloride salt of ethanolamine to prevent pH differences between the protein sample and the ligand. Titrations were performed at 298K using an iTC-200 calorimeter (GE Healthcare/MicroCal). Similar titrations were performed using ethanol, acetate, and acetyl-phosphate as ligands.

We collected ITC data for the chemically-reduced samples by adding 20mM TCEP to the protein and ligand solutions and repeating the titration described above.

Raw ITC data were processed using the *Origin* software. The raw data were integrated to give plots of ΔH versus molar ratio of ligand to protein. For the ethanolamine titrations, we used *Origin* to fit the titration curves to a two-site, sequential binding model and extract thermodynamic parameters for the EutL-ethanolamine interaction.

2.8.13 Calculation of electrostatic surface potential

Electrostatic surface potentials were calculated using the Poisson-Boltzmann equation, as implemented in the *APBS* plugin for *PyMol*[126].

2.8.14 Time-averaged X-ray refinement of structural ensembles

Time-averaged structural ensembles were refined against X-ray diffraction data using the method developed by Burnley, et al. and implemented in *phenix.ensemble_refinement*[98]. We used the original observed X-ray diffraction data and the final single-copy atomic models of specified cpEutL structures as inputs for the ensemble refinement procedure. We optimized the ensemble refinement protocol by first optimizing the *ptls* parameter, followed by the *wxray_coupled_tbath_offset* parameter, and then finally the *tx* parameter, choosing the models that gave the best R_{work} and R_{free}. For the final ensemble refinements, we used values of *ptls*=0.90 for both the cryogenic and room temperature apo structures, and *ptls*=1.0 for the cryogenic, ethanolamine-bound structures, and the default parameters for *wxray_coupled_tbath_offset* and *tx* were found to yield the best results.

CHAPTER 3

A Challenging Interpretation of a Layered Hexagonal Protein Structure

3.1 Introduction

The β-carboxysome is a well-studied MCP[12]. It is involved in the carbon fixation reactions of the cyanobacterial Calvin cycle (Figure 3.1). This MCP sequesters the enzymes carbonic anhydrase and RuBisCO, thereby increasing the local concentration of carbon dioxide in the vicinity of RuBisCO, an enzyme with notoriously poor catalytic efficiency and substrate selectivity. This chapter describes the structure determination of a CcmK1 shell protein mutant (L11K) from the β-carboxysome of the cyanobacterium *Synechocystis* PCC6803. Within the β-carboxysome, the CcmK paralogs (K1, K2, K3, K4) are the primary homohexameric components of the shell, and it has been hypothesized that the size, shape, and chemical properties of their central pores promote the diffusion of substrates and products over other molecules.

Consistent with the form of their natural biological assemblies, BMC shell proteins often crystallize in hexagonally packed layers[29,33,127,128]. Layered structures are generally prone to various crystal growth pathologies[129], and this has proven to be especially true of BMC shell proteins, where twinning and lattice translocation disorders have been observed often[22,130]. The present study involved another such case in which a specific shift of hexagonal layers in a crystal led to a challenging space group interpretation. This case demonstrates that when broken crystallographic symmetry leads to doubling of a unit cell axis, it is sometimes necessary to consider an alternate unit cell origin in order to identify the highest symmetry space group.

In addition to providing an instructive example for crystallographers, the work presented here offers interesting insight into the function of the CcmK1 BMC-domain shell protein. The alternating

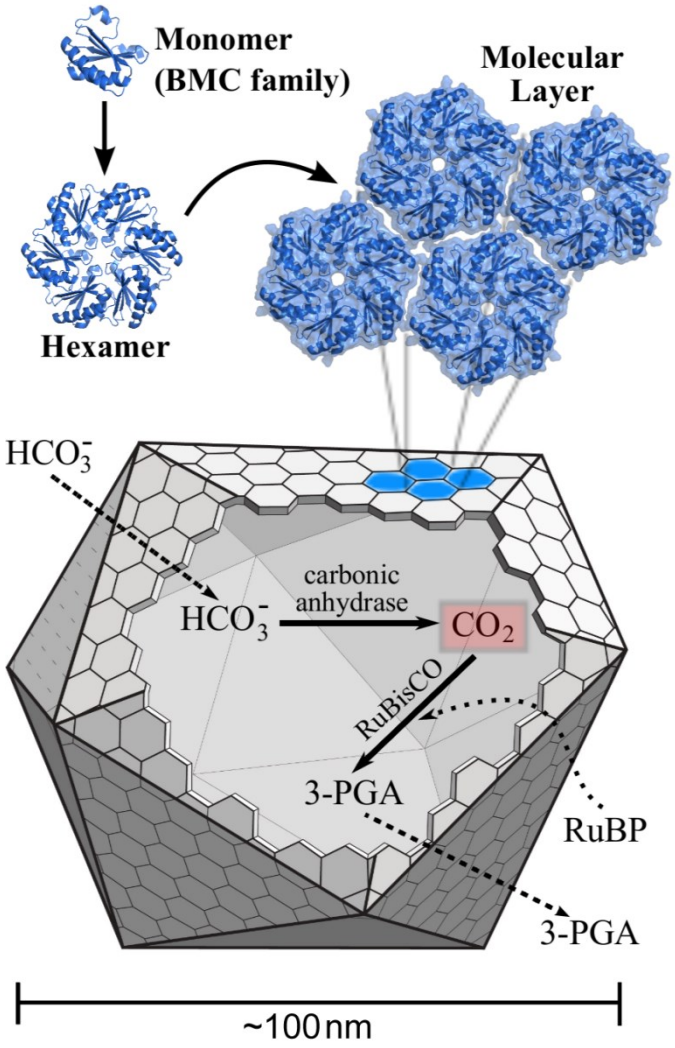


Figure 3.1. A model for bacterial microcompartment (MCP) structure and function in the carboxysome. The illustration highlights the assembly of shell protein monomers into hexamers, as well as the edgewise association of hexamers to form the tightly-packed facets of the polyhedral MCP shell. Within the shell, a reaction scheme depicts the metabolic events that occur in the lumen of the carboxysome. The critical intermediate, CO2, is highlighted in red. In the reaction scheme, solid black lines depict enzymatic reactions, while dashed lines indicate transport events. (RuBP = ribulose-1,6-bisphosphate; 3-PGA = 3-phosphoglyceric acid; RuBisCO = ribulose-1,6-bisphosphate carboxylase/oxygenase).

conformations of individual shell protein monomers in the crystal form used for this study provide potential connections to function, implying a certain amount of flexibility and dynamic motion in the central region of the CcmK1 hexamer. Also, analysis of the electrostatic surface potential of the L11K mutant in comparison to the wild-type protein demonstrates that the electrostatics of the central pore are influenced by amino acids outside of the β2-β3 loop.

3.2 Materials & Methods

3.2.1 Cloning, Overexpression, and Purification of Recombinant Protein

The wild-type DNA sequence for residues 1-91 of the CcmK1 protein from the cyanobacterium *Synechococystis sp.* PCC 6803 was cloned into the pET-22b expression vector incorporating a C-terminal hexahistidine tag consisting of amino acids –LeuGluHis$_6$. Specific details of the cloning protocol have been described previously[33]. The L11K mutation was introduced by site-directed mutagenesis using the QuickChange method (Stratagene). The sequence of the mutated plasmid was verified by dideoxy chain termination sequencing.

We expressed recombinant protein using transformed *Escherichia coli* BL21 (DE3) Rosetta cells (Novagen). During the exponential phase of cell growth in selective Luria-Bertani (LB) Broth, 1mM isopropyl-β-D-thiogalactopyranoside was added to induce protein expression for four hours at 37°C. Cells were pelleted by centrifugation for 15min at 5,000x*g* and then lysed by sonication. The lysis buffer contained 20mM Tris buffer and 300mM sodium chloride at pH 8.0 with a protease inhibitor additive (Sigma Aldrich), 10mM $MgCl_2$, 1mg/mL lysozyme, and 100units/mL of both DNase and RNase. We clarified the cell lysate by centrifugation at 30,000x*g* for 30min, and then used a HisTrap nickel affinity column (GE Healthcare) to purify the protein from clarified lysate. The bound protein was eluted with lysis buffer containing 300mM imidazole, and then dialyzed against a buffer containing 20mM Tris and 100mM sodium chloride at pH 8.0. This single purification step resulted in a highly pure protein sample, as demonstrated by SDS-PAGE.

3.2.2 Protein Crystallization

Following purification, the protein was concentrated to approximately 10mg/mL. We crystallized the CcmK1 L11K mutant by hanging-drop vapor diffusion in mother liquor containing 1.26M ammonium sulfate and 0.15M sodium chloride, buffered at pH 9.5 with 0.1M N-cyclohexyl-2-aminoethanesulfonic acid. Crystallization drops were prepared by mixing a 1:1 ratio of the concentrated protein solution and the mother liquor in a total volume of 1μL, using a Mosquito pipetting robot (TTP Labtech). The drops were sealed above 100μL reservoirs and allowed to equilibrate at 23°C. Our initial crystals diffracted poorly, and we found that addition of 10mM guanidinium chloride to the mother liquor prior to crystallization produced a large, high-quality crystal that diffracted well.

3.2.3 X-ray Data Collection and Processing

We collected single-crystal X-ray diffraction data using a Rigaku FR-E+ rotating anode X-ray source operating at the CuKα wavelength (1.54Å), equipped with VariMax HF optics and an R-axis HTC detector. Crystals were harvested and frozen directly in the liquid nitrogen cryostream without additional cryoprotection, and were subsequently maintained at cryogenic temperature (100K) throughout the course of the data collection. We indexed and integrated the reflection data to 1.6 Å resolution using *XDS*, performed scaling with *XSCALE*, and converted intensities to structure factor amplitudes using *XDSCONV*[112]. The free set of reflections was assigned using *phenix.reflection_file_converter*[113] such that reflections related by lattice symmetry would belong to the same set. Information regarding data collection and processing is presented in Table 3.1. We note that the data are strong out to the 1.6 Å resolution limit used in our X-ray experiment; the geometry of the detector setup precluded the collection of higher resolution data.

Parameter	Value
X-ray Wavelength	1.54Å
Nominal Resolution Range	41.24-1.6Å (1.64-1.60Å)
Unit Cell	a=b=70.0Å ; c=56.2Å α=β=90° ; γ=120°
Space Group	$P6_3$
Total Reflections	353,118 (7262)
Unique Reflections	20,387 (1293)
Multiplicity	17.3 (5.6)
Completeness	98.0% (84.6%)
$\langle I/\sigma I \rangle$	34.07 (5.55)
$R_{r.i.m.}$[131]	5.7% (26.8%)
$CC_{½}$[86]	100.0 (96.4)
Wilson B-factor	25.0Å^2
R_{work}[a]	18.0%
R_{free}[a]	19.5%
No. of Atoms	1441
Protein Residues	182
Water Molecules	51
Average B-factor: Protein Solvent	 19.7Å^2 26.5Å^2
RMSD(bonds)	0.012Å
RMSD(angles)	1.32°
Ramachandran Plot[87]: Favored Allowed Outliers	 97.8% 2.2% 0.0%
Molprobity Clashscore[88]	4.66

TABLE 3.1 Diffraction data and refinement statistics for CcmK1 L11K.

Values in parentheses reflect the highest resolution shell.

[a] R_{work} and R_{free} are given by the following equation, computed for the working and test sets of reflections respectively: $R = \frac{\sum_{hkl} \| F_o | - | F_c \|}{\sum_{hkl} | F_o |}$

The values of F_c used in the calculations include scattering contributions from the riding H atoms.

3.2.4 Analysis of Symmetry and Twinning

Automatic data indexing by *XDS* (as describe above) revealed that the crystal lattice was primitive hexagonal. During the early stages of data processing, we evaluated different Laue symmetries and also evaluated a self-rotation function using *MOLREP*[132,133], which suggested that the diffraction pattern obeyed 622 rotational symmetry (Figure 3.2a). We also ran the *phenix.xtriage* program[113,114] to check for crystal pathologies and non-crystallographic symmetry (NCS). The *Xtriage* analysis revealed a strong peak in the native Patterson map[134], as discussed subsequently. In addition to revealing translational NCS, the *Xtriage* analysis also suggested the presence of hemihedral twinning based on the results of the L-test (Figure 3.2c)[135].

3.2.5 Phasing by molecular replacement

In order to calculate initial phases, we used the method of molecular replacement (MR) with a structure of the CcmK2 BMC shell protein (PDB ID 3DNC) as a search model. The program *Phaser*[115] found a solution containing four monomers in a *P*3 asymmetric unit (2 hexamers per unit cell). This solution was used as a starting point for model building and refinement of our CcmK1 L11K structure.

3.2.6 Model building and refinement

As an initial step toward generating our final model, we subjected the MR solution to simulated-annealing torsion angle refinement with *phenix.refine*[113,118] in order to minimize potential model phase bias resulting from the MR procedure. We began refining our model in space group *P*3. However, during the course of the refinement process we noticed that our model was also consistent with the higher-symmetry space group $P6_3$ under a different choice of origin, as discussed subsequently. At that point we selected two of the four chains from our structure in *P*3, which corresponded to a $P6_3$ asymmetric unit under the appropriate choice of origin. For subsequent model refinement we merged our diffraction data to $P6_3$ using *XSCALE*, taking care to conserve the R_{free} flags from our original *P*3 data. Our early stages of refinement produced a relatively high-quality model, though some regions of the resulting electron

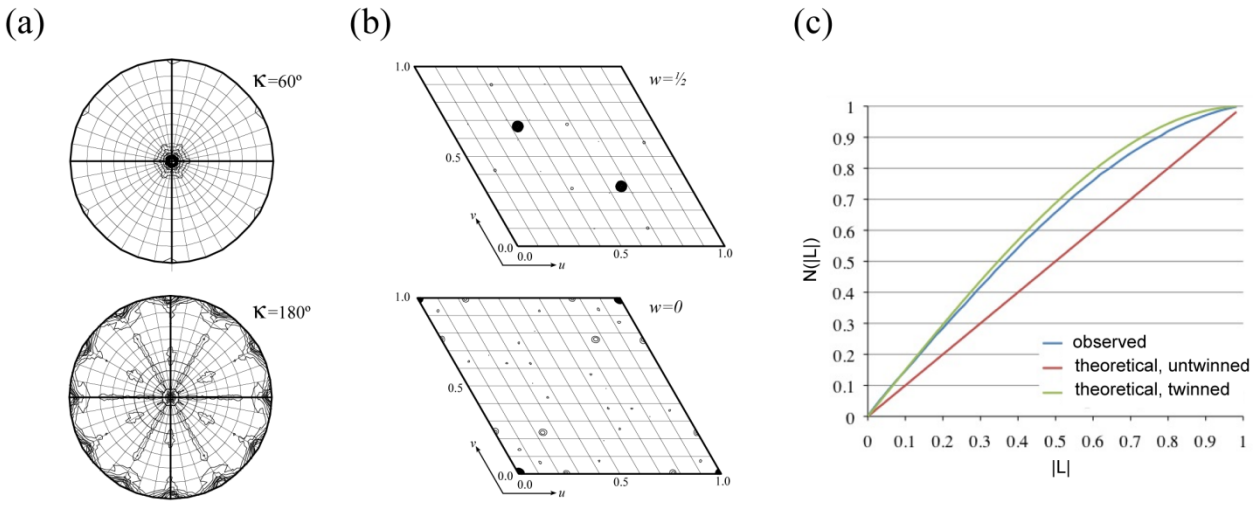


Figure 3.2. Analysis of symmetry and twinning. A) Sections of the self-rotation function (κ=180° and κ=60°) indicate that the underlying point group symmetry of the crystal is 622. B) Sections of a native Patterson map (w=0 and w=½) calculated from observed intensities show a prominent (47.5% of the origin) packing peak at =(⅓,⅔,½). C) Intensity statistics showing that the CcmK1 L11K crystal specimen is hemihedrally twinned. N(|L|) is the cumulative distribution of |L|.

density map appeared fragmented or featureless, and were not consistent with the model in those areas. In order to resolve these problematic regions of the model, we generated an electron density map in which questionable regions of the model were not included in the phase calculation. Using this map, we were able to manually fix incorrectly built regions of the model with the program *COOT*[117]. Finally, we refined the model to convergence against merged structure factor amplitudes with automatic weight optimization and a TLS model for atomic displacement parameters (10 groups per chain, as determined by a combination of automated *TLSMD* analysis[136,137] and visual analysis of secondary structural elements), as well as twin refinement with algebraic detwinning, twin operator (*k,h,-l*), and a twin fraction of 0.4. Hydrogen atoms were added to the model using *phenix.reduce* (Word et al., 1999), and were included in the riding positions throughout refinement for the purposes of maintaining good geometry and including their scattering contributions, but their positions were not independently refined against the X-ray data or included in the deposited coordinates. We performed the final refinement step using *phenix.refine* within version 1.8.2-dev1334 of the *PHENIX* suite[113,118].

Our final model includes two protein monomers in the asymmetric unit, including residues 3-94 out of 99 in chain A, and 3-92 out of 99 in chain B, as well as 51 water molecules and a single sulfate ion. Regions of the $2mF_o$-DF_c and mF_o-DF_c maps near the C-termini of the protein chains were somewhat noisy, indicating that the C-terminal hexahistidine tags might be partially ordered. However the features were not clear enough to justify extending the model into this density. Following model refinement, we evaluated the quality of the structure and compared the two crystallographically independent protein molecules in the unit cell using the *PHENIX* graphical interface[113]. The atomic model has been deposited in the Protein Data Bank under PDB ID 4LIW.

3.2.7 Calculation of Electrostatic Surface Potential

Electrostatic surface potentials were calculated using the Poisson-Boltzmann equation, as implemented in the *APBS* plugin for *PyMol*[126].

3.3 Results & Discussion

The CcmK1 L11K shell protein mutant *from Synechocystis sp.* PCC6803 was expressed and purified from *E. coli*, and crystallized by hanging drop vapor diffusion (see Methods). Diffraction data were collected to 1.6Å resolution. The unit cell was determined to be hexagonal (**a**=**b**=70.0Å , **c**=56.2Å), and data reduction suggested the Laue symmetry to be *P*6/*mmm*. Diffraction was strong in all directions, Bragg peaks were sharp, and no obvious signs of disorder were evident. However, a very strong native Patterson peak (47.5% of the origin height) was noted at fractional coordinates $\langle uvw \rangle$=(⅓, ⅔, ½) (Figure 3.2b). Although this peak appears to be a special position, it does not correspond to any centering operation in an alternate unit cell. Instead, such a translation interchanges the locations of six-fold and three-fold symmetry axes in *P*6.

3.3.1 Initial space group identification (*P*3)

Some of the essential features of the final structure could be discerned at the outset from the geometry and symmetry of the unit cell, based on prior experience with crystal structures of similar proteins from the BMC shell protein family. In many crystal structures, BMC shell proteins pack in hexagonal layers with a unit cell spacing along **a** and **b** in the range between 65Å and 70Å[29,33,127,128]. In their natural assembly state, the hexamers pack side by side, and in the same (i.e. upward facing) orientation, to give an essentially solid molecular layer, except for narrow pores at the centers of the hexamers. The thickness of one layer of proteins is generally just under 30 Å at its thickest point. In the present case, the unit cell value of **c**=56.25Å dictated that two layers of hexamers would be present within one unit cell. However, a challenge arose immediately from the observed Laue symmetry and the strong translational NCS peak in the native Patterson at **w**=1/2. The translational NCS peak required the two layers in the unit cell to be oriented likewise (i.e. both face up). However, this arrangement would not give the 622 rotational symmetry (apparent Laue symmetry *P*6/*mmm*) observed in the diffraction data. This conflict was resolved by realizing that the specimen was almost perfectly twinned by hemihedry. An

evaluation of the overall intensity statistics, using local differences, gave a value for $\langle|L|\rangle$=0.394 (Figure 3.2c). The theoretical expected value of $\langle|L|\rangle$ is 0.500 for untwinned data and 0.375 for perfectly twinned data[135]. This finding allowed for the two layers in the unit cell to share the same molecular orientation; the additional 2-fold rotational symmetry in the diffraction pattern arises from hemihedral twinning.

Given the two layers in the same orientation, what remained was to establish the shift between them. The relative shift was dictated by the native Patterson peak at (⅓, ⅔, ½). In this arrangement, the axis of local 6-fold symmetry in each layer is coincident with a local axis of 3-fold symmetry (where the corners of three hexamers meet) in the other layer (Figure 3.3a). As a result, the 6-fold symmetry axes through the centers of the hexamers in one layer are broken by the other layer, leading to a system of equally spaced 3-fold crystallographic axes of symmetry, as in space group *P*3. The apparent non-crystallographic, translational relationship between the two layers (evidenced by the native Patterson peak), together with the breakage of the 6-fold symmetry, led to a space group assignment of *P*3 and to a molecular replacement solution consistent with that symmetry. The asymmetric unit contained two copies of a third of a hexamer, one in each layer – a total of four polypeptide chains (Fig. 3a). Despite the absence of 6-fold rotational symmetry in the *P*3 space group, 6-fold symmetry was present in the calculated intensities as a result of the local 6-fold symmetry of the hexamers and the special translational shift between them (in agreement with the 6-fold symmetry of the observed intensities)[138,139]. After confirming the packing analysis above by molecular replacement, using CcmK2 (PDB ID 3DNC) as the search model, we were able to refine an atomic model in space group *P*3 that had good final statistics (R_{work}/R_{free}=0.171/0.199).

3.3.2 Identification of higher-symmetry (*P*6$_3$) by shifting the unit cell origin

During the refinement process, careful analysis of the refined model in space group *P*3 illuminated an element of symmetry that had been overlooked. A single layer of molecules in our crystal supports *P*6 symmetry, with the origin at the center of a hexamer. Our initial space group assignment for

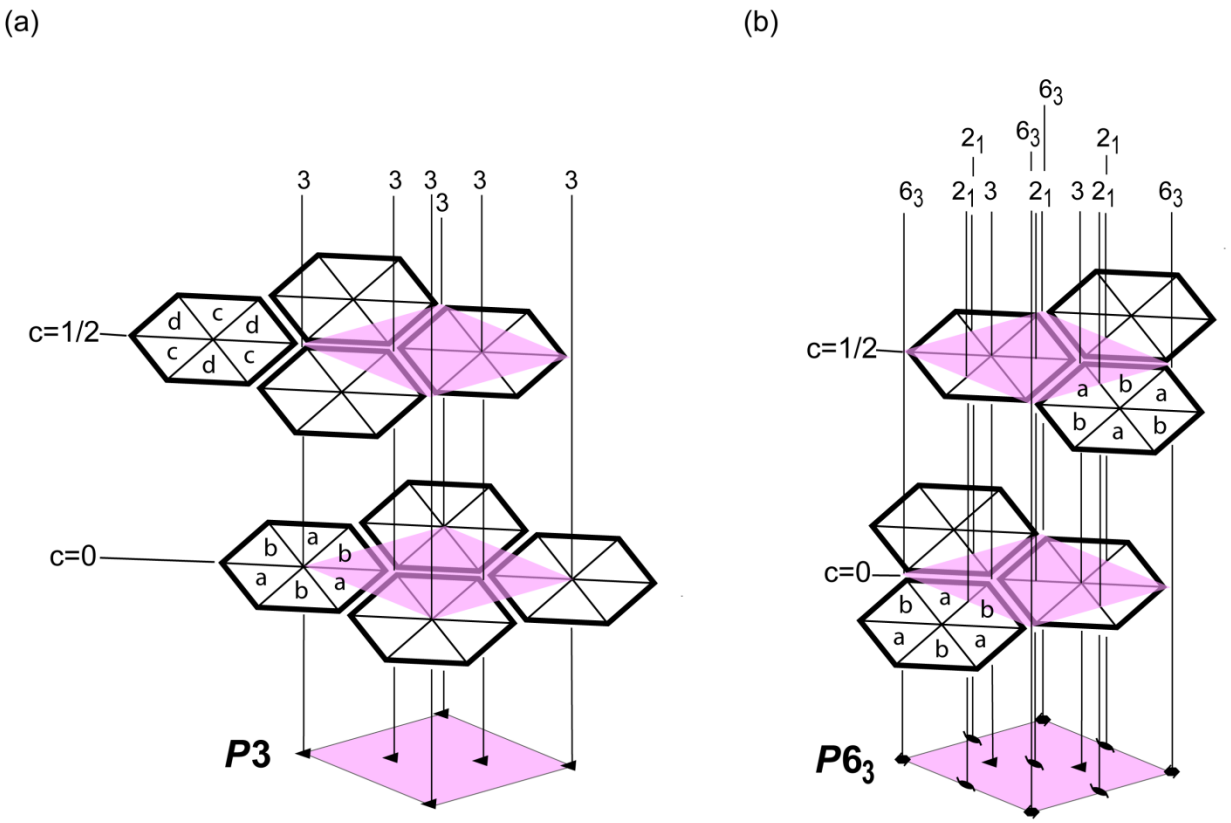


Figure 3.3. Space group assignment and packing arrangement of the two hexameric layers in the unit cell. A) An initial assignment of space group P3, with one hexameric layer having its center at the origin and the other layer having its center at (1/3, 2/3, 1/2), leading to an essentially correct structure, but with incompletely assigned symmetry. The four crystallographically independent subunits based on this space group assignment are labeled a-d. B) The correct assignment of P63 symmetry, after shifting the initial structure to a different origin. Note that the 63 screw axis contains within it a pure 3-fold axis and a 21 screw axis of symmetry. Only two crystallographically independent subunits (a and b) are present. Symmetry elements are illustrated with their conventional symbols.

the crystal (*P*3) assumed that although the *P*6 unit cell doubled along **c** and the 6-fold symmetry was broken due to the translational shift between layers of molecules, the origin would remain at the center of a hexamer. Additionally, when performing the initial space group assignment, we failed to recognize that although the translational shift breaks a pure 6-fold crystallographic axis, it introduces a 6_3 axis. Consequently, our initial MR search was performed in space group *P*3, rather than in any member of the $P6_x$ family. Upon closer analysis, we realized that an alternate choice for the origin of the unit cell would allow for a higher symmetry to be assigned, without significant modification of the structure.

The correct space group conformed to $P6_3$ (in its standard setting) only when the origin of the unit cell was placed on a 3-fold axis of symmetry that passes through points in both layers where three hexamers meet at a corner (Figure 3.3b). The two layers are then related by a 2_1 screw axis (contained within the 6_3 screw axis) through the origin. Under the space group assignment of $P6_3$, the asymmetric unit contains just one-third of a hexamer, or two adjacent subunits whose slightly different conformations give the hexamers 3-fold, but not 6-fold, symmetry.

The NCS packing peaks between the two layers (at **w**=½ in the native Patterson map) arise from a combination of the 2_1 crystallographic axis with local 2-fold NCS axes through the centers of the hexamers (along **c**), which are slightly broken by subtle deviations between the two monomers in the asymmetric unit. The combination of two rotational symmetry elements to produce this translational NCS introduces another source of confusion in indentifying the presence of 6_3 and 2_1 screw axes in $P6_3$; these screw axes are typically identified by systematically absent (0,0,l=odd) reflections, which would also be absent as a result of a translational NCS operation with *z*=½.

The correct space group assignment was further confirmed retrospectively by rerunning molecular replacement in all space groups having Laue symmetry *P*6/*m*, where $P6_3$ was readily identified as the correct space group. Additionally, a *post hoc* analysis of the *P*3 coordinates with the *LABELIT* software[140] also identified the higher symmetry $P6_3$ unit cell, further confirming our manual analysis. With this correct space group (and choice of origin), and application of the necessary twin law (*k, h, -l*), a final model was successfully refined (Table 3.1).

3.3.3 Broken crystallographic symmetry often requires a new choice of unit cell origin

The space group complication in the present study shares some similarities with another recently described case of symmetry breaking. While studying a new crystal form of human carbonic anhydrase II, it was found that alternating protein conformations led to a unit cell doubling compared to a previously characterized, simpler crystal form[141]. Unexpectedly, however, the $P2_1$ space group symmetry of the simpler crystal form was also broken under the most obvious choice of origin for the doubled unit cell, leading to apparent disorder. The correct structure, which was fully ordered, was obtained in retrospect when it was found that shifting the origin of the doubled unit cell led to recovery of the required symmetry elements in $P2_1$[142]. Likewise in our study, the shifting of a second hexagonal layer relative to the first layer breaks the *P*6 symmetry that would have been present in a single layer, leaving what appears to be *P*3 symmetry under the original choice of unit cell (Figure 3.3a). The assignment of P3 symmetry is not incorrect, and in our case no apparent disorder resulted from refining the structure in P3, but higher symmetry ($P6_3$) is in fact present, and is recognized under a different choice of origin (Figure 3.3b). Cases such as these serve as reminders that when symmetry-breaking gives rise to a new crystal form, identifying the highest possible symmetry in the new form may require a different choice of origin in order to match the correct space group in its standard setting. This requires either careful manual intervention or a reanalysis by computational methods able to automatically evaluate new space group possibilities[140,143].

3.3.4 Symmetry-breaking may play a role in shell protein function

The final model was mainly consistent with previous structures of homologous BMC shell proteins. However, the packing arrangement of hexamers in this crystal provides potential insight into the natural symmetry of these cyclic homo-oligomers. If we assume the hexamers are 6-fold symmetric, and that the hexamers in one layer prefer to not pack directly on top of hexamers from adjacent layers, then there would be six energetically equivalent translational shifts between hexamers in adjacent layers.

These six translational shifts admit two distinct positions for the second layer relative to the first. The energetic equivalence of those two outcomes would lead to growth of a crystal exhibiting a lattice translocation disorder, with hexamers occupying mutually exclusive (partially occupied) positions within one layer of molecules. This lattice translocation disorder would produce strong packing peaks in a native Patterson map at **w**=0, similar to a case previously observed in crystals of another BMC shell protein[130]. As evidenced by the lack of packing peaks in the **w**=0 section (Figure 3.2b), our structure does not exhibit such a pathology. This indicates that the hexamers add to the crystal in a way that breaks their 6-fold symmetry. Deviations from 6-fold symmetry, although subtle, were indeed observed in the refined atomic coordinates. These deviations include small differences in rotameric configurations and backbone torsion angles between the two crystallographically independent protein molecules in the unit cell. These rotamer and torsion angle differences cause a departure from perfect β-sheet geometry in chain A versus chain B, as well as a small difference in the conformation of the C-terminal helix in each chain, which appears to be a 3_{10} helix in chain A, but not in chain B. The breakage of 6-fold symmetry also brings up an element of potential biological interest. The broken symmetry implies – by exchange of alternating conformations – a certain amount of dynamic motion in this family of proteins, which may relate to their roles in molecular transport. The origin of the alternating conformations might be a need to alleviate a steric clash created by the crowding of six protein chains into the center of a cyclic oligomer. The role of symmetry-breaking in these systems is the subject of ongoing analysis.

3.3.5 The L11K Mutation has a Significant Effect on CcmK1 Pore Electrostatics

In order to examine the effect of the L11K mutation on the structure and function of CcmK1, we compared the coordinate positions and electrostatic surface potentials of the mutant and wild-type proteins. First, we superimposed the structures of the mutant and wild-type proteins, and found very low coordinate RMSD values for the structural alignment (0.32Å on all atoms). The alignment revealed that even in the immediate vicinity of the L11K mutation, there is virtually no structural change between the mutant and wild-type proteins. Next we calculated electrostatic surface potentials for the mutant and wild-

type proteins using the Poisson-Boltzmann equation[126]. These calculations showed that the L11K mutation has a significant effect on the electrostatics of the central pore. The pore electrostatics of the CcmK1 L11K mutant are very similar to the electrostatics of CcmK2, whose sequence is highly identical to CcmK1, but has arginine at the eleventh position. The electrostatic properties of the CcmK pores have been hypothesized to influence the diffusion selectivity of the carboxysome shell[22,25,27,29,33,127]. This work begins to highlight important residues, although a more exhaustive study, which includes experimental verification of structural observations, will be required to develop a complete understanding of how electrostatics and other chemical properties effect the transport functions of the CcmK proteins.

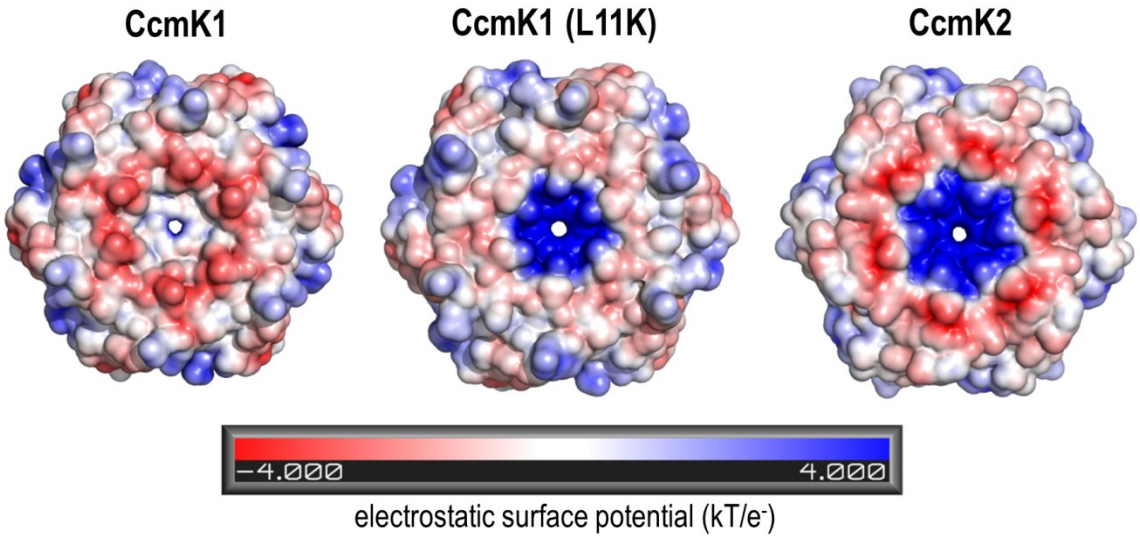


Figure 3.4. Electrostatic surface representations of CcmK proteins. Electrostatic surfaces, generated by solving the Poisson-Boltzmann equation, of CcmK1, CcmK1 with the L11K mutation, and CcmK2. The images demonstrate that a single point mutation can have a relatively dramatic effect on the pore electrostatics. The L11K mutation makes the CcmK1 surface more similar to CcmK2, which has arginine at the same position as the L11K mutation.

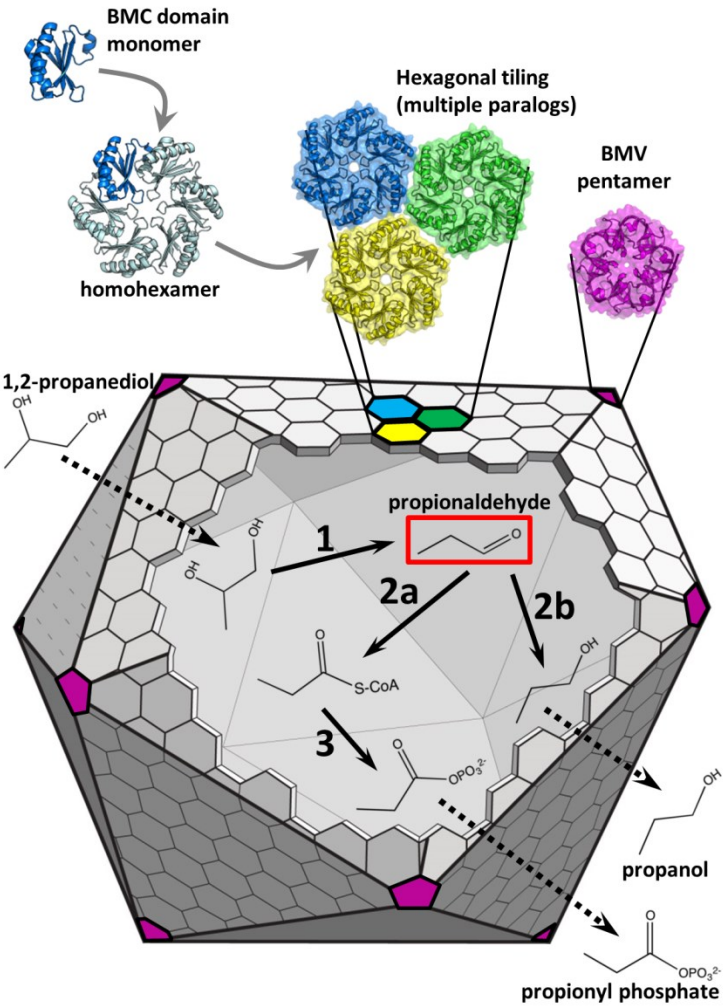
CHAPTER 4

Identification of a Unique Fe-S Cluster Binding Site in a Glycyl-Radical Type Microcompartment Shell Protein

4.1 Introduction

Prokaryotic organisms have evolved remarkable metabolic diversity through adaptation to highly varied environments. Many organisms have specialized metabolic pathways that allow them to utilize small molecules that are abundant within their specific niches as key carbon sources. In a number of intriguing cases, however, these specialized metabolic pathways produce intermediate compounds that are cytotoxic, mutagenic, and/or diffuse freely across the cell membrane at physiological pressure and temperature[7,15,21]. In such cases, the bacterial cell gains an advantage be being able to confine the intermediate within a MCP. An example of one such MCP sequesters the initial steps of 1,2-propanediol catabolism in saccharolytic pathogens (Figure 4.1)[15,28]. When these pathogenic bacteria infect their host, they are capable of hydrolyzing cell-surface glycans from the host tissue, releasing monosaccharides, which become a primary source of metabolic energy for the proliferating bacteria[144]. These carbohydrate molecules, notably fucose and rhamnose, are first degraded to 1,2-propanediol[145–147]. Propanediol is then metabolized to propionate, which enters the methylcitrate cycle, providing a source of metabolic energy[15]. The conversion of 1,2-propanediol to propionate proceeds through a reactive propionaldehyde intermediate, which is both cytotoxic and mutagenic if released into the cytoplasm[15]. Organisms that are capable of propanediol catabolism, therefore, must encapsulate several steps of the pathway within an MCP to sequester the aldehyde intermediate.

Two distinct types of 1,2-propanediol catabolic MCPs have been identified and studied previously (Figure 4.1). The first of these two types, the Pdu MCP (for propanediol utilization), has been studied for a number of years and is relatively well-understood. Within the Pdu MCP, 1,2-propanediol is



	Pdu	Grp
Reaction 1	B_{12}-dependent diol dehydratase	glycyl-radical diol dehydratase
Reaction 2a	phosphotransacylase	
Reaction 2b	propionaldehyde dehydrogenase	
Reaction 3	alcohol dehydrogenase	

Figure 4.1. Schematic of propanediol catabolic MCPs. Degradation of 1,2-propanediol occurs in a MCP. Homohexameric BMC-domain shell protein paralogs assemble into a proteinaceous shell surrounding a series of internalized enzymes that convert the substrate, 1,2-propanediol, into the products, ethanol and acetyl phosphate. The chemistry involves a toxic intermediate (acetaldehyde), which is highlighted by a red box. The first of these enzymatic steps differs in Pdu-type vs. Grp-type MCPs. Dashed black arrows represent molecular transport events, while solid black arrows represent enzymatic reactions.

first converted to propionaldehyde by a vitamin B12-dependent diol dehydratase enzyme (PduCDE)[15,145]. The aldehyde is then converted to propionyl-CoA by an aldehyde dehydrogenase enzyme, or to propanol by an alcohol dehydrogenase enzyme, which facilitates recycling of the NAD+ cofactor within the MCP[82]. In the final enzymatic step that occurs within the MCP lumen, propionyl-CoA is converted to propionyl-phosphate. The second type of propanediol catabolic MCP has been implicated only recently through microarray analysis[146] and a comparative genomic analysis[10]. This MCP degrades 1,2-propanediol using a series of enzymatic reactions similar to those found in the Pdu MCP, except the initial dehydration reaction is catalyzed by a glycyl-radical diol dehydratase enzyme, rather than a vitamin B12-dependent enzyme. Consequently, this MCP has been named Grp, for glycyl-radical propanediol. In the case of the Grp MCP, it has been suggested that in addition to retaining the propionaldehyde intermediate, the MCP shell might also maintain an anaerobic environment for the glycyl-radical enzyme, whose activated state is sensitive to oxygen-mediated cleavage of the polypeptide backbone[148].

In contrast to the Pdu MCP, which has been studied in detail, only limited experimental work has been done to characterize the Grp MCP. Along with a genomic analysis that highlighted the Grp MCP operon, Jorda, et al. provided limited biochemical analysis of a BMC domain shell protein from the Grp MCP[10]. Almost simultaneously, Petit, et al. demonstrated that *Clostridium phytofermentans* expresses Grp-type MCPs during the fermentation of fucose and rhamnose[146]. Shortly after, Wheatley, et al. determined the structure of a pentameric vertex protein, GrpN, from the Grp MCP of Rhodospirillum rubrum[26]. These three reports provide the only experimental characterization of the Grp MCP to date. Because of its overall similarity to the Pdu MCP, some of the functional details of the Grp MCP can be inferred by analogy. However, the difference in the first enzymatic step between these two systems also implies a certain degree of functional divergence. For example, the Pdu MCP contains a system of enzymes (PduGHOS) that regenerates inactivated B12 cofactors[28,149,150], while the Grp MCP instead contains a glycyl-radical enzyme activase[10]. Each of these two biochemical systems requires access to a different set of substrates, cofactors, and reducing equivalents, which indicates that the protein shells of the Pdu and Grp systems must differ to some extent with respect to their function as a diffusion barrier.

Therefore, further experimental characterization of the protein shell of the Grp MCP should highlight the similarities and differences between the Pdu and Grp MCP systems, and will provide insight into how the Grp MCP supports the activation of an encapsulated glycyl-radical enzyme.

Some Grp operons contain a gene encoding a particularly divergent BMC domain shell protein. Analysis of various MCP operons revealed that it is fairly common for a particular type of MCP to vary in gene composition from one species to another[6,10]. These species-to-species variations in MCP composition are presumed to represent functional adaptations that modify the "core" MCP machinery, which is invariant between species. In the genomic analysis of the Grp MCP, Jorda, et al. counted that 9 out of the 23 Grp operons they examined contain a gene corresponding to a divergent BMC-domain shell protein[10]. The closest homolog of this polypeptide in the Pdu MCP system is the PduU shell protein, and consequently we have named this divergent Grp shell protein "GrpU." The GrpU protein is an interesting subject of inquiry for several reasons. First, because it is not universally conserved within Grp operons, it represents a functional variation on the core Grp MCP system. Second, the GrpU sequence is highly divergent from other BMC-domain proteins, and therefore the degree of structural similarity between this BMC-domain protein and other members of the protein family is unclear.

This chapter describes a structural study of two GrpU orthologs from very different bacteria. The first of these GrpU proteins is from *Clostridiales bacterium* 1_7_47FAA, and is referred to hereafter as "*Clost*_GrpU." While little information is available about this particular species, the *Clost*_GrpU sequence is closely related to GrpU sequences from other *Clostridia* that are human pathogens, including *C. botulinum* and *C. tetani*, the causative agents of botulism and tetanus. The other GrpU sequence used in this work is from *Pectobacterium wasabiae*, a plant pathogen that causes stem rot[151]. This sequence is hereafter referred to as "*Pecwa*_GrpU." Our work combines X-ray crystallography, absorbance spectroscopy, and molecular modeling to demonstrate that the GrpU shell protein is a novel type of BMC-domain metalloprotein that binds an iron-sulfur (Fe-S) cluster using a conserved, but uniquely polymorphic, binding motif that has not been observed before. This study represents an important step

towards a complete understanding of how BMC-domain shell proteins function within the context of the Grp MCP.

4.2 Results

4.2.1 Spectroscopic evidence of Fe-S clusters in GrpU

Solutions of purified GrpU proteins appear brownish in color, and have spectral features consistent with the presence of an iron-sulfur cluster. Absorption spectra of purified GrpU solutions reveal broad peaks near 420nm (Figure 4.2). Specifically, the absorption maxima of these peaks occur at 411nm for *Clost*_GrpU, and at 415nm for *Pecwa*_GrpU. Absorbance in this spectral region is known to result from iron-sulfur charge-transfer bands[1]. We note that similar absorbance features have been observed for the PduT BMC protein[31], which were demonstrated (by electron paramagnetic resonance spectroscopy) to result from the presence of a $[4Fe\text{-}4S]^{+}$ cluster. Based on the similarities of the Grp and Pdu MCP systems, it is likely that GrpU and PduT coordinate similar metal clusters. Furthermore, we observed that these absorbance features disappeared after storage of the samples for several days at 4°C, indicating that the Fe-S cluster is somewhat labile and prone to dissociation upon protein oxidation.

4.2.2 GrpU structure determination

We determined X-ray crystal structures of two GrpU homologs, *Clost*_GrpU and *Pecwa*_GrpU, at 2.5Å and 2.8Å resolution respectively. In both cases, structure determination presented considerable difficulty due to high symmetry and high Wilson B-factors for the X-ray data. Despite these difficulties, solutions were found in space group $P2_13$ for *Clost*_GrpU and space group $H3$ for *Pecwa*_GrpU. In both crystal structures, GrpU forms homohexamers which sit upon 3-fold crystallographic symmetry axes. Additionally, 2-fold non-crystallographic symmetry axes lie perpendicular to the crystallographic 3-folds, resulting in what appear to be homododecameric arrangements of monomers with apparent 622 point group symmetry. Following molecular replacement, we performed restrained refinement of the atomic

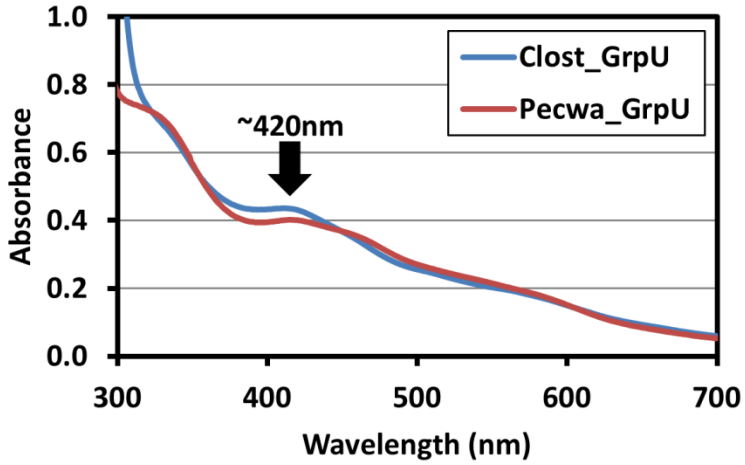


Figure 4.2. Absorption spectra of GrpU proteins. Absorbance spectra are shown for Clost_GrpU and Pecwa_GrpU, demonstrating broad peaks at 411nm and 415nm respectively. Absorption maxima in this region, near 420nm, are characteristic of Fe-S charge transfer bands.

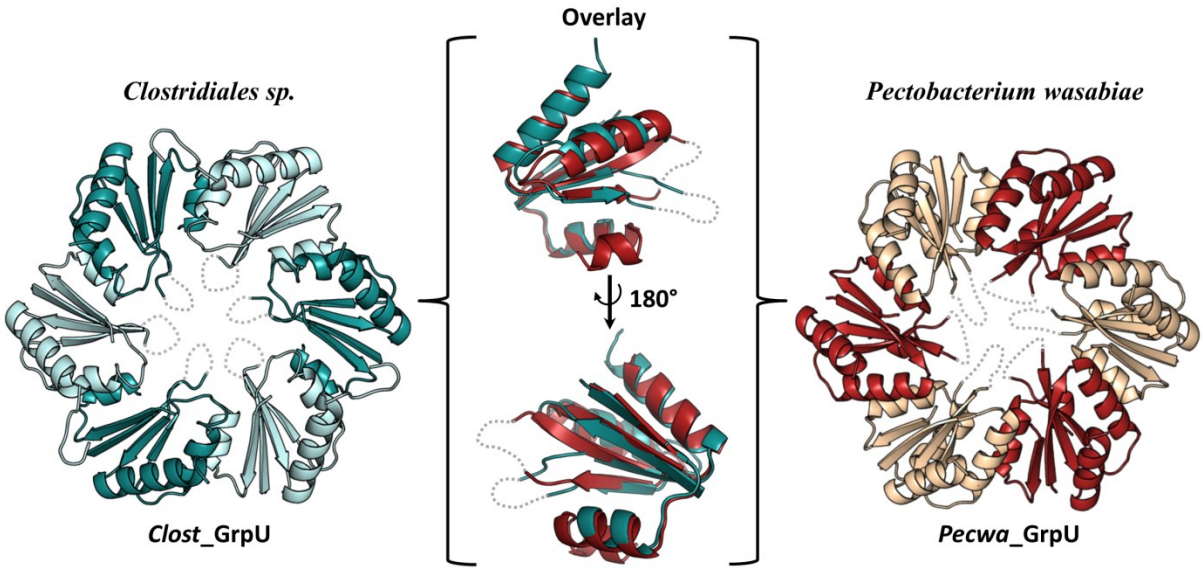


Figure 4.3. X-ray crystal structures of GrpU proteins. Hexameric assemblies of Clost_GrpU (blue, left) and Pecwa_GrpU (red, right) are shown. Monomers that are crystallographically unrelated are colored in different shades, and the disordered β3-β4 loops are represented as dashed gray lines. In the center, an overlay of Clost_GrpU and Pecwa_GrpU monomers illustrates their structural similarity.

Parameter	*Clost*_GrpU	*Pecwa*_GrpU
X-ray Wavelength	0.9791Å	0.9792Å
Nominal Resolution Range	19.84-2.50Å	60.97-2.79Å
Unit Cell	a=b=c=130.10 α=β=γ=90°	a=b=117.85; c=76.02 α=β=90°; γ=120°
Space Group	$P2_13$	$H3$
Unique Reflections	25,255	9,715
Multiplicity	4.1	12.9
Completeness	98.3% (99.8%)	99.5% (97.5%)
$\langle I/\sigma I \rangle$	17.55 (1.06)	10.2 (2.8)
Wilson B-factor	80.94Å^2	94.70Å^2
$CC_{½}$ [86]	0.999 (0.446)	0.997 (0.897)
CC^* [86]	1.00 (0.768)	0.999 (0.972)
CC_{work} [86]	0.958 (0.664)	0.925 (0.759)
CC_{free} [86]	0.945 (0.650)	0.948 (0.461)
R_{work} [a]	0.1901	0.2154
R_{free} [a]	0.2207	0.2671
No. Atoms	2536	2637
Protein Residues	302	317
Solvent Molecules	80	4
Average Atomic B-factor	87.07Å^2	76.42Å^2
RMSD (bonds)	0.01Å	0.01Å
RMSD (angles)	1.26°	1.60°
Ramachandran Plot: Favored Allowed Outliers	 99.01% 0.99% 0.00%	 97.48% 2.52% 0.00%
Molprobity Clashscore[3]	4.07	5.20

Table 4.1 Diffraction data and refinement statistics for GrpU structures.

Values in parenthesis reflect the highest resolution shell. *R*-factors are calculated with riding hydrogens present in the model.

[a] R_{work} and R_{free} are given by the following equation, computed for the working and test sets of reflections respectively: $R = \frac{\sum_{hkl} ||F_o| - |F_c||}{\sum_{hkl} |F_o|}$

coordinates, which converged to yield final models that agreed reasonably well with the observed X-ray data and had excellent molecular geometry (Table 4.1).

In general, the *Clost*_GrpU and *Pecwa*_GrpU structures have similar features (Figure 4.3). Likelihood-weighted electron density maps (*2mFo-DFc* and *mFo-DFc*) used to generate our structural models were somewhat noisy, likely due to the large proportion of disordered polypeptide in the crystal lattice (Figure 4.4). Nevertheless, we were able to reliably model residues 1-19, 32-64, and 70-102 of *Clost*_GrpU, and residues 2-67 and 72-94 of *Pecwa*_GrpU. Unfortunately, there is no electron density corresponding to the presumptive Fe-S ligands in either electron density map, likely because the Fe-S clusters in GrpU are unstable under the crystallization conditions over the time course of the experiment (approximately 1 week), even in an anaerobic environment. The *Clost*_GrpU and *Pecwa*_GrpU sequences are 40% identical, and consequently our two structures have a low coordinate RMSD; any of the four independent protein molecules in our *Clost*_GrpU structure can be superimposed upon any of the four molecules in our *Pecwa*_GrpU structure with coordinate RMSDs of 1-2Å calculated on Cα atoms. GrpU molecules that are part of the same hexamer, but are crystallographically independent, can be superimposed with coordinate RMSDs of less than 1Å.

4.2.3 Comparison of GrpU with other BMC domain proteins

Although GrpU has a highly divergent sequence, its structure is quite similar to other BMC proteins. Our structures of GrpU reveal that the polypeptide assumes a typical BMC domain fold, with the "circularly permuted" topology previously observed in a number of other BMC domain proteins (Figure 4.3)[27,39,103]. Furthermore, GrpU monomers oligomerize to form homohexamers, consistent with previous knowledge about BMC protein assembly[23–25]. Interestingly, GrpU structures have an average RMSD of 2.1Å (calculated on Cα atoms), when compared to a representative set of 17 other BMC domain protein structures from the PDB, but the average sequence identities between these proteins is only 24%, and the alignments typically have poor coverage due to sequence gaps and circular permutation. Furthermore, a BLAST search[152] against sequences in the Protein Data Bank using a GrpU query sequence does not find

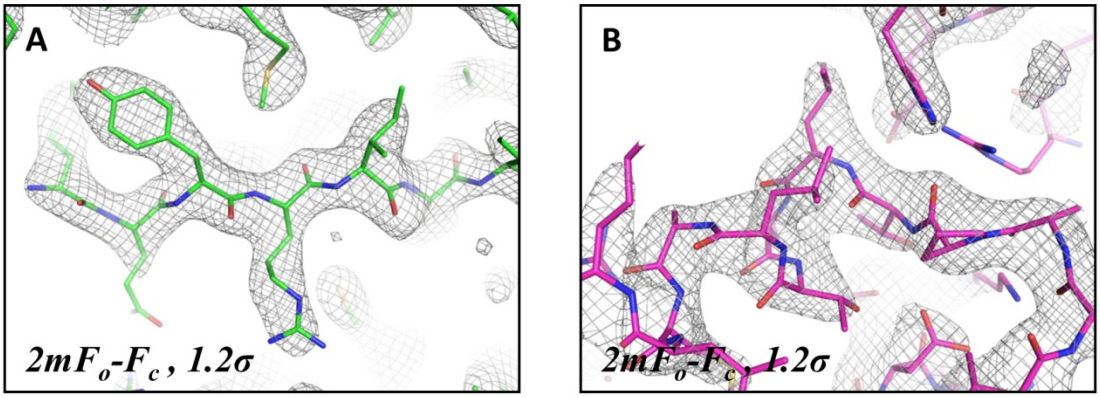


Figure 4.4. Representative electron density. Representative images of σA-weighted (2mFo-DFc) electron density maps for Clost_GrpU (A) and Pecwa_GrpU (B), also showing the final models.

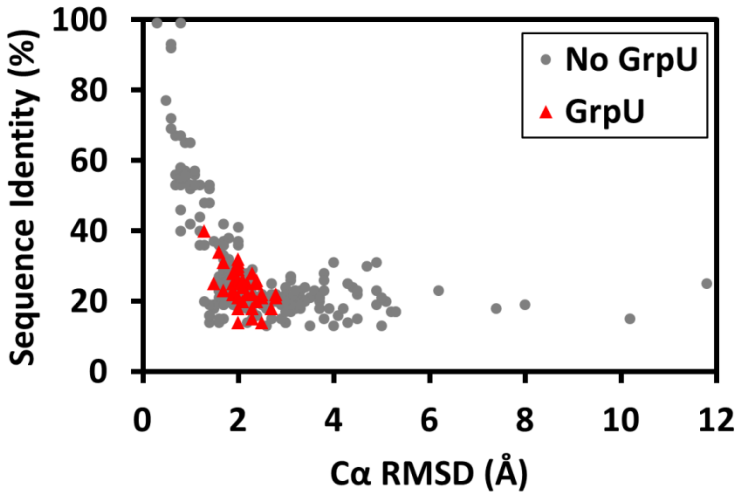


Figure 4.5. Comparison of BMC domain sequences and structures. A plot of sequence identity vs. Cα RMSD for pairwise alignment of 24 BMC-domain proteins, including Clost_GrpU and Pecwa_GrpU, demonstrates that GrpU proteins are structurally similar to the rest of the BMC superfamily, despite having highly divergent sequences. Comparisons involving GrpU are shown in red. Note that the pairwise sequence identities reported tend to overestimate the degree of similarity between proteins when there are substantial gaps in alignment, as is the case for many of the GpuU alignments.

a single hit from the BMC domain family with a significant E-value, indicating that the GrpU sequence is highly divergent from other members of the BMC protein family. A plot of Cα RMSD versus sequence identity for the pairwise comparison of 19 BMC-domain structures, including *Clost*_GrpU and *Pecwa*_GrpU, illustrates that while GrpU sequences have low sequence similarity to other BMC-domain proteins their structures are not particularly different (Figure 4.5).

4.2.4 Disorder in GrpU crystal structures

A striking feature of the GrpU structure is the large percentage of disordered residues, including the loops that occupy the centers of the individual hexamers (Figure 4.3). Residues 65-69 in each of the four chains within the asymmetric unit of our *Clost*_GrpU structure, which occupy the central pore region of the hexamer, lack interpretable electron density. Likewise, in all four chains of the asymmetric unit of our *Pecwa*_GrpU structure, residues 68-71 at the centers of the hexamers also lack interpretable electron density. In addition to the pore residues described above, our GrpU structures suffer from poor or absent electron density for a segment of residues that lie on the luminal side of the hexamer. As a result, our structures have additional unmodeled segments corresponding to residues 20-31 in *Clost*_GrpU and 29-32 (chains A and D only) in *Pecwa*_GrpU.

4.2.5 A conserved sequence motif in GrpU

The disordered loops that occupy the centers of the GrpU hexamers contain a conserved sequence motif. We performed a multiple sequence alignment of 15 non-redundant GrpU sequences, taken from Grp operons analyzed by Jorda, et al. in their comparative genomic study[10]. We observed that within this set of representative GrpU sequences, there is a perfectly conserved GXCPQ sequence motif (Figure 4.6a). By mapping the position of the conserved GXCPQ motif onto our structural models of GrpU, we determined that this motif is located in the loop connecting β-strands 3 and 4. In the context of the GrpU hexamer, six copies of this loop, one per monomer, protrude toward the central pore region of the

oligomer. In both of our GrpU crystal structures, this loop region lacks interpretable electron density and appears to be disordered, as described above.

4.2.6 Identification of cysteine residues that bind the Fe-S ligand in GrpU

Site-directed mutagenesis confirms that the conserved cysteine in the GXCPQ motif (Cys67 in *Clost*_GrpU and Cys69 in *Pecwa*_GrpU) coordinates the putative iron-sulfur cluster. We systematically mutated each of the cysteine residues in the sequence of *Clost*_GrpU to serine, to determine the effects of these mutations on the ability of GrpU to bind to its Fe-S ligand. We immediately observed that the purified C67S mutant lacked the brownish appearance characteristic of the wild-type protein (Figure 4.6b). Cysteine 67 is the conserved cysteine residue found in the GXCPQ motif of *Clost*_GrpU. In contrast, the other cysteine mutants, C18S and C47S, retained the brownish color seen for the wild-type protein. Absorbance spectra of these mutants confirmed our visual observation, showing that the C18S and C47S mutants absorb light at 411nm, while the C67S mutant does not (Figure 4.6c). This result demonstrates that in *Clost*_GrpU, Cys67 is required for Fe-S cluster binding. Furthermore, we note that the *Pecwa*_GrpU sequence contains only one cysteine (C69), which is found in the GXCPQ motif, and therefore the Fe-S cluster must be ligated by this residue.

4.2.7 Computational modeling of metal-bound GrpU

Using our crystal structure of *Clost*_GrpU as a starting point, we were able to prepare a computational model of a *Clost*_GrpU hexamer with a cubic 4Fe-4S cluster bound at its central pore. We selected the 4Fe-4S cluster over other types of Fe-S clusters based on the probable functional analogy between GrpU and PduT, a BMC domain protein from the Pdu MCP that has been shown to coordinate a 4Fe-4S cluster[31]. Our model demonstrates that a GrpU hexamer is indeed capable of accommodating its presumptive metal cluster at the proposed binding site. The computational procedure allowed us to model the pore loops that contain the ligand-binding residues in the context of a complete hexamer, without substantially altering the portions of the structure that were well-resolved in the crystal structure. Our

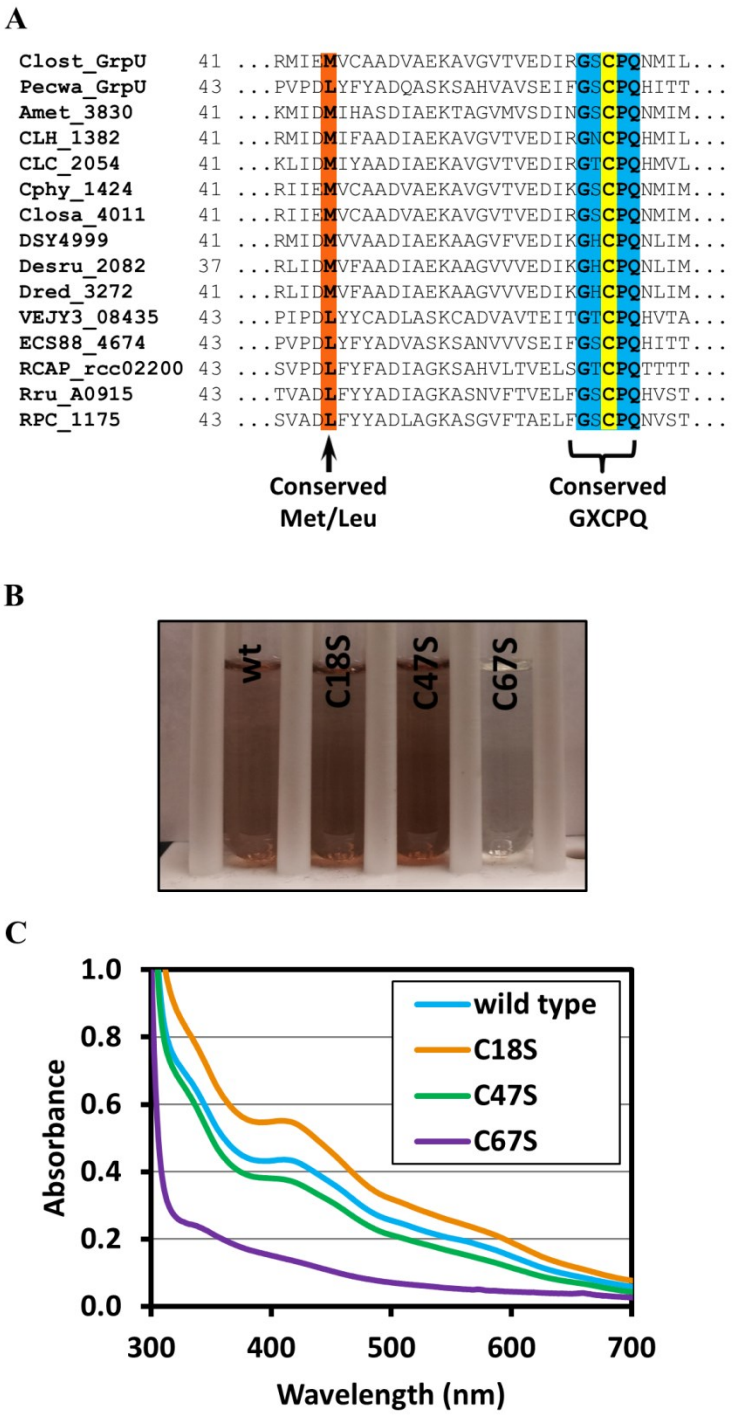


Figure 4.6. A conserved cysteine coordinates the Fe-S cluster. A sequence alignment of 15 GrpU orthologs (A) shows perfect conservation of a GXCPQ sequence motif, and a conserved methionine or leucine at another sequence position. Solutions of the purified C67S mutant lack the characteristic brownish color (B), and absorption spectra of Clost_GrpU and three point mutants (C18S, C47S, C67S) demonstrate that only the C67S mutation eliminates the signal attributed to the Fe-S cluster (C).

modeling protocol began with a 6-fold symmetric model of a hexamer, and subsequently introduced ligand restraints between alternating Cys67 residues, while simultaneously relaxing the oligomeric symmetry from 6-fold to 3-fold. The final, energy-minimized model of the *Clost*_GrpU hexamer with the cubic 4Fe-4S cluster showed signifcant differences in the conformations of the two independent β3-β4 loops. Breakage of 6-fold oligomeric symmetry to give a 3-fold symmetric homohexamer accompanies binding of a cubic Fe-S cluster in GrpU (Figure 4.7), largely as a result of steric constraints. In our model of metal-bound GrpU, the Fe-S cluster is positioned very low in the pore, making the luminal side of the GrpU hexamer appear as a flat surface, in contrast to the concave feature that is typically found on the luminal side of most BMC hexamers. Additionally, because of its position close to the luminal side of the hexamer, the Fe-S cluster appears to lie at the bottom of a funnel-shaped cavity when viewed from the cytosolic face of the hexamer.

4.2.8 A GrpU-like protein from a Pdu MCP

Somewhat surprisingly, an analysis of Pdu operons (i.e. those that encode B12-dependent propanediol-utilizing MCPs rather than the glycyl-radical type) revealed that at least one organism, Anaerobaculum mobile, contains a shell protein sequence closely resembling GrpU. This operon can be identified as a true Pdu operon based on the presence of a B12-dependent diol dehydratase enzyme (PduCDE), and we note that in addition to the GrpU-like shell protein, this operon contains the normal repertoire of BMC-domain shell proteins typically found in the Pdu MCP (PduABKJTU)[28]. The GrpU-like sequence from this Pdu operon also bears some similarity to the shell protein PduU, including a small β-barrel formed by extended N-termini; however, a phylogenetic tree comparing this sequence to known PduU and GrpU sequences confirms that this sequence is more closely related to GrpU homologs than it is to PduU homologs (Figure 4.8a). Furthermore, sequence alignment demonstrates that this polypeptide contains the conserved GXCPQ sequence motif characteristic of GrpU proteins. Using our *Clost*_GrpU crystal structure as a template, we were able to create a computational model of this GrpU-like sequence as well (Figure 4.8b). This model superimposes upon our model of metal-bound *Clost*_GrpU with a

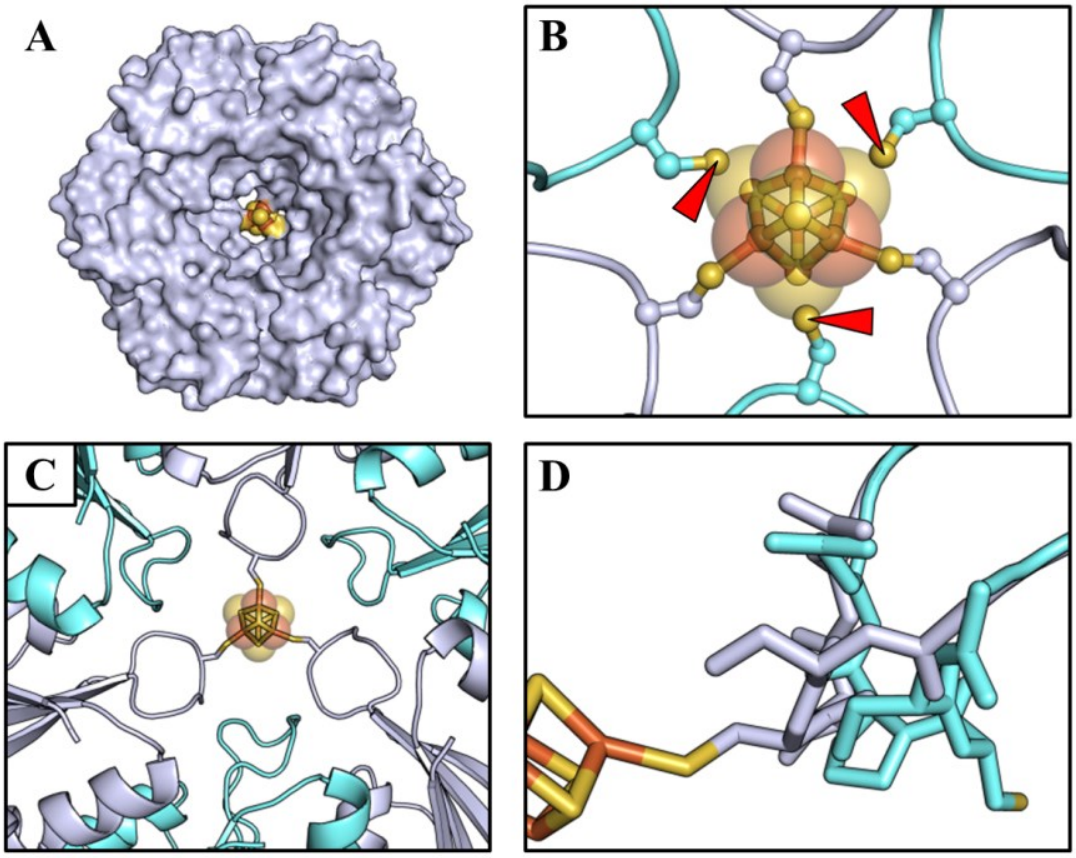


Figure 4.7. Computational modeling of Clost_GrpU bound to a 4Fe-4S cluster. Using our crystal structure as a starting point, we prepared a computational model of Clost_GrpU bound to a 4Fe-4S cluster (A). A model with 6-fold symmetry was judged to be impossible, based on a steric clash, involving the Fe-S cluster and the Sγ-atoms of alternating Cys67 residues at the center of the hexamer (red arrowheads, B). Our model demonstrates that this clash can be relieved if the symmetry of the hexamer is relaxed from 6-fold to 3-fold (C) via alternating conformations of the conserved GXCPQ motif, which is shown in sticks (D).

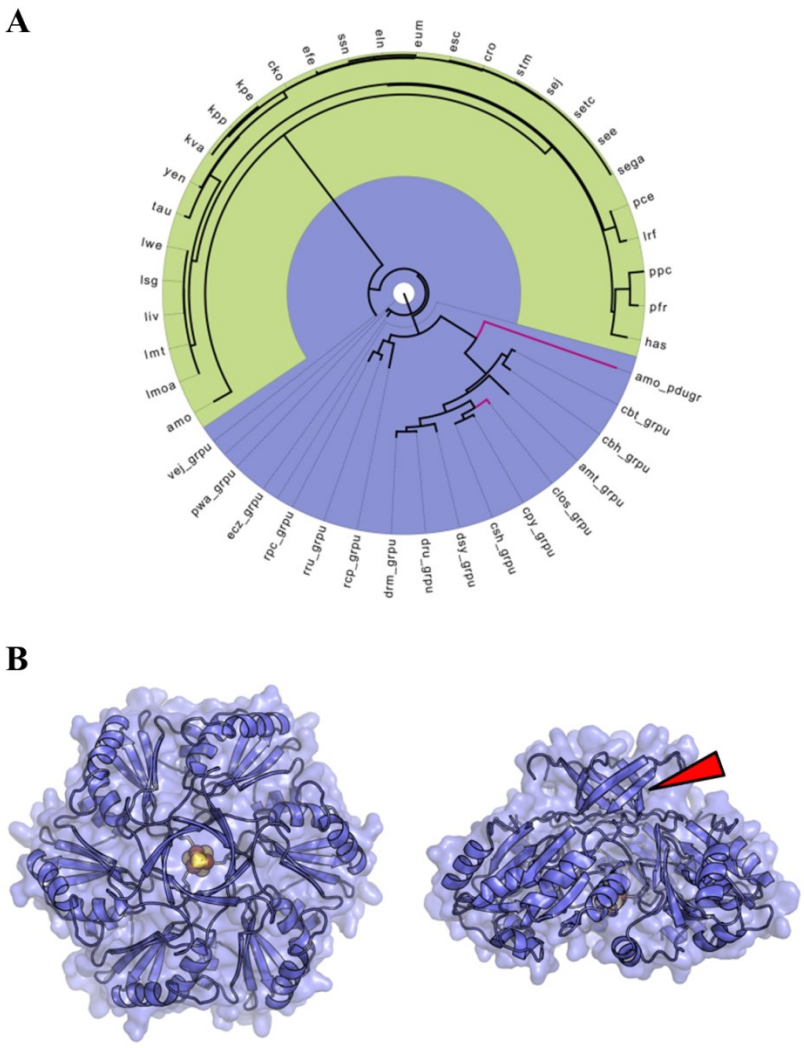


Figure 4.8. A GrpU ortholog from a Pdu operon. A phylogenetic tree (A) demonstrates the relationship of GrpU (blue background) and PduU (green background) protein sequences. Pink lines highlight Clost_GrpU and another GrpU homolog that is found in a true Pdu-type operon, but clearly clusters with GrpU sequences. A computational model of this sequence is shown (B), revealing that although clearly a GrpU homolog, this protein has notable structural similarities to PduU, including a small β-barrel atop the hexamer (red arrowhead).

relatively low coordinate RMSD (2.5Å, calculated on Cα atoms for comparison of two monomers), and demonstrates that this GrpU-like protein also has the capacity to bind a Fe-S cluster at the central pore of an assembled hexamer. Interestingly, this GrpU-like sequence has a short N-terminal extension that forms a small 6-stranded β-barrel in the context of the hexamer (Figure 4.8b), similar to the PduU and EutS shell proteins[27,103,153]. This small β-barrel appears to occlude the central Fe-S cluster from the cytosolic side of the MCP shell.

4.3 Discussion

Here we provide the first structural insight into the homohexameric BMC domain shell proteins from the recently-discovered Grp MCP. BMC domain proteins are the major component of MCP shells, and previous studies of other MCP systems have highlighted the mechanistic understanding that can be gained from their structural elucidation[22,27,29]. Towards the goal of extracting similar types of insights into glycyl-radical type (Grp) MCPs, we have determined X-ray crystal structures of two GrpU homologs (Figure 4.3). One of these homologs is taken from an organism whose genomic sequence is annotated only as *Clostridiales bacterium* 1_7_47FAA. This sequence is 82% identical to the GrpU sequence from *Clostridium saccharolyticum*, and GrpU sequences with high levels of sequence similarity (54-67% identity) are found in species of *Clostridia* that are human pathogens, including *Clostridium botulinum* and *Clostridium tetani*. The other GrpU homolog used in this study originates from the plant pathogen *Pectobacterium wasabiae*. It is interesting to note that although the genera *Clostridium* and *Pectobacterium* both contain a number of pathogenic organisms with Grp operons, they are evolutionarily distant from one another; *Pectobacteria* are gram-negative Proteobacteria, while Clostridia are gram-positive Firmicutes. The conservation of the Grp MCP in such distantly related pathogens suggests its potential importance in infection or other host-pathogen interactions.

4.3.1 Structural overview of GrpU

The structure of GrpU is quite similar to other BMC domain proteins, despite its highly divergent sequence. One feature of GrpU that initially attracted our attention was its apparent dissimilarity to other BMC-domain proteins. BLAST searches do not identify any alignments between GrpU sequences and other members of the BMC domain superfamily as having significant E-values, which we initially interpreted as evidence that GrpU structures might be quite different from other known BMC domain structures. Surprisingly, our X-ray crystal structures of *Clost*_GrpU and *Pecwa*_GrpU reveal that these proteins exhibit only minor, though important, structural variation relative to the other members of the BMC superfamily. In order to quantify the similarities between GrpU and divergent members of the BMC superfamily, we calculated sequence identities and coordinate RMSDs (calculated on Cα atoms) for pairwise comparisons of 24 BMC proteins of known structure, including *Clost*_GrpU and *Pecwa*_GrpU. A plot of sequence identity vs. Cα RMSD for the 276 pairwise comparisons demonstrates that while GrpU sequences typically have low sequence identities when compared to other BMC proteins, they also tend to have low Cα RMSDs for the corresponding structural alignments (Figure 4.5), confirming that, although GrpU proteins have relatively divergent sequences, they are structurally similar to other members of the BMC superfamily.

Although GrpU appears to adopt a typical BMC domain fold, our X-ray crystal structures reveal that GrpU also has some unique structural features (Figure 4.3). Notably, GrpU seems to have a large percentage of unstructured residues relative to other members of the BMC domain superfamily. For example, in most single-domain BMC proteins, the β3-β4 loop segment is well-defined and imparts the respective hexamer with a narrow pore through its center. These narrow pores have been hypothesized to regulate the diffusion of substrates and products through the MCP shell[25,27,29,127]. In contrast, both of our GrpU crystal structures reveal disordered β3-β4 loop segments protruding toward the centers of the hexamers, resulting in an apparent void at the center of the oligomer (Figure 4.3). As discussed subsequently, the intrinsic flexibility of these loop segments is likely to facilitate the Fe-S cluster-binding activity of GrpU. In addition to disorder at the center of the hexamers, GrpU monomers in our crystal

structures also exhibit disordered polypeptide segments connecting helix-αA and the β2-strand on the luminal side of the hexamer (Figure 4.3). Other studies of BMC-domain proteins have demonstrated that this surface is a likely interaction point between the MCP shell and the internalized enzymes[24,56]; therefore, we suggest that these unstructured loops might be involved in protein-protein interactions that enable the assembly of the complete MCP.

In both of our GrpU crystal structures, hexamers appear to pack within the lattice as dodecameric assemblies with local D6 symmetry. The possibility that MCP shells may consist of a double layer of oppositely faced BMC-domain shell proteins has been raised in previous studies[40,128,154]. The dodecameric assemblies we observed in crystals of *Pecwa*_GrpU and *Clost*_GrpU are distinct from each other, and neither appears to be arranged in a way that would be compatible with formation of extended double-layer sheets, though this does not discount the possible relevance of local dodecameric assemblies.

4.3.2 Identification of a Fe-S cluster binding site in GrpU

GrpU likely binds a cubic iron-sulfur cluster at its central pore, similar to the PduT shell protein from the Pdu MCP. Absorbance spectra of GrpU shell protein samples show features that are consistent with the presence of an iron-sulfur cluster (Figure 4.2), particularly a broad absorbance peak near 420nm[1]. The absorption spectra we observed for GrpU samples are nearly identical to those observed by Parsons, et al. from samples of the PduT BMC domain shell protein[31]. Additional structural and spectroscopic characterization of PduT revealed that, in contrast to GrpU, it is a tandem BMC domain polypeptide, which assembles into a pseudohexameric trimer and binds a cubic 4Fe-4S cluster at its central pore[29,31,35]. In order to bind the Fe-S cluster at the center of a symmetric trimer, each of the three PduT monomers contributes one cysteine, whose sulfur atoms ligate the Fe-S cluster at the three-fold axis of oligomeric symmetry. In the GrpU shell protein, mutation of a single, perfectly-conserved cysteine residue ablates the spectroscopic signal assigned to the Fe-S cluster (Figure 4.6b,c), indicating that GrpU binds its Fe-S ligand in a manner similar to PduT, in which multiple monomers each contribute a single cysteine residue to a binding site that lies at the center of a homo-oligomer. The key difference is that in GrpU, only three

of the six chemically identical subunits contribute their cysteines to ligate the metal cluster. Mapping of these critical cysteine residues onto our structural models confirms their location within loops (which appear disordered in the absence of the bound metal cluster) near the oligomeric symmetry axis. Although our absorbance spectra cannot distinguish the specific type of Fe-S cluster that is present in GrpU, we suggest that it is likely to be a cubic cluster (3Fe-4S or 4Fe-4S) based on analogy with PduT.

4.3.3 Structural features of the ligand binding site

Disorder at the center of the GrpU hexamer, in the β3-β4 loop, breaks the 6-fold oligomeric symmetry, as required for binding of the Fe-S cluster at this position. There is no known example of a biological Fe-S cluster with 6-fold symmetry, and to our knowledge there have been no previous observations of any type of Fe-S cluster present at a coordination site with 6-fold symmetry. However, because the cubic Fe-S cluster does have 3-fold symmetry, which is also contained within the 6-fold oligomeric symmetry of GrpU, we propose that the GrpU hexamer binds a 3-fold symmetric metal cluster in one of two degenerate orientations using a coordination site where the 6-fold symmetry is decisively broken. In order to explore the feasibility of such an arrangement, we performed a computational modeling exercise. Our model of *Clost*_GrpU with a cubic 4Fe-4S cluster revealed that a ligation site with cysteine residues in a 6-fold-symmetric arrangement is not possible, because the Sγ atoms of the cysteine residues that do not participate in the Fe-Sγ bonds would clash with the sulfur atoms that are part of the Fe-S cluster itself (Figure 4.7b). Consequently, we concluded that a GrpU hexamer cannot retain strict 6-fold oligomeric symmetry when bound to its presumptive 4Fe-4S ligand. By relaxing the oligomeric symmetry (from 6-fold to 3-fold) while constraining alternating Cys67 residues to form a 4Fe-4S coordination site, we were able to obtain a model that demonstrates the ability of a GrpU hexamer to bind its presumptive Fe-S ligand without steric overlap, allowing for a modest structural difference between alternating chains in the cyclic hexamer (Figure 4.7c,d). Furthermore, our computational model suggests that rearrangement of the conserved GXCPQ motif yields the required conformational heterogeneity (Figure 4.7d). Our crystallographic and computational results, taken together, suggest that

intrinsic disorder at the center of the GrpU hexamer in the absence of the ligand serves to alleviate a steric clash resulting from symmetry mismatch between GrpU and the cubic Fe-S cluster. Upon ligand binding, this disordered region likely adopts a much more rigid, 3-fold symmetric conformation, owing to the structural constraints imposed by the Fe-S cluster.

In GrpU, a structural feature of the β-sheet (within each monomer) appears to promote structural disorder at the center of the oligomer. In particular, three successive strands, β2, β3, and β4 have an unusual geometry that leads to conformational variation of the β3-β4 loop, which occupies the center of the oligomer. The outer two of these three strands, β2 and β3, appear to splay unusually far apart from one another at the edge of the sheet. This arrangement of β2 and β3 creates a situation in which the middle strand, β4, cannot simultaneously form hydrogen-bonds with both of its neighboring strands in the sheet, leading to two alternative conformations (Figure 4.9a,b) and implying the existence of dynamic disorder in the absence of the Fe-S cluster. Evidence for both of these conformations can be found in our *Clost*_GrpU model and in corresponding electron density maps. After modeling the major conformation of the β4 strand in *Clost*_GrpU, a strong positive density feature in the mFo-DFc difference electron density map indicated a second, minor conformation (Figure 4.9c). This difference electron density was present near all four crystallographically independent molecules in the *Clost*_GrpU structure. We were able to successfully model the second conformation in one of the four crystallographically independent molecules, although our attempt to model a second conformation of the β4 strand in the other three *Clost*_GrpU molecules led to a drop in Rwork, but not in Rfree, and consequently these conformers were omitted from the final coordinates. Additionally, in *Clost*_GrpU a methionine residue (Met45) adjacent to the β4 strand also occupies two alternative rotamers in a fashion that correlates with the alternative positions of the strand (Figure 4.9c). Failure to model either of these alternate side-chain conformations results in strong positive peaks in the mFo-DFc difference electron density map. We note that this methionine residue is well-conserved in GrpU; all GrpU sequences have either methionine or leucine at this position, indicating that motion of this residue may also play an important role in promoting structural heterogeneity at the metal binding site.

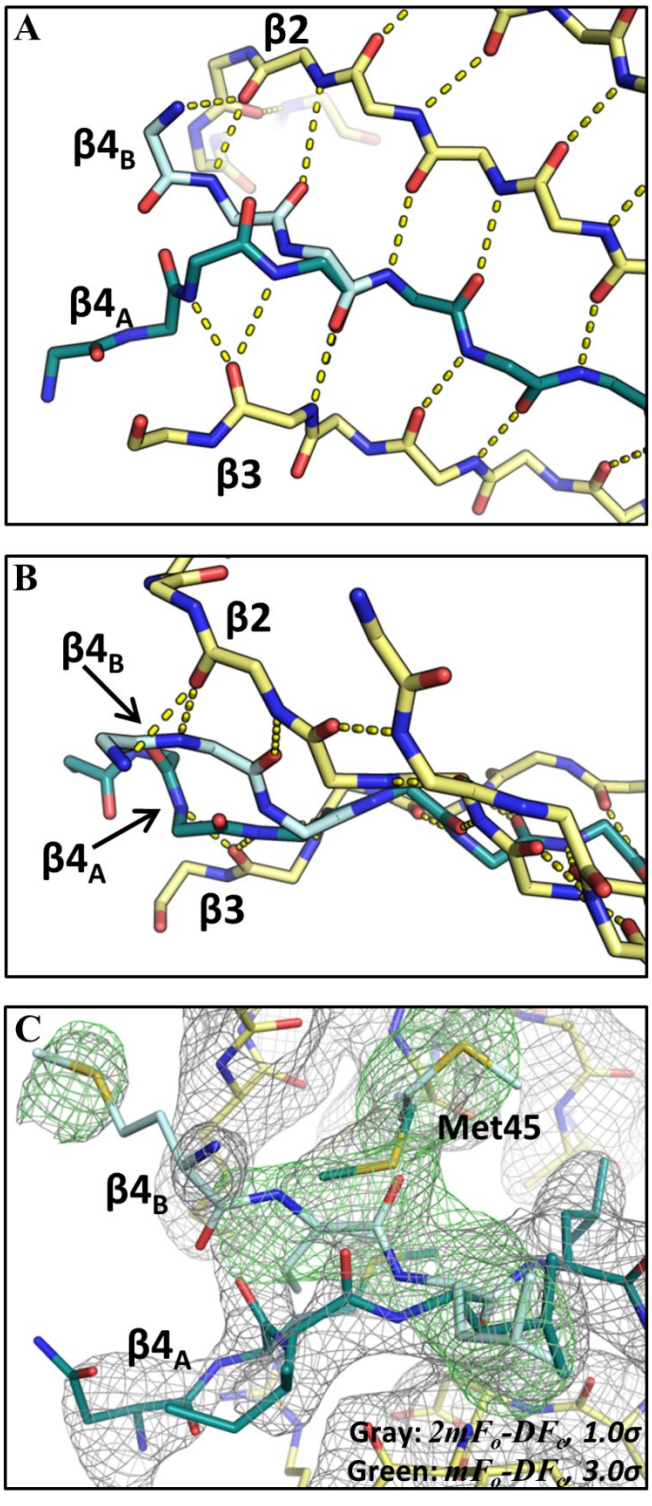


Figure 4.9. Multiple conformations of the β3-β4 loop. A unique splaying of the β2 and β3 strands in GrpU leads to a situation in which the strand between them, β4, cannot simultaneously form H-bonds with both of its partners, leading to two alternative conformations (A,B – side and top views respectively, side chains omitted for clarity). Positive electron density features in both 2mFo-DFc and mFo-DFc electron density maps (C) indicate that the β4 strand and Met45 occupy alternate conformations.

4.3.4 Insights into GrpU function

Although we have provided detailed structural characterization of GrpU, our work reveals little information about the function of this shell protein within the Grp MCP. Based on the presence of the Fe-S ligand in GrpU, we have developed several hypotheses about its function. Our first hypothesis is that GrpU facilitates electron transport through the shell of the Grp MCP in order to help generate the glycyl-radical at the active site of the diol dehydratase enzyme. The glycyl-radical within the diol dehydratase enzyme is generated with the help of a secondary activase enzyme, which requires S-adenosylmethionine and a single electron donor[10,148,155]. It is possible that the single electron required for this reaction is generated by oxidation of a cytoplasmic reductant and then transported through the MCP shell via GrpU. An alternative hypothesis is that GrpU functions to transport intact Fe-S clusters through the MCP shell. The activase in the Grp system requires a 4Fe-4S cluster for its enzymatic activity[148], and it is possible that GrpU transports 4Fe-4S clusters to supply the activase enzyme with fresh cofactor. GrpU might be a well-suited transport protein for a cubic Fe-S cluster, because it would bind the cluster with only three cysteine ligands. This might lead to a weaker affinity for the Fe-S cluster relative to another protein that binds the same cluster with four cysteine residues, leading to an affinity gradient compatible with directional transport.

Genetic variability between Grp operons creates further confusion about the function of GrpU. That not every Grp operon contains GrpU indicates that its specialized function may represent a modification to the core Grp MCP[10]. On the other hand, there are Grp operons that contain GrpU alongside other presumptive Fe-S-containing BMC shell proteins, namely orthologs of PduT, indicating some functional redundancy in these systems. One such Grp operon that contains GrpU and a PduT homolog is from the organism *Desulfitobacterium halfniense*. Additionally, we have identified at least one true Pdu operon (i.e. encoding PduCDE enzyme), from *Anaerobaculum mobile*, that contains a GrpU-like shell protein. We have shown, by creation of a phylogenetic tree, that this GrpU-like sequence is more similar to GrpU than it is to PduU, another closely-related shell protein from the Pdu MCP system (Figure 4.8a). A computational modeling exercise illustrates that this GrpU-like sequence is similar to

GrpU in its capacity to bind a cubic Fe-S cluster, although it appears to also have a small β-barrel structure on the cytosolic face of the hexamer like the PduU shell protein (Figure 4.8b). The anomalous distribution of GrpU across MCP operons is evidence that GrpU plays a complicated role in these MCP systems. Hopefully, future studies of GrpU will integrate our structural characterization with the results of genetic and biochemical assays designed to unravel the function of this unusual shell protein.

4.4 Conclusions

The results presented here provide unexpected evidence that a previously uncharacterized GrpU microcompartment shell protein binds a metal cluster at a unique site of coordination. We used absorption spectroscopy to demonstrate the presence of a Fe-S cluster in GrpU, and confirmed that a conserved sequence motif ligates the cluster at the center of a GrpU hexamer. Two X-ray crystal structures of GrpU proteins at moderate resolutions reveal a disordered binding motif in the absence of the Fe-S cluster. A subsequent computational study of GrpU demonstrated that 6-fold oligomeric symmetry must break down in order to accommodate the presumptive 4Fe-4S cluster. Our work with GrpU provides the first insight into the BMC domain shell proteins of the Grp MCP. Based on the chemistry associated with the Grp MCP, it is likely that the presence of the Fe-S cluster in this shell protein facilitates the transport of reducing equivalents across the shell, although more experiments will be required to develop a full understanding of GrpU function.

In addition to being the first BMC domain homohexamer from the Grp MCP to be experimentally characterized, GrpU represents a new variation on BMC domain metalloproteins. The PduT shell protein is another known Fe-S binding BMC domain protein. The functions of GrpU and PduT are likely similar, yet the two proteins have dramatic structural differences. GrpU and PduT have different topology (by virtue of circular permutation), and GrpU is a single-domain homohexamer, while PduT is a tandem-domain homotrimer. This difference in oligomerization reflects a difference in how these two proteins break the 6-fold oligomeric symmetry typical of the BMC domain proteins in order to accommodate their

Fe-S ligands; PduT breaks 6-fold symmetry by domain duplication and fusion (i.e. sequence polymorphism), while 6-fold symmetry is broken by conformational polymorphism in GrpU. As far as we are aware, the symmetry-breaking mechanism by which GrpU binds its Fe-S ligand is indeed unique among metalloproteins studied to date.

4.5 Materials & Methods

4.5.1 Cloning of recombinant GrpU proteins

The codon-optimized DNA sequences corresponding to amino acids 1-101 of the GrpU protein from *Clostridiales bacterium* 1_7_47FAA (*Clost*_GrpU) and amino acids 1-99 of the GrpU protein from *Pectobacterium wasabiae* (*Pecwa*_GrpU) were designed using the *DNAWORKS* web server[156]. Segments of linear DNA containing the designed sequences were independently assembled from small oligonucleotides (Integrated DNA Technologies), as suggested by *DNAWORKS*, using recursive PCR[157]. The PCR products were digested with NdeI and XhoI restriction enzymes (New England Biolabs), purified by agarose gel electrophoresis, and ligated into the pET-22b+ expression vector, which had also been digested with NdeI and XhoI restriction enzymes, using Quick Ligase (New England Biolabs). The restriction sites selected for incorporation of the GrpU sequences into pET-22b+ append a hexahistadine purification tag (sequence: -LEHHHHHH) at the c-terminus of the polypeptides, and place the GrpU genes under the control of the T7 promoter region. Point mutations (C18S, C47S, and C67S) were introduced into the *Clostridiales* GrpU sequence using the QuickChange method (Stratagene). All DNA constructs were verified by dideoxy chain termination sequencing[109].

4.5.2 Protein overexpression and purification

Expression plasmids (pET-22b+) bearing the GrpU genes were independently transformed into *Escherichia coli* BL21 (DE3) chemically competent cells (New England Biolabs) for protein expression. For expression of both protein constructs (*Clost*_GrpU and *Pecwa*_GrpU), we added 1mM isopropyl-β-

D-thiogalactopyranoside to cell cultures grown in selective Luria-Bertani (LB) broth during the exponential phase of growth. Growth media was supplemented with 1mM L-cysteine and 50mg/L ferric ammonium citrate when protein overexpression was induced, in order to supply the cells with additional iron and sulfur. After four hours of protein overexpression at 37°C, the cells were harvested by centrifugation for 15 minutes at 5,000x*g* and stored at -20°C.

Cell pellets containing overexpressed *Clost*_GrpU protein were resuspended in buffer containing 20mM Tris, 300mM sodium chloride, and 0.03% polysorbate 20 at pH 8.0 with a protease inhibitor additive (Roche), 10mM MgCl2, 1mg/mL lysozyme, and 100units/mL of both DNase and RNase, and then lysed using a high-pressure emulsifier (EmulsiFlex C3, Avestin). We clarified the cell lysate by centrifugation at 30,000x*g* for 30 minutes, and then used a HisTrap nickel affinity column (GE Healthcare) to purify the protein from clarified lysate. The bound protein was eluted with lysis buffer containing 300mM imidazole, and then dialyzed against a buffer containing 20mM Tris and 50mM sodium chloride at pH 8.0. The dialyzed sample was then loaded onto a HiTrapQ anion exchange column (GE Healthcare) that was pre-equilibrated with the dialysis buffer, and subsequently eluted with a linear gradient of buffer containing 20mM Tris and 1M sodium chloride at pH 8.0. Finally, the eluted protein sample was dialyzed once again against a buffer consisting of 20mM Tris and 50mM sodium chloride at pH 8.0.

*Pecwa*_GrpU protein was purified from cell pellets by first resuspending the cells in buffer containing 50mM Tris, and 300mM sodium chloride at pH=8.0 with a protease inhibitor additive (Roche). Cells were lysed by ultrasonic disruption (Vibra-Cell VCX, Sonics & Materials, Inc.) and the cell lysate was clarified by centrifugation at 30,000xg for 30 minutes. The lysate was loaded onto a HisTrap nickel affinity column and the bound protein was eluted with lysis buffer containing 350mM imidazole. Finally, the Pecwa_GrpU sample was subjected to gel filtration using a Superdex75 column (GE Healthcare), which provided a pure protein sample in buffer consisting of 18mM Tris and 100mM sodium chloride at pH=7.6.

4.5.3 Absorption spectroscopy

Absorption spectra of *Clost*_GrpU and corresponding point mutants (C18S, C47S, and C67S) were measured using a Cary-60 spectrophotometer (Agilent Technologies). We performed a wavelength scan from 250nm to 800nm using a 0.5nm step size and 0.1sec integration time. Spectra were collected immediately after protein purification and dialysis, in order to minimize loss of the Fe-S cluster upon exposure to oxygen in the atmosphere. For comparison of the wild-type protein with the various point mutants, individual spectra of the mutant constructs were scaled to the wild-type spectrum according to protein concentration.

The absorption spectrum of *Pecwa*_GrpU was measured using a Cary-300Bio spectrophotometer (Agilent Technologies). We performed a wavelength scan from 240nm to 700nm using a 1nm step size and 0.2sec integration time. Prior to measuring the absorption spectrum, we added 5mM dithiothreitol (DTT) in order to minimize loss of the Fe-S cluster due to protein oxidation.

4.5.4 Protein Crystallization

Purified *Clost*_GrpU protein was concentrated to approximately 12.5mg/mL and crystallized by hanging-drop vapor diffusion in mother liquor consisting of 2.0M ammonium sulfate, 0.1M sodium acetate at pH=4.5. Crystallization drops were prepared by mixing 1µL of protein solution with 1µL of mother liquor, sealing the drop above a 0.5mL reservoir (Hampton VDX crystallization plate), and allowing the system to equilibrate at 296K. The highest quality Clost_GrpU crystals were obtained when 10mM oxidized glutathione was added to the protein sample at least 30 minutes prior to setting up the crystallization experiment.

Purified *Pecwa*_GrpU protein was crystallized using a hanging-drop protocol similar to the one used for crystallization of *Clost*_GrpU. In the case of *Pecwa*_GrpU, the protein sample was concentrated to approximately 60mg/mL, and crystallized in mother liquor consisting of 30% 2-methyl-2,4-pentanediol, 0.025M sodium potassium phosphate at pH=5.8.

We attempted to crystallize both *Clost*_GrpU and *Pecwa*_GrpU in an anaerobic environment, in order to preserve the Fe-S clusters. Unfortunately, crystals obtained under anaerobic conditions were identical to similar crystals grown in ambient conditions, and did not provide additional information about the Fe-S ligands.

4.5.5 X-ray data collection and processing

Prior to X-ray data collection, crystals of *Clost*_GrpU were harvested, cryoprotected using a 1:1 mixture of crystallization mother liquor and 4.0M trimethylamine-N-oxide, and frozen in a liquid nitrogen cryostream at 100K. We collected single-crystal X-ray diffraction data at the Advanced Photon Source using beamline 24-ID-C equipped with a Pilatus 6M-F detector and operating at 12663eV. Crystals were maintained at cryogenic temperature (100K) throughout the course of data collection. We indexed and integrated reflections to 2.5Å resolution using *XDS*, performed scaling and merging with *XSCALE*, and converted intensities to structure factors using *XDSCONV*[112]. The free set of reflections (10%) was designated by *XDSCONV*.

For X-ray data collection from our *Pecwa*_GrpU crystals, harvested samples were cryoprotected using a mixture of 30% glycerol and 70% crystallization mother liquor and then frozen in a liquid nitrogen cryostream at 100K. We collected single-crystal X-ray diffraction data at the Advanced Photon Source using beamline 24-ID-C equipped with a Pilatus 6M-F detector and operating at 12662eV. Crystals were maintained at cryogenic temperature (100K) throughout the course of data collection. We indexed and integrated reflections to 2.8Å resolution using *DENZO*, and performed scaling and merging with *SCALEPACK*[111]. Intensities were converted to structure factors using *Ctruncate* and the free set of reflections (5%) was assigned using *Uniqueify*, both within the *CCP4* utility *sca2mtz*[116].

Information regarding data collection and processing for both *Clost*_GrpU and *Pecwa*_GrpU is provided in Table 4.1.

4.5.6 Phase calculation by molecular replacement

Initial phase determination for the *Clost*_GrpU diffraction data was performed using the *phenix.mr_rosetta* software[158], using an initial, lower-resolution (3.2Å) dataset. Fragment files and alignment information required for *phenix.mr_rosetta* were obtained from the *Robetta* and *hhpred* web servers respectively. The solution obtained by *phenix.mr_rosetta* at 3.2Å was subsequently used as a search model in a straightforward molecular replacement procedure with the 2.5Å data set, implemented using the *Phaser* software[115].

In order to find an acceptable molecular replacement solution for *Pecwa*_GrpU, we first had to expand our data from space group *H*3 to the lower-symmetry space group *P*1. Using the *P*1 data, we performed a molecular replacement search using the *Phaser* software[115], with a search model consisting of a full BMC hexamer constructed from PDB: 3IA0. Once a solution was found in *P*1, we selected four polypeptide chains corresponding to an asymmetric unit in space group *H*3, and proceeded with a model corresponding to the trigonal space group.

4.5.7 Model building and refinement

Following molecular replacement, atomic models of GrpU proteins were prepared by iterative steps of model-building and automated refinement of the model coordinates against the observed X-ray data. Model building was performed with Coot v0.7[117]. The final model of *Clost*_GrpU was refined using *BUSTER* (version 2.10.0) with automatic NCS restraints and automatic refinement of TLS parameters[159]. The final model of *Pecwa*_GrpU was generated by restrained refinement using *REFMAC* (version 5.8.0049) with automatic NCS restraints and a single TLS group for each chain[160]. The final models were deposited in the Protein Data Bank (PDB)[125] under accession codes PDB: 4OLO and PDB: 4OLP for *Clost*_GrpU and *Pecwa*_GrpU respectively. Refinement and model validation statistics are provided in Table 4.1.

4.5.8 Sequence and structural alignment of BMC-domain proteins

Pairwise sequence and structure alignments of 24 BMC-domain proteins, including both *Clost*_GrpU and *Pecwa*_GrpU, were performed using the *DALI* web server[161]. The PDB accession codes for the 22 non-GrpU proteins used in the comparison are: 3BN4, 2A1B, 2A10, 2EWH, 3H8Y, 3FCH, 3I96, 4AXI, 3I6P, 4AXJ, 3N79, 3PAC, 3NWG, 3I82, 4EDI, 3U27, 4FAY, 3NGK, 3CGI, 4HT5, 3IO0, 3IA0. These 22 BMC-domain homologs were selected in order to represent a variety of different species and MCP systems.

4.5.9 Multiple sequence alignment of GrpU homologs

For multiple sequence analysis of GrpU homologs, we used a set of 15 non-redundant GrpU sequences. Fourteen of these sequences, including the *Pecwa*_GrpU sequence, were taken from the original bioinformatics analysis of the Grp operon performed by Jorda, et al.[10], and the fifteenth sequence is that of *Clost*_GrpU. We used the *MUSCLE* software to perform the multiple sequence alignment[162].

4.5.10 Computational modeling of metal-bound GrpU

A computational model of *Clost*_GrpU bound to a cubic 4Fe-4S cluster (which was present after purification, but lost during crystallization) was prepared in several steps. First, we used *MODELLER* (version 9.10) to model the disordered loop regions for a single monomer from the *Clost*_GrpU crystal structure[163]. The model of the complete monomer was then assembled back into a hexamer by applying a 6-fold symmetry operator. This hexameric model was subjected to two cycles of structure relaxation using *Rosetta*[164], followed by symmetry-restrained relaxation using *CNS* (version 1.2)[165,166]. This hexameric model was then subjected to energy minimization allowing backbone flexibility, using *MODELLER* (version 9.10) [163]. During this energy minimization step, we constrained the hexamer to have 3-fold symmetry (thereby modeling two independent chains), and we harmonically restrained the distances of alternating cysteine Sγ atoms so that they would form an equilateral triangle with a side length of approximately 6.5Å. This distance restraint places the sulfur atoms in positions which can

accommodate ligation of an Fe-S cluster. After obtaining this 3-fold symmetric model, we attached the Fe-S group to the corresponding cysteine residues, and performed one last energy minimization step with *CNS* (version 1.2)[165,166] in the presence of the ligated Fe-S cluster.

In addition to the model of ligand-bound *Clost*_GrpU, we also prepared a model of a GrpU-like protein from the phylogenetically remote bacterium *Anaerobaculum mobile*, also bound to a cubic 4Fe-4S cluster. This model was prepared using the same protocol that was used for *Clost*_GrpU, except in this case the crystal structure was not known, so we first prepared a homology model using *MODELLER* with a template (PDB: 4AXI, 23% sequence identity) identified by *MUSTER*[167].

The computational models were judged to be plausible, based on above-average scores from the *MolProbity* geometry analysis web server[88].

4.5.11 Analysis of Pdu operons

A reliable set of 28 Pdu operons was defined by computationally identifying operons containing both PduCDE genes and genes for BMC-domain shell proteins. Computationally identified operons were confirmed by manual inspection. By carefully analyzing shell protein sequences within this manually-curated set of Pdu operons, we identified a GrpU homolog in the Pdu operon of *Anaerobaculum mobile* (KEGG: Anamo_0124)[168].

4.5.12 Calculation of a phylogenetic tree comparing GrpU and PduU

As a first step towards generating a phylogenetic tree comparing GrpU and PduU sequences, we created a multiple sequence alignment including 28 PduU sequences, 15 GrpU sequences, and the GrpU-like sequence from the Pdu operon of *A. mobile*. The phylogenetic tree was generated from this multiple sequence alignment using the *PhyML* software[169], and the results were visualized using *FigTree*.

CHAPTER 5

Detection of Non-Uniform Structural Polymorphism in a Single Protein Crystal: from Non-Isomorphous Unit Cells to Conformational Fluctuations

5.1 Introduction

The fact that protein molecules exhibit dynamic movement is firmly established[170–175], and there are many examples of how conformational fluctuations are critical for protein function[176]. Interconversion of conformational substates is critical for biological phenomena such as enzymatic turnover[177–179], allostery[180,181], signal transduction[182,183], force generation[184], and even evolution[185]. Despite the obvious functional importance, characterizing the atomic details of conformational heterogeneity in protein molecules remains a major challenge in structural biology.

The preeminent technique used to obtain structural information for protein molecules is X-ray crystallography, which provides atomic models with a high level of coordinate precision[186]; however, traditional crystallographic studies are not well-suited to describe conformational heterogeneity. With several notable exceptions[98,187–189], the models derived from crystallography typically represent the single, lowest-energy conformation of the molecules in the crystal lattice. The general singularity of crystallographic models results from the nature of the diffraction experiment – it is not instantaneous and it requires a large ensemble of molecules, leading to a spatiotemporal averaging of the derived electron density over all the unit cells in the crystal and throughout the duration of the data collection. The effect of this averaging is that only highly-populated conformational states are visible in the electron density. Because individual X-ray diffraction experiments tend to yield "snapshots" of molecules, the study of conformational heterogeneity by crystallographic methods can be laborious. In order to provide the correct series of snapshots to elucidate structural fluctuations and their underlying mechanisms,

crystallographers are often required to obtain multiple crystal forms of the same protein, apply various different crystal treatments, and collect individual data sets from many different crystals[187,190].

In contrast to the laborious process of solving multiple independent structures using different crystals, we hypothesize that it might be possible to detect significant structural differences by comparing multiple X-ray data sets collected from different regions of a single crystal. It is well know that the physical manipulation required to remove a crystal from its growth environment and prepare it for a diffraction experiment can alter the crystal in observable ways. For example, it has been documented that dehydration and freezing can both alter the unit cell dimensions and packing interactions in macromolecular crystals[96,97,191,192]. The fact that these sorts of structural perturbations to the crystal lattice are not uniform is also known. Crystallographers often calculate the "mosaicity" of their X-ray diffraction data, which is essentially a measure of long-range rotational disorder throughout the irradiated volume of the crystal. The work presented in this chapter serves to explore the possibility that long-range disorder in protein crystals might not only represent rigid body movements of molecules that perturb the lattice, but also might also include interesting conformational rearrangements.

The method that we propose is intended to detect and utilize structural differences within a single crystal, and the work presented here is intended to demonstrate the feasibility of this idea. First, we demonstrate that a microfocal X-ray beam can be used to probe the structural landscape of an individual protein crystal and identify significant differences between different spatial regions. Next, we attempt to outline a procedure for detecting conformational differences in electron density maps when numerous isomorphous X-ray data sets are available. In our case, the isomorphous data sets are independent, but collected from the same crystal, connecting our electron density analysis to the idea that conformational heterogeneity can be identified within a single crystal. The methodology outlined here relies on calculating pairwise correlations (R-factors) between data sets, and subsequently using principal component analysis to provide useful geometric representations of the differences between them. Using a specific test case, the EutL microcompartment shell protein from Clostridium perfringens, our initial

results confirm the viability of this analysis for detecting and visualizing structural polymorphism in crystalline protein samples.

5.2 Materials & Methods

5.2.1 Protein preparation and crystallization

For the present study, we sought to collect diffraction data while exposing the smallest possible volume of the crystal to the X-ray beam. In order to achieve this goal, we wanted to grow crystals that diffracted well, but were thin in at least one dimension. The monoclinic crystal form of the EutL tandem BMC-domain protein from *Clostridium perfringens*, cpEutL, described in detail in Chapter 2, fulfills this requirement; these crystals are only about 5μm in their thinnest dimension (Figure 5.1a). We prepared a protein sample following the same protocol described in Chapter 2, and crystallized the protein in hanging-drop format from 0.1M sodium/potassium phosphate buffer pH 6.6, 0.25M sodium chloride, and 10% (w/v) PEG-8000.

5.2.2 X-ray data collection

Prior to X-ray data collection, we harvested a single, plate-like crystal from a crystallization drop and cryoprotected it using 50% mother liquor with 2M trimethylamine-N-oxide (Figure 5.1b).

X-ray diffraction data were collected at the Advanced Photon Source on microfocal beamline 24-ID-E, equipped with a 5μm beam aperture and an ADSC Quantum315 CCD detector. Using this highly focused X-ray beam, we collected 18 data sets from essentially random, non-overlapping positions on the crystal (Figure 5.1c). Each data set was collected using the same strategy, which measured a 75° wedge of reciprocal space with 90% theoretical completeness. We maintained the crystal at cryogenic temperature (100K) throughout the course of the data collection.

We performed indexing, integration, and scaling of the X-ray data using *XDS* and *XSCALE*, and then converted intensities to structure factors with *XDSCONV*[112]. A brief description of data quality

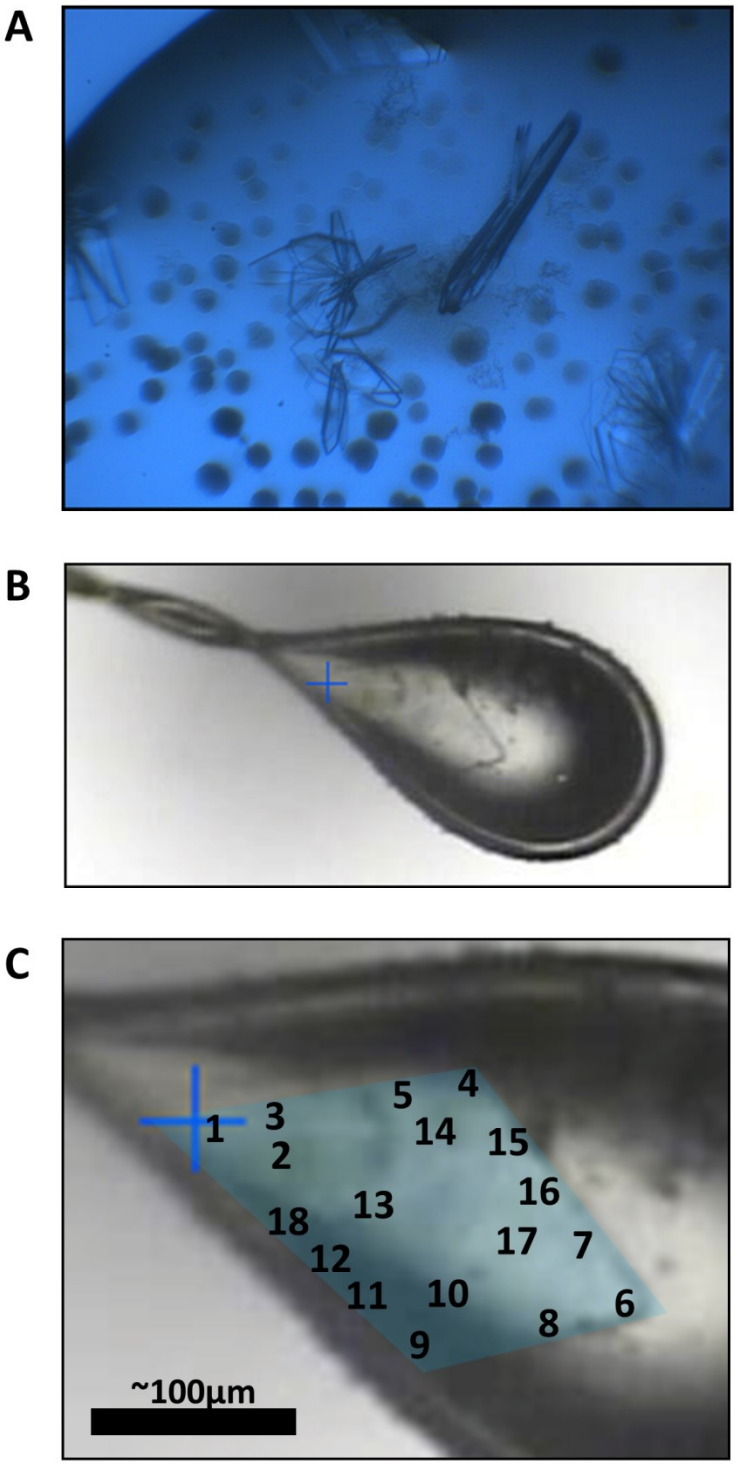


Figure 5.1. Crystallization and data collection. A) Monoclinic crystals of cpEutL grow as flat plates, which are 5µm or less in their thinnest dimension, but still appreciably large in their other two dimensions, making them ideal for our study. B) The single cpEutL crystal used to collect our 18 data sets is shown mounted and frozen in a nylon cryoloop. C) Locations on the crystal from which the 18 data sets were collected. The crystal has been false-colored to help with viewing.

Parameter	Worst	Best
Max. Resolution	2.5Å	1.9Å
Completeness	83.3%	90.8%
$\langle I/\sigma \rangle$ overall	7.37	10.54
$\langle I/\sigma \rangle$ last shell	2.01	3.14
R_{meas} overall	15.7	7.6
R_{meas} last shell	54.4	32.2

Table 5.1. Data collection statistics. A table showing some data statistics that demonstrate the range of quality for our 18 data sets. While there is some variation, all of the data are of sufficient quality.

statistics describing the group of data sets is provided in Table 5.1. All data sets were scaled to the same reference set during data reduction.

5.2.3 Pairwise comparison of X-ray data sets

Following data reduction, we performed several calculations to compare each of the data sets to one another. First, using structure factors, we calculated crystallographic *R*-factors for all possible pairs of data sets, according to Equation 1.

$$R = \frac{\sum_{hkl} \left| |F_{obs(1)}| - |F_{obs(2)}| \right|}{\sum_{hkl} \langle F_{obs} \rangle} \qquad \text{(Eq. 1)}$$

Next, we wanted to compare unit cell geometries, in order to check for potential non-isomorphism between the data sets. This test was based on comparison of unit cell metric tensors for each pair of data sets. First, for each pair of metric tensors, we calculated a 3x3 distortion matrix, $\boldsymbol{D}$. This distortion matrix, $\boldsymbol{D}$, operates by multiplication on the second metric tensor, $\boldsymbol{G}_2$, to yield the first metric tensor, $\boldsymbol{G}_1$, according to Equation 2.

$$\boldsymbol{G}_1 = \boldsymbol{D}\boldsymbol{G}_2 \qquad \text{(Eq. 2)}$$

To obtain the matrix D, we simply multiply both sides of Equation 2 by the inverse of the second metric tensor, G_2^{-1}, yielding Equation 3.

$$\boldsymbol{D} = \boldsymbol{G}_1 \boldsymbol{G}_2^{-1} \qquad \text{(Eq. 3)}$$

We note that when the unit cells are perfectly isomorphous their metric tensors are necessarily equivalent ($\boldsymbol{G}_1$=$\boldsymbol{G}_2$), and the distortion matrix becomes the identity matrix ($\boldsymbol{D}$=$\boldsymbol{I}$).

In order to analyze the unit cell non-isomorphism captured by the distortion matrices in terms of a single metric, we calculated a "pairwise distortion index" by comparing the distortion matrices to the identity matrix and computing the sum described by Equation 4, where d_{ij} and i_{ij} represent equivalent elements in the distortion matrix and the identity matrix respectively.

$$distortion\ index = \sum_{all\ ij} \left(|d_{ij} - i_{ij}| / 2 \right) \qquad \text{(Eq. 4)}$$

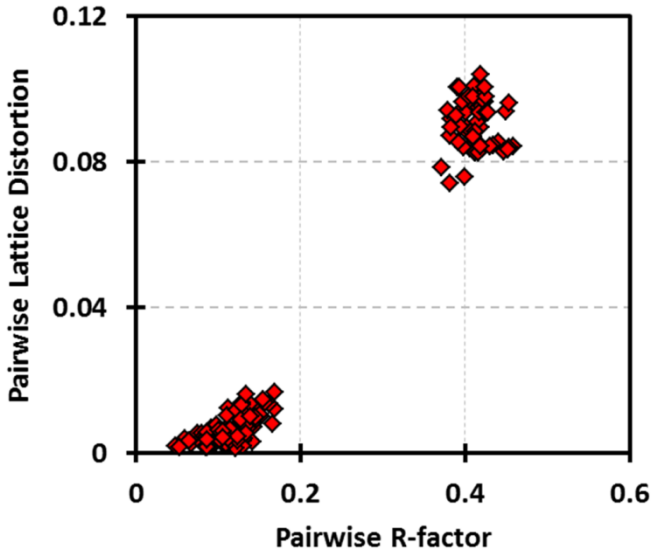


Figure 5.2. High pairwise R-factors are correlated with non-isomorphism. A plot comparing pairwise crystallographic R-factors between our 18 data sets with the corresponding pairwise measurements of isomorphism (distortion index) between the data sets shows that high R-factors are due to non-isomorphous crystal lattices.

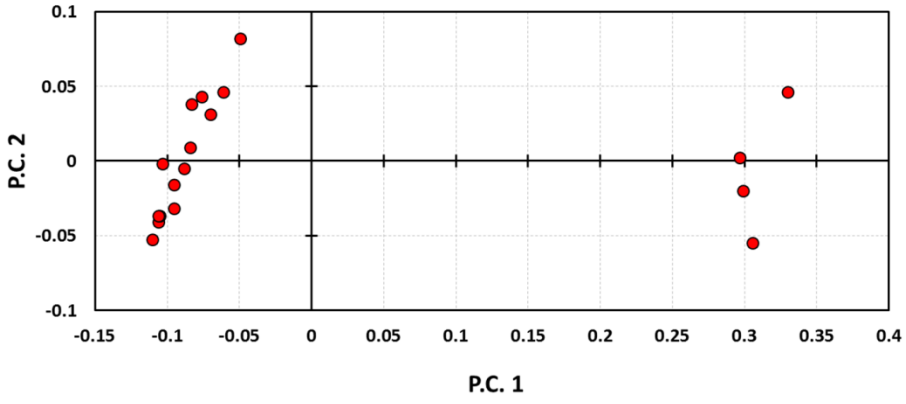


Figure 5.3. PCA clustering of 18 data sets. Projection of the 18 data sets onto the first two principal components following PCA. Two clear groups are evident.

The metric described by Equation 4 was found to be a useful estimator of the fractional difference between two unit cells.

For each pairwise comparison of data sets, we plotted the calculated R-factor against the calculated distortion index (Figure 5.2).

5.2.4 Data set clustering by principal component analysis

In order to visualize the relationships between the 18 individual data sets, we performed principal component analysis (PCA).

By formulating a geometric representation of our 18 X-ray data sets as 18 points in 18-dimensional space, we could represent the pairwise "distances" between these points in terms of their pairwise *R*-factors. Using this representation, we constructed an 18x18 "distance matrix," ***D***, whose elements were equal to the pairwise *R*-factors between the corresponding data sets *i* and *j* (see Equation 5).

$$D = \begin{bmatrix} d_{11} & \cdots & d_{i1} \\ \vdots & \ddots & \vdots \\ d_{1j} & \cdots & d_{ij} \end{bmatrix} \; ; \; d_{ij} = \frac{\sum_{hkl} \left| |F_{obs(i)}| - |F_{obs(j)}| \right|}{\sum_{hkl} \langle F_{obs} \rangle} \qquad \text{(Eq. 5)}$$

Note that this distance matrix is square, symmetric, and contains zeroes along its diagonal. Furthermore, using the information in the distance matrix, ***D***, we can calculate the distance from each of the data points to the center of mass of the 18-dimensional set of coordinates, according to Equation 6.

$$d_{oi} = \sqrt{\frac{1}{N}\left[\sum_{j=1}^{N} d_{ij}^2 - \left(\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} d_{ij}^2}{2N}\right)\right]} \qquad \text{(Eq. 6)}$$

Next, given the set of pairwise distances, we sought to place the data points within an 18-dimensional coordinate system whose origin lies at the center of mass, with orthogonal axes representing the principal components (eigenvectors) of the data. The coordinates of the data points in this 18-dimensional coordinate system can be expressed as column vectors ($\overrightarrow{x_i}$) in an 18x18 matrix, ***X***. Given the

definition of $\boldsymbol{X}$, the matrix $\boldsymbol{X}^{\mathrm{T}}$ contains row vectors ($\overrightarrow{x_j}$) expressing the same 18-dimensional coordinates. Consequently, $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ is also a symmetric, 18x18 matrix, whose elements are equal to $\overrightarrow{x_i}\cdot\overrightarrow{x_j}$ (Equation 7).

$$\boldsymbol{X} = \begin{bmatrix} \overrightarrow{x_1} \\ \vdots \\ \overrightarrow{x_i} \end{bmatrix} \quad \therefore \quad \boldsymbol{X}^T = [\overrightarrow{x_1} \quad \dots \quad \overrightarrow{x_j}] \quad \text{and} \quad \boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} \overrightarrow{x_1}\cdot\overrightarrow{x_1} & \cdots & \overrightarrow{x_j}\cdot\overrightarrow{x_1} \\ \vdots & \ddots & \vdots \\ \overrightarrow{x_1}\cdot\overrightarrow{x_i} & \cdots & \overrightarrow{x_j}\cdot\overrightarrow{x_i} \end{bmatrix} \qquad \text{(Eq. 7)}$$

Defining the matrix $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ would allow us to factorize by eigenvalue decomposition to obtain $\boldsymbol{X}$. Using the elements in our distance matrix, $\boldsymbol{D}$, and the calculated distance of each data point to the center of mass, we can compute the dot products $\overrightarrow{x_i}\cdot\overrightarrow{x_j}$ using the law of cosines as described by Equation 8.

$$\overrightarrow{x_i}\cdot\overrightarrow{x_j} = \frac{1}{2}\left(d_{oi}^2 + d_{oj}^2 - d_{ij}^2\right) \qquad \text{(Eq. 8)}$$

Equation 8 can be used to define $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$, allowing subsequent factorization by eigenvalue decomposition to obtain $\boldsymbol{X}$.

After obtaining $\boldsymbol{X}$, we projected the 18 data points representing our X-ray data sets onto the first two principal components (eigenvectors), in order to most effectively visualize the differences between them (Figure 5.3). Projection of the data onto these first two principal components captures 86% of the variation between the X-ray data sets, based on their eigenvalues. This analysis revealed that the 18 data sets could be clustered into two groups, a major group of 14 members, and a minor group of 4 members.

5.2.5 Determination of "average" structures from clustered data sets

Following our PCA, we determined that the main difference between the two groups of data sets was unit cell non-isomorphism. Subsequently, we were able to successfully merge measured reflection intensities from X-ray data sets belonging to the same group. We merged data sets 1-8, 10, and 13-17 to create a "group 1 average" data set, and we merged data sets 9, 11, 12, and 18 to create a "group 2 average" data set. The merging was performed with *XSCALE*, and the merged intensities were converted to structure factors with *XDSCONV*[112]. The differences between the two unit cells in the two "average" data sets are summarized in Table 5.2. The free set of reflections was designated according to the original data set used to determine the monoclinic crystal structure of cpEutL.

Group	*N* datasets	a	b	c	α	β	γ	R_{merge}
1	14	118.74Å	66.14Å	80.14Å	90°	108.53°	90°	13.2%
2	4	118.72Å	66.14Å	78.58Å	90°	111.22°	90°	13.5%

Table 5.2. Comparison of non-isomorphous unit cells. The table compares unit cell parameters derived for the two non-isomorphous unit cells present in our crystal. The minor group (group 2) corresponds to a unit cell with a shorter *c*-axis and a wider β angle. The *R*-factors provided show that the data sets are reasonably consistent within the two individual groups.

We used our "average" data sets to determine the structures of the two different, non-isomorphous packing arrangements. First, a molecular replacement search was carried out using *Phaser*[115], with the original monoclinic cpEutL as a search model. Solutions were readily found for both average data sets, refined briefly with *phenix.refine*[113,118], and inspected to determine the differences in crystal packing.

5.2.6 Principal component analysis of the major data set group

In order to further search for conformational differences within our single crystal specimen, we performed a second round of PCA, this time considering only the 14 data sets from the major group identified in the first round of PCA (data sets 1-8, 10, and 13-17). The mathematical procedure used to perform this analysis was identical to the procedure described for the initial round of PCA, except this time we started with a 14x14 distance matrix describing the correlation (*R*-factors) between only the 14 data sets that comprise the major group.

We again projected the 14 data points representing our X-ray data sets onto the first two principal components (eigenvectors) (Figure 5.4). For this round of PCA, projection of the data onto these first two principal components captures 52% of the variation between the X-ray data sets, based on their eigenvalues. The second round of PCA did not reveal clear sub-groupings of the 14 data sets within the major group.

5.2.7 Novel electron density calculations

As a next step towards using our single crystal analysis to define relevant conformational polymorphism, we have started to develop two new types of electron density maps that are intended to highlight the most significant structural differences between isomorphous data sets. Both of these novel map types rely on PCA, as discussed above, to define the contributions of many different data sets to the final electron density.

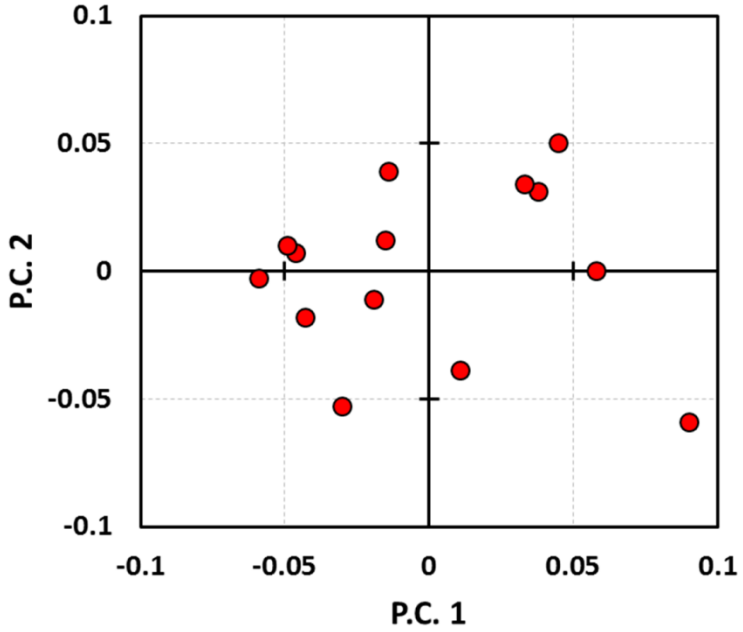


Figure 5.4. PCA of 14 isomorphous data sets. Projection of the 14 isomorphous data sets which comprise the major group onto the first two principal components following PCA. No clear groupings can be found.

We refer to the first of these new map types as the "eigen-difference" map. This type of map is essentially a generalization of the classic F_o-F_o isomorphous difference map, however it is generalized to the case of N data sets rather than just two. To calculate this eigen-difference map, we first select an eigenvector calculated from our PCA. The structure factor amplitudes used for the map synthesis are calculated according to Equation 9, where x_n denotes the coordinate of a particular data set (n) projected onto the selected eigenvector, and $F_{obs(n)}$ denotes a structure factor from data set n.

$$|F_{hkl}| = \sum_{n=1}^{N} x_n |F_{obs(n)}| \qquad \text{(Eq. 9)}$$

Phases for these maps are generated using the atomic coordinates of the average structure, determined from a single data set containing merged structure factors. The effect of this map should be to cancel out all molecular features that are the same in each of the data sets while highlighting, as positive and negative map features, the conformational differences represented by the differences in the structure factors. Using our major group of 14 data sets, we calculated two eigen-difference maps as described above. These two maps correspond to the first two eigenvectors identified during PCA.

Our second type of new electron density map is referred to as the eigen-delta map. Once again, to begin the map calculation, we select a single eigenvector calculated from PCA. The structure factor amplitudes for these maps are then derived according to Equation 10, where x_n denotes the coordinate of a particular data set (n) projected onto the selected eigenvector, $F_{obs(n)}$ denotes a structure factor from data set n, and λ denotes the eigenvalue associated with the selected eigenvector.

$$|F_{hkl}| = \langle |F_{obs}| \rangle \pm \Delta|F_{hkl}| \text{, where } \Delta|F_{hkl}| = (1/\lambda) \times \sum_{n=1}^{N} x_n |F_{obs(n)}| \qquad \text{(Eq. 10)}$$

As with the eigen-difference maps, phases are generated using the atomic coordinates of the average structure. We note that when calculating structure factor amplitudes for the eigen-delta maps, it is possible to either add or subtract the $\Delta|F_{hkl}|$ term from the average value of $|F_{obs}|$. Both of these strategies are equally valid mathematically, and it is possible that certain structural differences might be more apparent in the density under one operation versus the other. We used our major group of 14 data sets to

calculate four eigen-delta maps. These four maps represent the first two eigenvectors identified by PCA, and for each eigenvector we calculated the eigen-delta map using both positive and negative values of $\Delta|F_{hkl}|$.

For both map types, we experimented with various resolution cutoffs and scaling protocols. We found that both the eigen-difference and eigen-delta maps looked best at 2.5-3.0Å resolution, after global isotropic scaling of the 14 data sets with *ScaleIt* in *CCP4*[116].

5.3 Results & Discussion

The overall goal of the experimental strategy presented here is to expose and exploit the potential nonuniformity of structural perturbations within a single crystalline specimen to obtain information about structural polymorphism in proteins. Our strategy is based on the idea that multiple crystal forms containing different molecular conformations can be simultaneously present in a single macroscopic crystal. Our approach requires collecting multiple independent data sets from a single crystal using a microfocal synchrotron X-ray beam, and subsequently comparing the data sets using principal component analysis (PCA). The results described in this section begin to demonstrate that the information derived from PCA of diffraction data can be utilized to highlight structural differences between the underlying crystal structures.

5.3.1 Eighteen independent data sets from a single crystal

In order to perform our designed structural analysis, the first step was to select an appropriate crystal specimen. Our assumption was that we would obtain the best chance of finding differences within an individual crystal by collecting diffraction data while irradiating the smallest number of unit cells, thereby averaging the resulting diffraction experiment over a smaller number of molecules. Consequently, we sought crystals that were thin in one dimension (i.e. plate-like crystals), but diffracted well and were reasonably robust to radiation damage. Fortunately, we had identified such crystals in our earlier work

with the bacterial microcompartment shell protein EutL. The monoclinic crystal form of EutL from Clostridium perfringens (cpEutL) forms thin plates that are only about 5μm in their thinnest dimension, making them ideal for this study (Figure 5.1a).

Using a single, monoclinic cpEutL crystal, we collected 18 independent data sets. The data sets were collected using a microfocal X-ray beam with a 5μm cross-sectional diameter, capable of delivering a very large number of photons to a very small region of the crystal. The high intensity of the microfocal X-ray beam was the impetus for using crystals that were robust to radiation damage. Because the cross-section of the X-ray beam was so small, it was easy to collect 18 data sets from completely non-overlapping positions on the crystal (Figure 5.1c), making each of these individual data sets truly independent. Additionally, we note that the original crystal structure of monoclinic cpEutL was determined using a 70μm beam. A simple geometric calculation demonstrates that this 70μm X-ray beam would expose roughly 200 times the volume of the crystal that was exposed by the 5μm beam used in the present study.

All 18 of the data sets collected are of relatively high quality, although there is some statistical variation between sets (Table 5.1). These statistical differences were early indications of the non-uniformity of the crystalline specimen. All of the data sets have a resolution of 2.5Å or better, with some as high as 1.9Å, and all of the data sets have $\langle I/\sigma I \rangle$ greater than 2.0, and merging *R*-factors that indicate reasonable internal consistency. We chose to collect data over a 75° wedge of reciprocal space, centered on a vector normal to the flat face of the plate-like crystal, giving approximately 90% completeness (theoretical) in space group *C*2. We collected this wedge of data, because we determined that it provided the best compromise between data completeness and irradiation of the smallest possible crystal volume.

After scaling and merging the 18 data sets independently, we converted intensities to structure factor amplitudes, which we used to perform the rest of our analyses. Additionally, it is important to note that the same set of free reflections was designated for each of the 18 data sets, consistent with the free set used for initial structure determination of monoclinic cpEutL.

5.3.2 Pairwise comparison of data sets reveals non-isomorphism

As a first step towards identifying differences between our 18 data sets, we calculated all possible pairwise *R*-factors between them. We noticed that the calculated *R*-factors tended to be either small (<20%), or rather large (~40%), with virtually no intermediate values. This observation prompted us to explore the possibility that large *R*-factors result from unit cell non-isomorphism between the two data sets, rather than from conformational differences. We answered this question by also calculating a lattice "distortion index," which was intended to serve as a single metric describing the fractional deviations between two unit cells. By plotting these two calculations, *R*-factors and distortion indices, against one another, we observed that *R*-factors and distortion indices are highly correlated, confirming that differences in structure factor magnitudes do indeed result from unit cell non-isomorphism (Figure 5.2). While this simplistic analysis clearly correlates *R*-factors with lattice distortions and demonstrates the presence of non-isomorphism in the crystal, it does not indicate how many different unit cells are present.

5.3.3 Principal component analysis reveals two unit cells in the same macroscopic crystal

While the pairwise comparisons of our 18 data sets demonstrated that there were non-isomorphous regions within our single crystal, we sought a more detailed understanding of how this structural variation was distributed. For this purpose, we turned to principal component analysis (PCA), to cluster our data sets into isomorphous groups. Starting with a correlation matrix defined by pairwise *R*-factors, we could formulate a geometric construction that allowed us to treat our 18 data sets as points in 18-dimensional space. PCA then allowed us to place those points into a coordinate system defined by the principal components (eigenvectors) of the data. Projection of those 18 points onto the first two principal components captured 82% of the variation between the X-ray data sets, and revealed two very obvious groups (Figure 5.3). The major group consisted of 14 data sets, including sets 1-8, 10, and 13-17. The minor group consisted of 4 data sets, including sets 9, 11-12, and 18.

When we identify the beam position used for each data set with respect to the crystal, it can be seen that the minor group of data sets represents four locations along the same side of the crystal (Figure

5.5). This mapping of unit cell dimensions onto the crystal shows that the perturbations to the unit cell are not spatially uniform, but instead that there are two distinct domains within the crystal. Based on the location of the domain representing the minor group of data sets, near the edge of the crystal and nylon loop, it is tempting to speculate that the cause of the lattice perturbation might be dehydration, faster cooling, or mechanical interaction with the loop, however there is no evidence for any of these mechanisms. Ultimately, the cause of the lattice perturbation is unimportant for our study; what matters is that our mapping provides evidence that this type of multiple data set X-ray analysis can be used to find structural polymorphism within a "single" crystal specimen.

After identifying the presence of two non-isomorphous domains in our crystal, we wanted to define the specific structural differences between the two domains. We started by merging the individual data sets within the two groups, to give two "average" data sets, one representing each of the two domains. Comparison of the unit cell parameters calculated for each average data set show that the minor unit cell has a wider β angle and a shorter c-axis (Table 5.2). We solved the two structures by molecular replacement, and performed preliminary atomic refinement of the coordinates. We analyzed the two structures, and noticed that the differences between the two unit cells are attributed to a subtle shearing of the monoclinic cell, which makes the β angle wider in one crystal form, as well as to a slight compression between the layers of molecules along a vector normal to the *ab* plane of the unit cell (Figure 5.6).

5.3.4 Steps toward identifying conformational polymorphism using principal component analysis

While it is interesting that we were able to identify non-isomorphism within a single crystal, we ultimately want to identify functionally relevant conformational polymorphism. Therefore, we applied our PCA-based strategy to the major group of 14 data sets, hoping that we could identify sub-groups that represented different crystal structures. Unfortunately, for this group of 14 data sets, the PCA does not result in such an obvious clustering (Figure 5.4).

Even in the absence of obvious substructures within the large domain of our crystal, we reasoned that it might be possible to still exploit differences between the data sets by looking at electron density

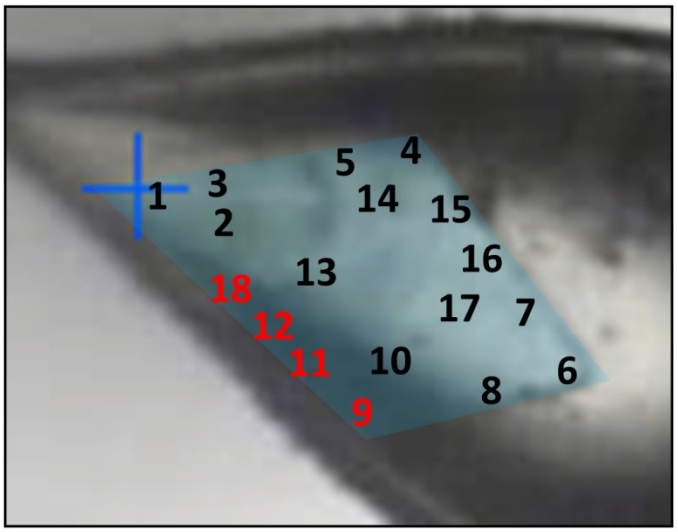


Figure 5.5. Non-isomorphous domains in a single crystal. Mapping the locations of the data sets onto the crystal relative to their unit cells reveals that there appear to be two discreet, non-isomorphous domains.

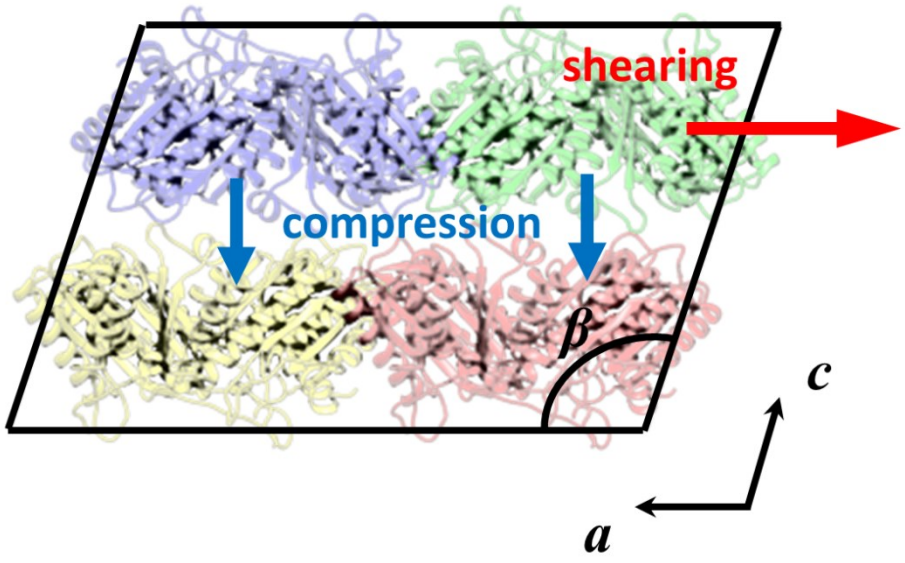


Figure 5.6. Displacements resulting in unit cell non-isomorphism. The two non-isomorphous unit cells identified in our crystal differ as a result of unit cell shearing and compression between the layers of molecules.

maps generated from algebraic sums of structure factor amplitudes that were weighted according to the results of our PCA. To this end, we generated two novel types of electron density maps.

The first of these maps, the eigen-difference map, is essentially a generalization of the classic Fo-Fo map to the case of *N* data sets rather than two. The F_o-F_o map is an established method for identifying differences between isomorphous data sets[193], and our eigen-difference map is intended to provide a similar representation of structural heterogeneity, wherein positive and negative features in the map correspond to alternative conformations of the molecules in the crystal. We generated two of these eigen-difference maps, according to Equation 9, and analyzed the features therein. We noticed that in general, these maps contain more positive features than negative ones, and also that these features do most often cover the protein molecule (Figure 5.7). Admittedly, the features in these maps do not highlight any obvious conformational polymorphism, however it is interesting to note that some of the density features lie on segments of the protein molecules that have been shown to occupy multiple conformations[27,34].

The second new type of electron density map we generated is referred to as the eigen-delta map. This map, generated according to Equation 10, is intended to have features that look like a protein molecule (i.e. more like a $2F_o$-F_c map), but be modified in a way that reflects conformational differences between the data sets. We generated four of these maps, and inspected them. Unfortunately, we found that these maps were relatively uninformative, as they were nearly identical to the $2F_o$-F_c map calculated for the average structure.

The electron density analysis presented here falls short of the goal of identifying conformational polymorphism between isomorphous data sets. Although there are some interesting features in our eigen-difference maps, which is encouraging, they are largely uninterpretable, and our eigen-delta maps are also uninformative with regard to conformational heterogeneity. At the present stage, it is not clear whether our maps are uninformative because the method is not a viable strategy for identifying conformational heterogeneity, or because our system does not display the fluctuations we seek to find. A proper test of the eigen-difference and eigen-delta maps will require synthetic data wherein the extent and nature of the conformational heterogeneity is known *a priori*.

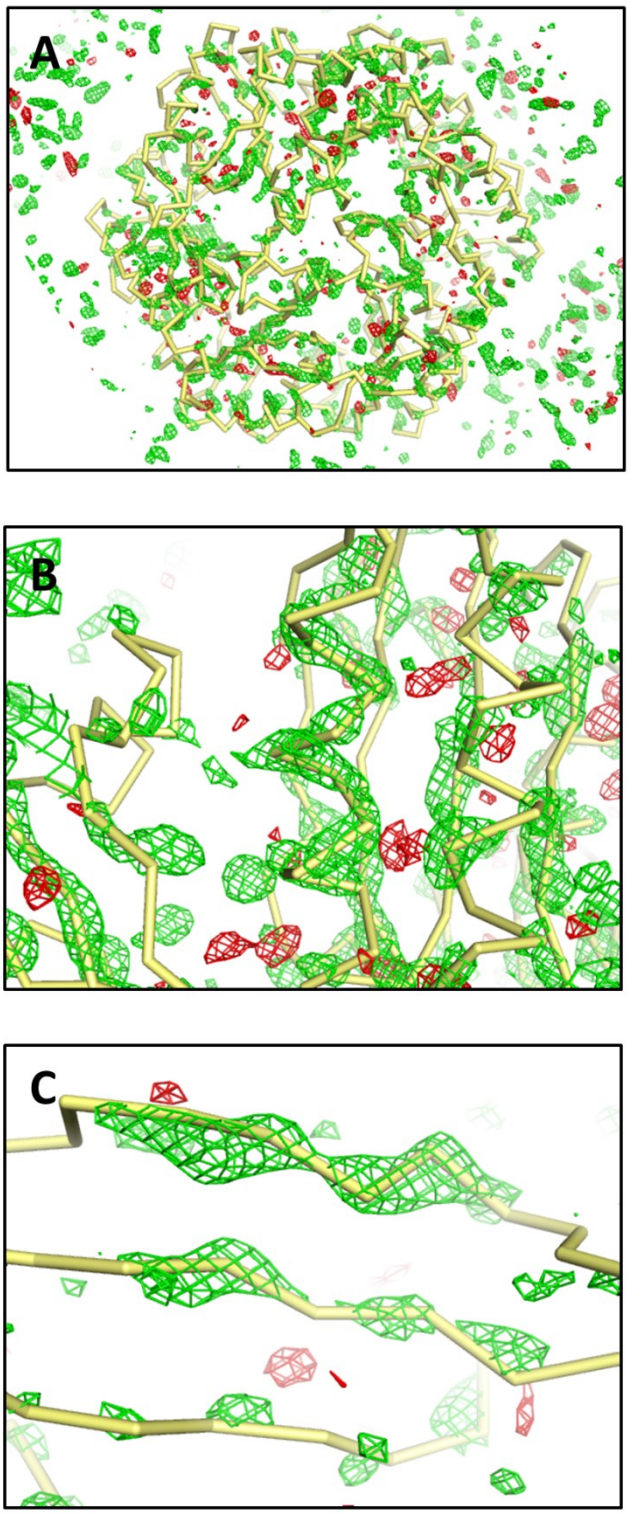


Figure 5.7. An eigen-difference map. Our initial attempt to calculate eigen-difference maps based on the first principal component of our 14 isomorphous data sets produced electron density exemplified by these images. While many of the peaks in the density overlap with the molecular model, the features are difficult to interpret in terms of conformational polymorphism. The maps were calculated using a 3.0Å resolution cutoff, and are contoured at 3.0σ.

5.4 Conclusions

The ideas and results described in Chapter 5 serve to provide new insight into methodology for understanding conformational fluctuations in protein molecules using X-ray crystallography. Development of such techniques addresses a significant challenge in structural biology. Specifically, we have demonstrated that due to non-uniform perturbations of the crystalline molecules, information about multiple conformational states might be extracted by collecting and comparing multiple independent data sets from a single crystalline sample. We used PCA to highlight the differences between multiple data sets, and used the relationships that emerged from PCA to guide our search for structural variation and map the structural domain landscape of our crystal specimen. Our analysis revealed two spatially distinct structural domains in the crystal, with non-isomorphous unit cells. This work highlights another use for microfocal X-ray beams, which have emerged as useful tools for specialized X-ray diffraction experiments[194]. Finally, we have laid out the fundamental mathematics for generating new types of electron density maps that might further illuminate conformational polymorphism in proteins; however, the true utility of these maps has yet to be tested rigorously.

CHAPTER 6

Conclusions

The study of bacterial microcompartments (MCPs) over the last several decades has contributed to a paradigm shift in field of prokaryotic biology. The cytoplasm of the prokaryotic cell was once thought of as being more or less homogenous, without any sophisticated internal organization. In contrast to this outdated view, we now recognize that prokaryotic cells do, in fact, have a repertoire of complex subcellular structures that fulfill specific physiological roles[3]. These structures effectively function as organelles in prokaryotic cells, even if they are morphologically different from their membrane-bound, eukaryotic counterparts. Given their role in sequestering specific metabolic reaction pathways, MCPs are among the group of prokaryotic structures that can be classified as functional organelles[3,6,8,9]. The studies presented here further our understanding of MCPs and provide a greater comprehension of the historically underappreciated complexity of prokaryotic cells.

Owing to their structural and functional complexity, the study of MCPs provides a wealth of opportunities to explore biochemical and biophysical phenomena. MCPs are beautiful, highly-ordered structures that self-assemble from thousands of protein subunits, forming a semi-permeable, polyhedral shell that surrounds a specific set of metabolic enzymes[8,9,22]. Despite several decades of research, there are still many unanswered questions about the molecular mechanisms of MCP function and assembly. Recently, the maturation of the field of synthetic biology has fueled interest in MCPs because of their potential to be engineered as genetically-encoded, biochemical reaction vessels[52,53,55].

The selective permeability of the MCP shell is critical for the function of these elaborate protein complexes, and we have yet to fully elucidate the structural mechanisms responsible for this feature[23,24]. X-ray crystallography has been instrumental in developing our understanding of molecular transport through the MCP shell. Early crystal structures showed simple pores through the centers of tightly-packed, symmetric BMC-domain hexamers could support transport of small substrates and products[25,27,29].

Subsequently, further structural characterization of the pseudohexameric tandem BMC-domain shell proteins demonstrated that the broken symmetry resulting from domain duplication confers an important functional advantage on the tandem BMC domain proteins relative to their single-domain counterparts, giving them the ability to form more structurally and chemically diverse pores with apparently more complex transport properties[27,29,34,35]. One example of such a discovery was the observation that the EutL tandem BMC-domain pseudohexamer can populate two distinct conformations, which differ in the presence or absence of a large central pore, indicating that large cofactors likely traverse the shell through gated pores[27,34]. In this particular case, the pseudosymmetry of the protein oligomer allows for conformational heterogeneity that imparts EutL with its unique function as a gated pore into the MCP lumen.

The goal of this dissertation is to expand our knowledge of the functional mechanisms that permit the MCP shell to act as a semi-permeable diffusion barrier. The work undertaken in fulfillment of that goal began with the exploration of the mechanism by which pseudosymmetry and other molecular phenomena facilitate conformational polymorphism of the EutL shell protein. In this case, the presence of conformational polymorphism was known, and we sought to define the structural basis of the functional transition between conformational states. Several additional, but initially unrelated, structure determination efforts fortuitously revealed additional roles for conformational polymorphism and broken symmetry in facilitating the specialized molecular transport functions of BMC-domain proteins. In the cases of these projects, we identified an unanticipated role for conformational polymorphism in promoting BMC-domain function.

Collectively, the research projects presented here illustrate the intimate connection between conformational polymorphism and broken symmetry in BMC-domain proteins, and demonstrate that the functional mechanisms of the shell proteins within this family universally display these two features. Two main ideas have been explored, which are somewhat opposite in nature. The first is the idea that in the EutL tandem BMC-domain protein and its homologs, boken symmetry in the form of pseudosymmetry gives the oligomers the capacity for functional conformational polymorphism. On the other hand, the

second idea is that for single BMC-domain shell proteins, conformational heterogeneity leads to quasisymmetry which is required for function.

The studies presented in Chapter 2 were conducted in attempt to better understand the atomic basis for the conformational fluctuations that arise from pseudosymmetry of the EutL tandem BMC-domain protein from the Eut MCP. The Eut MCP is of intrinsic interest, because associations of genotypes and phenotypes have implicated the Eut MCP as being the molecular entity responsible for bacterial enteropathogenesis (food poisoning)[19,68,73,74]. EutL is likely to be critical for the function of the Eut MCP, acting as a gated pore for large cofactor molecules[27,34]. Although it is recognized that EutL is structurally polymorphic it is not known how the conformational equilibrium between the open and closed states is regulated. Our results support a model of negative allosteric regulation of pore opening by ethanolamine binding. Furthermore, some evidence suggests that the allosteric mechanism might be modulated by electrochemical environment; however this hypothesis has yet to be tested rigorously. The model of allosteric regulation by the substrate ethanolamine is consistent with our hypotheses about when and why the EutL pore needs to be open or closed. The EutL pore is assumed to allow the transport of large cofactor molecules that need to be exchanged with the cytosol when the absence of ethanolamine leads to inactivation of the cofactors in the active site of the ethanolamine-ammonia lyase enzyme. Additionally, it has been shown that the oxidative burst created by the host immune response is necessary for ethanolamine utilization, which provides a possible connection to our observation that EutL contains a disulfide bond and oxidation of the protein appears to stabilize the closed conformation of the pore[78,79].

Next, Chapter 3 describes the structure determination of the L11K point mutant of the CcmK1 single BMC-domain protein from the β-carboxysome. The primary goal of this work was to highlight an instructive example of how to handle a challenging situation in crystallographic structure determination. Specifically, we wanted to point out that sometimes broken crystallographic symmetry requires a new choice of unit cell origin in order to identify the highest symmetry space group. Although primarily intended for a crystallographic audience, this work also revealed an interesting several interesting features of the CcmK1 protein that relate to its biological function. First, the L11K mutation gave some insight

into the electrostatic properties of the small pore at the center of the CcmK1 hexamer. More interestingly, the crystal packing arrangement of the molecules in our CcmK1 crystal revealed that even the highly-symmetric CcmK1 hexamer might actually have subtle deviations from perfect 6-fold symmetry. This potentially broken symmetry implies, by exchange of alternative conformations, that CcmK1 might display dynamic motion leading to some type of reciprocating mechanism. If the broken symmetry in CcmK1 involves the pore region, it might result from a need to alleviate steric clashes at the center of the symmetric oligomer.

Chapter 4 of the dissertation describes studies of another single BMC-domain shell protein, GrpU from the Grp MCP, which revealed unexpected conformational heterogeneity and broken symmetry with functional implications. Initially, purification and spectroscopic characterization of this protein demonstrated that it binds to an iron-sulfur cluster. While tandem BMC-domain, iron-sulfur metalloproteins are known to exist in the shell of the Pdu MCP[29,31,35], we were surprised to find that GrpU bound to a similar metal cluster, because it is a single BMC-domain hexamer. Crystallographic structure determination and molecular modeling demonstrated that GrpU has a structurally disordered central pore that contains a conserved iron-sulfur cluster binding motif. The disorder allows GrpU to assume a quasisymmetric configuration upon binding an iron-sulfur ligand with lower symmetry. In addition to revealing another connection between conformational polymorphism, broken symmetry, and function, these studies provided the first structures of BMC-domain proteins from the Grp MCP, and also identified a unique type of iron-sulfur cluster binding site, previously unobserved in metalloproteins.

In Chapter 5, the work described departs from exploring the functional mechanisms of specific proteins, and instead provides details about progress toward developing a novel method for identifying structural heterogeneity in macromolecular crystallography. As demonstrated by Chapters 2-4, the ability to detect and characterize conformational polymorphism is critical in structural biology, which is often achieved through the laborious process of solving multiple structures using different crystal specimens. Chapter 5 presents methodology that attemps to circumvent the bottleneck in the crystallographic analysis of structural polymorphism by collecting multiple independent X-ray datasets from a single crystal and

using principal component analysis (PCA) to cleverly search for significant differences between those data sets. Using a specific test case wherein multiple nonisomorphous domains existed in a single crystal, we demonstrated that data collection with a microfocal X-ray beam followed by dataset clustering with PCA can illuminate significant structural differences within a single crystal. Additionally, we have outlined the mathematical formalism for computing crystallographic electron density maps that we feel may find utility in revealing conformational polymorphism, however the practical utility of these maps has not been explored extensively.

The work presented in this dissertation represents a comprehensive discussion of the role of conformational polymorphism in BMC-domain shell protein function. Forming a series of illustrative examples, this work demonstrates that conformational polymorphism and associated broken oligomeric symmetry are defining features of the functional mechasims that allow these proteins to collectively form a semi-permeable diffusion barrier in the context of a bacterial organelle. BMC-domain shell proteins are not, however, unique in the emergence of function from conformational polymorphism and broken symmetry[41]. Indeed, we have seen further into the world of MCPs because we have stood on the shoulders of such giants as Max Perutz, Richard Feynman, Daniel Koshland, and Paul Boyer, who, among others, pioneered the idea that structural fluctuations and broken symmetry are important determinants of protein function, especially for oligomeric proteins. Following in this legacy, we have provided further examples of this phenomenon, and also attempted to contribute methodology to facilitate future explorations of the connection between conformational polymorphism and protein function.

REFERENCES

1. Sweeney, W.V. & Rabinowitz, J.C. (1980). Proteins Containing 4Fe-4S Clusters: An Overview. *Annu. Rev. Biochem.* **49**, 139–161

2. Shorte, S.L. & Frischknecht, F. (Springer: 2007). *Imaging Cellular and Molecular Biological Functions: With 13 Tables*.

3. Shively, J.M. (Springer: 2006). *Complex Intracellular Structures in Prokaryotes*.

4. Shively, J.M., Ball, F., Brown, D.H. & Saunders, R.E. (1973). Functional Organelles in Prokaryotes: Polyhedral Inclusions (Carboxysomes) of Thiobacillus neapolitanus. *Science* **182**, 584–586

5. Shively, J.M., Ball, F.L. & Kline, B.W. (1973). Electron Microscopy of the Carboxysomes (Polyhedral Bodies) of Thiobacillus neapolitanus. *J. Bacteriol.* **116**, 1405–1411

6. Yeates, T.O., Kerfeld, C.A., Heinhorst, S., Cannon, G.C. & Shively, J.M. (2008). Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat. Rev. Microbiol.* **6**, 681–691

7. Cheng, S., Liu, Y., Crowley, C.S., Yeates, T.O. & Bobik, T.A. (2008). Bacterial microcompartments: their properties and paradoxes. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **30**, 1084–1095

8. Yeates, T.O., Crowley, C.S. & Tanaka, S. (2010). Bacterial microcompartment organelles: protein shell structure and evolution. *Annu. Rev. Biophys.* **39**, 185–205

9. Kerfeld, C.A., Heinhorst, S. & Cannon, G.C. (2010). Bacterial microcompartments. *Annu. Rev. Microbiol.* **64**, 391–408

10. Jorda, J., Lopez, D., Wheatley, N.M. & Yeates, T.O. (2013). Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci. Publ. Protein Soc.* **22**, 179–195

11. Cannon, G.C., Bradburne, C.E., Aldrich, H.C., Baker, S.H., Heinhorst, S. & Shively, J.M. (2001). Microcompartments in Prokaryotes: Carboxysomes and Related Polyhedra. *Appl. Environ. Microbiol.* **67**, 5351–5361

12. Badger, M.R. & Price, G.D. (2003). CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J. Exp. Bot.* **54**, 609–622

13. The Cells That Rule the Seas. at <http://www.scientificamerican.com/article/the-cells-that-rule-the-s/>

14. Ellis, R.J. (2010). Biochemistry: Tackling unintelligent design. *Nature* **463**, 164–165

15. Sampson, E.M. & Bobik, T.A. (2008). Microcompartments for B12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. *J. Bacteriol.* **190**, 2966–2971

16. Havemann, G.D. & Bobik, T.A. (2003). Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in Salmonella enterica serovar Typhimurium LT2. *J. Bacteriol.* **185**, 5086–5095

17. Kofoid, E., Rappleye, C., Stojiljkovic, I. & Roth, J. (1999). The 17-gene ethanolamine (eut) operon of Salmonella typhimurium encodes five homologues of carboxysome shell proteins. *J. Bacteriol.* **181**, 5317–5329

18. Stojiljkovic, I., Bäumler, A.J. & Heffron, F. (1995). Ethanolamine utilization in Salmonella typhimurium: nucleotide sequence, protein expression, and mutational analysis of the cchA cchB eutE eutJ eutG eutH gene cluster. *J. Bacteriol.* **177**, 1357–1366

19. Garsin, D.A. (2010). Ethanolamine utilization in bacterial pathogens: roles and regulation. *Nat. Rev. Microbiol.* **8**, 290–295

20. Rondon, M.R., Horswill, A.R. & Escalante-Semerena, J.C. (1995). DNA polymerase I function is required for the utilization of ethanolamine, 1,2-propanediol, and propionate by Salmonella typhimurium LT2. *J. Bacteriol.* **177**, 7119–7124

21. Penrod, J.T. & Roth, J.R. (2006). Conserving a volatile metabolite: a role for carboxysome-like organelles in Salmonella enterica. *J. Bacteriol.* **188**, 2865–2874

22. Tanaka, S., Kerfeld, C.A., Sawaya, M.R., Cai, F., Heinhorst, S., Cannon, G.C. & Yeates, T.O. (2008). Atomic-level models of the bacterial carboxysome shell. *Science* **319**, 1083–1086

23. Yeates, T.O., Thompson, M.C. & Bobik, T.A. (2011). The protein shells of bacterial microcompartment organelles. *Curr. Opin. Struct. Biol.* **21**, 223–231

24. Yeates, T.O., Jorda, J. & Bobik, T.A. (2013). The shells of BMC-type microcompartment organelles in bacteria. *J. Mol. Microbiol. Biotechnol.* **23**, 290–299

25. Kerfeld, C.A., Sawaya, M.R., Tanaka, S., Nguyen, C.V., Phillips, M., Beeby, M. & Yeates, T.O. (2005). Protein structures forming the shell of primitive bacterial organelles. *Science* **309**, 936–938

26. Wheatley, N.M., Gidaniyan, S.D., Liu, Y., Cascio, D. & Yeates, T.O. (2013). Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. *Protein Sci. Publ. Protein Soc.* **22**, 660–665

27. Tanaka, S., Sawaya, M.R. & Yeates, T.O. (2010). Structure and mechanisms of a protein-based organelle in Escherichia coli. *Science* **327**, 81–84

28. Bobik, T.A., Havemann, G.D., Busch, R.J., Williams, D.S. & Aldrich, H.C. (1999). The Propanediol Utilization (pdu) Operon of Salmonella enterica Serovar Typhimurium LT2 Includes Genes Necessary for Formation of Polyhedral Organelles Involved in Coenzyme B12-Dependent 1,2-Propanediol Degradation. *J. Bacteriol.* **181**, 5967–5975

29. Crowley, C.S., Cascio, D., Sawaya, M.R., Kopstein, J.S., Bobik, T.A. & Yeates, T.O. (2010). Structural insight into the mechanisms of transport across the Salmonella enterica Pdu microcompartment shell. *J. Biol. Chem.* **285**, 37838–37846

30. Cheng, S., Sinha, S., Fan, C., Liu, Y. & Bobik, T.A. (2011). Genetic analysis of the protein shell of the microcompartments involved in coenzyme B12-dependent 1,2-propanediol degradation by Salmonella. *J. Bacteriol.* **193**, 1385–1392

31. Parsons, J.B., Dinesh, S.D., Deery, E., Leech, H.K., Brindley, A.A., Heldt, D., Frank, S., Smales, C.M., Lünsdorf, H., Rambach, A., Gass, M.H., Bleloch, A., McClean, K.J., Munro, A.W., Rigby, S.E.J., Warren, M.J. & Prentice, M.B. (2008). Biochemical and Structural Insights into Bacterial Organelle Form and Biogenesis. *J. Biol. Chem.* **283**, 14366–14375

32. Parsons, J.B., Frank, S., Bhella, D., Liang, M., Prentice, M.B., Mulvihill, D.P. & Warren, M.J. (2010). Synthesis of Empty Bacterial Microcompartments, Directed Organelle Protein Incorporation, and Evidence of Filament-Associated Organelle Movement. *Mol. Cell* **38**, 305–315

33. Tanaka, S., Sawaya, M.R., Phillips, M. & Yeates, T.O. (2009). Insights from multiple structures of the shell proteins from the beta-carboxysome. *Protein Sci. Publ. Protein Soc.* **18**, 108–120

34. Takenoya, M., Nikolakakis, K. & Sagermann, M. (2010). Crystallographic insights into the pore structures and mechanisms of the EutL and EutM shell proteins of the ethanolamine-utilizing microcompartment of Escherichia coli. *J. Bacteriol.* **192**, 6056–6063

35. Pang, A., Warren, M.J. & Pickersgill, R.W. (2011). Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe-4S cluster-binding site. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 91–96

36. Dryden, K.A., Crowley, C.S., Tanaka, S., Yeates, T.O. & Yeager, M. (2009). Two-dimensional crystals of carboxysome shell proteins recapitulate the hexagonal packing of three-dimensional crystals. *Protein Sci.* **18**, 2629–2635

37. Sinha, S., Cheng, S., Sung, Y.W., McNamara, D.E., Sawaya, M.R., Yeates, T.O. & Bobik, T.A. (2014). Alanine Scanning Mutagenesis Identifies an Asparagine-Arginine-Lysine Triad Essential to Assembly of the Shell of the Pdu Microcompartment. *J. Mol. Biol.* doi:10.1016/j.jmb.2014.04.012

38. Heldt, D., Frank, S., Seyedarabi, A., Ladikis, D., Parsons, J.B., Warren, M.J. & Pickersgill, R.W. (2009). Structure of a trimeric bacterial microcompartment shell protein, EtuB, associated with ethanol utilization in Clostridium kluyveri. *Biochem. J.* **423**, 199–207

39. Pang, A., Liang, M., Prentice, M.B. & Pickersgill, R.W. (2012). Substrate channels revealed in the trimeric *Lactobacillus reuteri* bacterial microcompartment shell protein PduB. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 1642–1652

40. Klein, M.G., Zwart, P., Bagby, S.C., Cai, F., Chisholm, S.W., Heinhorst, S., Cannon, G.C. & Kerfeld, C.A. (2009). Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. *J. Mol. Biol.* **392**, 319–333

41. Goodsell, D.S. & Olson, A.J. (2000). Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153

42. Perutz, M.F. (1970). Stereochemistry of cooperative effects in haemoglobin. *Nature* **228**, 726–739

43. Perutz, M.F. (1972). Nature of haem-haem interaction. *Nature* **237**, 495–499

44. Perutz, M.F. (1963). X-ray analysis of hemoglobin. *Science* **140**, 863–869

45. Boyer, P.D. (1993). The binding change mechanism for ATP synthase--some probabilities and possibilities. *Biochim. Biophys. Acta* **1140**, 215–250

46. Boyer, P.D. (2001). Toward an adequate scheme for the ATP synthase catalysis. *Biochem. Biokhimiia* **66**, 1058–1066

47. Prabu-Jeyabalan, M., Nalivaika, E. & Schiffer, C.A. (2000). How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *J. Mol. Biol.* **301**, 1207–1220

48. Brown, J.H., Kim, K.H., Jun, G., Greenfield, N.J., Dominguez, R., Volkmann, N., Hitchcock-DeGregori, S.E. & Cohen, C. (2001). Deciphering the design of the tropomyosin molecule. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8496–8501

49. Iwata, S., Kamata, K., Yoshida, S., Minowa, T. & Ohta, T. (1994). T and R states in the crystals of bacterial L-lactate dehydrogenase reveal the mechanism for allosteric control. *Nat. Struct. Biol.* **1**, 176–185

50. De Vos, A.M., Ultsch, M. & Kossiakoff, A.A. (1992). Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* **255**, 306–312

51. Lassila, J.K., Bernstein, S.L., Kinney, J.N., Axen, S.D. & Kerfeld, C.A. (2014). Assembly of robust bacterial microcompartment shells using building blocks from an organelle of unknown function. *J. Mol. Biol.* **426**, 2217–2228

52. Tsai, S.J. & Yeates, T.O. (2011). Bacterial microcompartments insights into the structure, mechanism, and engineering applications. *Prog. Mol. Biol. Transl. Sci.* **103**, 1–20

53. Howorka, S. (2011). Rationally engineering natural protein assemblies in nanobiotechnology. *Curr. Opin. Biotechnol.* **22**, 485–491

54. Lin, M.T., Occhialini, A., Andralojc, J.P., Devonshire, J., Hines, K.M., Parry, M.A.J. & Hanson, M.R. (2014). β-carboxysomal proteins assemble into highly organized structures in Nicotiana chloroplasts. *Plant J. Cell Mol. Biol.* doi:10.1111/tpj.12536

55. Held, M., Quin, M.B. & Schmidt-Dannert, C. (2013). Eut bacterial microcompartments: insights into their function, structure, and bioengineering applications. *J. Mol. Microbiol. Biotechnol.* **23**, 308–320

56. Fan, C., Cheng, S., Sinha, S. & Bobik, T.A. (2012). Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 14995–15000

57. Fan, C., Cheng, S., Liu, Y., Escobar, C.M., Crowley, C.S., Jefferson, R.E., Yeates, T.O. & Bobik, T.A. (2010). Short N-terminal sequences package proteins into bacterial microcompartments. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 7509–7514

58. Sagermann, M., Ohtaki, A. & Nikolakakis, K. (2009). Crystal structure of the EutL shell protein of the ethanolamine ammonia lyase microcompartment. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 8883–8887

59. Savage, D.C. (1977). Microbial Ecology of the Gastrointestinal Tract. *Annu. Rev. Microbiol.* **31**, 107–133

60. Guarner, F. & Malagelada, J.-R. (2003). Gut flora in health and disease. *Lancet* **361**, 512–519

61. Sears, C.L. (2005). A dynamic partnership: celebrating our gut flora. *Anaerobe* **11**, 247–251

62. Cummings, J.H. & Macfarlane, G.T. (1997). Role of intestinal bacteria in nutrient metabolism. *Clin. Nutr.* **16**, 3–11

63. Steinhoff, U. (2005). Who controls the crowd? New findings and old questions about the intestinal microflora. *Immunol. Lett.* **99**, 12–16

64. Björkstén, B., Sepp, E., Julge, K., Voor, T. & Mikelsaar, M. (2001). Allergy development and the intestinal microflora during the first year of life. *J. Allergy Clin. Immunol.* **108**, 516–520

65. Ley, R.E., Turnbaugh, P.J., Klein, S. & Gordon, J.I. (2006). Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1023

66. Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R. & Gordon, J.I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031

67. Tillisch, K., Labus, J., Kilpatrick, L., Jiang, Z., Stains, J., Ebrat, B., Guyonnet, D., Legrain–Raspaud, S., Trotin, B., Naliboff, B. & Mayer, E.A. (2013). Consumption of Fermented Milk Product With Probiotic Modulates Brain Activity. *Gastroenterology* **144**, 1394–1401.e4

68. Tsoy, O., Ravcheev, D. & Mushegian, A. (2009). Comparative genomics of ethanolamine utilization. *J. Bacteriol.* **191**, 7157–7164

69. Roof, D.M. & Roth, J.R. (1988). Ethanolamine utilization in Salmonella typhimurium. *J. Bacteriol.* **170**, 3855–3863

70. Blackwell, C.M., Scarlett, F.A. & Turner, J.M. (1976). Ethanolamine catabolism by bacteria, including Escherichia coli. *Biochem. Soc. Trans.* **4**, 495–497

71. Cotton, P.B. (1972). Non-dietary lipid in the intestinal lumen. *Gut* **13**, 675–681

72. Randle, C.L., Albro, P.W. & Dittmer, J.C. (1969). The phosphoglyceride composition of Gram-negative bacteria and the changes in composition during growth. *Biochim. Biophys. Acta* **187**, 214–220

73. Li, H., Kristensen, D.M., Coleman, M.K. & Mushegian, A. (2009). Detection of Biochemical Pathways by Probabilistic Matching of Phyletic Vectors. *PLoS ONE* **4**, e5326

74. Korbel, J.O., Doerks, T., Jensen, L.J., Perez-Iratxeta, C., Kaczanowski, S., Hooper, S.D., Andrade, M.A. & Bork, P. (2005). Systematic association of genes to phenotypes by genome and literature mining. *PLoS Biol.* **3**, e134

75. Lawhon, S.D., Frye, J.G., Suyemoto, M., Porwollik, S., McClelland, M. & Altier, C. (2003). Global regulation by CsrA in Salmonella typhimurium. *Mol. Microbiol.* **48**, 1633–1645

76. Kelly, A., Goldberg, M.D., Carroll, R.K., Danino, V., Hinton, J.C.D. & Dorman, C.J. (2004). A global role for Fis in the transcriptional control of metabolism and type III secretion in Salmonella enterica serovar Typhimurium. *Microbiol. Read. Engl.* **150**, 2037–2053

77. Bourgogne, A., Hilsenbeck, S.G., Dunny, G.M. & Murray, B.E. (2006). Comparison of OG1RF and an Isogenic fsrB Deletion Mutant by Transcriptional Analysis: the Fsr System of Enterococcus faecalis Is More than the Activator of Gelatinase and Serine Protease. *J. Bacteriol.* **188**, 2875–2884

78. Winter, S.E., Thiennimitr, P., Winter, M.G., Butler, B.P., Huseby, D.L., Crawford, R.W., Russell, J.M., Bevins, C.L., Adams, L.G., Tsolis, R.M., Roth, J.R. & Bäumler, A.J. (2010). Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature* **467**, 426–429

79. Thiennimitr, P., Winter, S.E., Winter, M.G., Xavier, M.N., Tolstikov, V., Huseby, D.L., Sterzenbach, T., Tsolis, R.M., Roth, J.R. & Bäumler, A.J. (2011). Intestinal inflammation allows Salmonella to use ethanolamine to compete with the microbiota. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 17480–17485

80. Winter, S.E. & Bäumler, A.J. (2011). A breathtaking feat: to compete with the gut microbiota, Salmonella drives its host to provide a respiratory electron acceptor. *Gut Microbes* **2**, 58–60

81. Huseby, D.L. & Roth, J.R. (2013). Evidence that a metabolic microcompartment contains and recycles private cofactor pools. *J. Bacteriol.* **195**, 2864–2879

82. Cheng, S., Fan, C., Sinha, S. & Bobik, T.A. (2012). The PduQ Enzyme Is an Alcohol Dehydrogenase Used to Recycle NAD+ Internally within the Pdu Microcompartment of Salmonella enterica. *PLoS ONE* **7**, e47144

83. Kaplan, B.H. & Stadtman, E.R. (1968). Ethanolamine Deaminase, a Cobamide Coenzyme-dependent Enzyme II. PHYSICAL AND CHEMICAL PROPERTIES AND INTERACTION WITH COBAMIDES AND ETHANOLAMINE. *J. Biol. Chem.* **243**, 1794–1803

84. Mori, K., Bando, R., Hieda, N. & Toraya, T. (2004). Identification of a Reactivating Factor for Adenosylcobalamin-Dependent Ethanolamine Ammonia Lyase. *J. Bacteriol.* **186**, 6845–6854

85. Buan, N.R., Suh, S.-J. & Escalante-Semerena, J.C. (2004). The eutT gene of Salmonella enterica Encodes an oxygen-labile, metal-containing ATP:corrinoid adenosyltransferase enzyme. *J. Bacteriol.* **186**, 5708–5714

86. Karplus, P.A. & Diederichs, K. (2012). Linking crystallographic model and data quality. *Science* **336**, 1030–1033

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

88. Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. & Richardson, D.C. (2009). *MolProbity* : all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21

89. Smart, O.S., Goodfellow, J.M. & Wallace, B.A. (1993). The pore dimensions of gramicidin A. *Biophys. J.* **65**, 2455–2460

90. Burmeister, W.P. (2000). Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 328–341

91. Weik, M., Ravelli, R.B., Kryger, G., McSweeney, S., Raves, M.L., Harel, M., Gros, P., Silman, I., Kroon, J. & Sussman, J.L. (2000). Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 623–628

92. Weik, M., Bergès, J., Raves, M.L., Gros, P., McSweeney, S., Silman, I., Sussman, J.L., Houée-Levin, C. & Ravelli, R.B.G. (2002). Evidence for the formation of disulfide radicals in protein crystals upon X-ray irradiation. *J. Synchrotron Radiat.* **9**, 342–346

93. Garman, E.F. (2010). Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallogr. D Biol. Crystallogr.* **66**, 339–351

94. Mallick, P., Boutz, D.R., Eisenberg, D. & Yeates, T.O. (2002). Genomic evidence that the intracellular proteins of archaeal microbes contain disulfide bonds. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 9679–9684

95. Jorda, J. & Yeates, T.O. (2011). Widespread disulfide bonding in proteins from thermophilic archaea. *Archaea Vanc. BC* **2011**, 409156

96. Juers, D.H. & Matthews, B.W. (2001). Reversible lattice repacking illustrates the temperature dependence of macromolecular interactions. *J. Mol. Biol.* **311**, 851–862

97. Fraser, J.S., van den Bedem, H., Samelson, A.J., Lang, P.T., Holton, J.M., Echols, N. & Alber, T. (2011). Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16247–16252

98. Burnley, B.T., Afonine, P.V., Adams, P.D. & Gros, P. (2012). Modelling dynamics in protein crystal structures by ensemble refinement. *eLife* **1**, e00311–e00311

99. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, W529–W533

100. Glaser, F., Pupko, T., Paz, I., Bell, R.E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003). ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. *Bioinformatics* **19**, 163–164

101. Pellequer, J.-L. & Chen, S.W. (2006). Multi-template approach to modeling engineered disulfide bonds. *Proteins* **65**, 192–202

102. Marino, S.M. & Gladyshev, V.N. (2012). Analysis and functional prediction of reactive cysteine residues. *J. Biol. Chem.* **287**, 4419–4425

103. Crowley, C.S., Sawaya, M.R., Bobik, T.A. & Yeates, T.O. (2008). Structure of the PduU shell protein from the Pdu microcompartment of Salmonella. *Struct. Lond. Engl. 1993* **16**, 1324–1332

104. Rasmussen, B.F., Stock, A.M., Ringe, D. & Petsko, G.A. (1992). Crystalline ribonuclease A loses function below the dynamical transition at 220 K. *Nature* **357**, 423–424

105. Tilton, R.F., Jr, Dewan, J.C. & Petsko, G.A. (1992). Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320 K. *Biochemistry (Mosc.)* **31**, 2469–2481

106. Halle, B. (2004). Biomolecular cryocrystallography: Structural changes during flash-cooling. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4793–4798

107. Drennan, C.L., Matthews, R.G. & Ludwig, M.L. (1994). Cobalamin-dependent methionine synthase: the structure of a methylcobalamin-binding fragment and implications for other B12-dependent enzymes. *Curr. Opin. Struct. Biol.* **4**, 919–929

108. Toraya, T. (2003). Radical Catalysis in Coenzyme B12-Dependent Isomerization (Eliminating) Reactions. *Chem. Rev.* **103**, 2095–2128

109. Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467

110. Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J. & Klenk, D.C. (1985). Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **150**, 76–85

111. Otwinowski, Z. & Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* **276**, 307–326

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

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

114. P H Zwart, R.W.G.-K. Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation.

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

116. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A. & Wilson, K.S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242

117. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501

118. Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H. & Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367

119. Moriarty, N.W., Grosse-Kunstleve, R.W. & Adams, P.D. (2009). *electronic Ligand Builder and Optimization Workbench* (*eLBOW*): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074–1080

120. Roy, A., Kucukural, A. & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738

121. Boutz, D.R., Cascio, D., Whitelegge, J., Perry, L.J. & Yeates, T.O. (2007). Discovery of a thermophilic protein complex stabilized by topologically interlinked chains. *J. Mol. Biol.* **368**, 1332–1344

122. Schägger, H. (2006). Tricine-SDS-PAGE. *Nat. Protoc.* **1**, 16–22

123. Whitelegge, J.P., Coutre, J. le, Lee, J.C., Engel, C.K., Privé, G.G., Faull, K.F. & Kaback, H.R. (1999). Toward the bilayer proteome, electrospray ionization-mass spectrometry of large, intact transmembrane proteins. *Proc. Natl. Acad. Sci.* **96**, 10695–10698

124. Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L.L., Eddy, S.R., Bateman, A. & Finn, R.D. (2011). The Pfam protein families database. *Nucleic Acids Res.* **40**, D290–D301

125. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542

126. Baker, N.A., Sept, D., Joseph, S., Holst, M.J. & McCammon, J.A. (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci.* **98**, 10037–10041

127. Tsai, Y., Sawaya, M.R., Cannon, G.C., Cai, F., Williams, E.B., Heinhorst, S., Kerfeld, C.A. & Yeates, T.O. (2007). Structural Analysis of CsoS1A and the Protein Shell of the Halothiobacillus neapolitanus Carboxysome. *PLoS Biol* **5**, e144

128. Samborska, B. & Kimber, M.S. (2012). A Dodecameric CcmK2 Structure Suggests β-Carboxysomal Shell Facets Have a Double-Layered Organization. *Structure* **20**, 1353–1362

129. Zwart, P.H., Grosse-Kunstleve, R.W., Lebedev, A.A., Murshudov, G.N. & Adams, P.D. (2008). Surprises and pitfalls arising from (pseudo)symmetry. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 99–107

130. Tsai, Y., Sawaya, M.R. & Yeates, T.O. (2009). Analysis of lattice-translocation disorder in the layered hexagonal structure of carboxysome shell protein CsoS1C. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 980–988

131. Diederichs, K. & Karplus, P.A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Mol. Biol.* **4**, 269–275

132. Vagin, A. & Teplyakov, A. (1997). *MOLREP* : an Automated Program for Molecular Replacement. *J. Appl. Crystallogr.* **30**, 1022–1025

133. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1131–1137

134. Patterson, A.L. (1934). A Fourier Series Method for the Determination of the Components of Interatomic Distances in Crystals. *Phys. Rev.* **46**, 372–376

135. Padilla, J.E. & Yeates, T.O. (2003). A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1124–1130

136. Painter, J. & Merritt, E.A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 439–450

137. Painter, J. & Merritt, E.A. (2006). *TLSMD* web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **39**, 109–111

138. Iwasaki, H. (1972). On the diffraction enhancement of symmetry. *Acta Crystallogr. Sect. A* **28**, 253–260

139. Sadanaga, R. & Ohsumi, K. (1979). Basic theorems of vector symmetry in crystallography. *Acta Crystallogr. Sect. A* **35**, 115–122

140. Poon, B.K., Grosse-Kunstleve, R.W., Zwart, P.H. & Sauter, N.K. (2010). Detection and correction of underassigned rotational symmetry prior to structure deposition. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 503–513

141. Robbins, A.H., Domsic, J.F., Agbandje-McKenna, M. & McKenna, R. (2010). Structure of a monoclinic polymorph of human carbonic anhydrase II with a doubled a axis. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 628–634

142. Robbins, A.H., Domsic, J.F., Agbandje-McKenna, M. & McKenna, R. (2010). Emerging from pseudo-symmetry: the redetermination of human carbonic anhydrase II in monoclinic P2(1) with a doubled a axis. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 950–952

143. Lebedev, A.A. & Isupov, M.N. Space group validation with Zanuda. *CCP4 Newsl.* at <http://www.ccp4.ac.uk/newsletters/newsletter48/articles/Zanuda/zanuda.html>

144. J R Vercellotti, A.A.S. (1977). Breakdown of mucin and plant polysaccharides in the human colon. *Can. J. Biochem.* **55**, 1190–6

145. Obradors, N., Badía, J., Baldomà, L. & Aguilar, J. (1988). Anaerobic metabolism of the L-rhamnose fermentation product 1,2-propanediol in Salmonella typhimurium. *J. Bacteriol.* **170**, 2159–2162

146. Petit, E., LaTouf, W.G., Coppi, M.V., Warnick, T.A., Currie, D., Romashko, I., Deshpande, S., Haas, K., Alvelo-Maurosa, J.G., Wardman, C., Schnell, D.J., Leschine, S.B. & Blanchard, J.L. (2013). Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by Clostridium phytofermentans. *PloS One* **8**, e54337

147. Saxena, R.K., Anand, P., Saran, S., Isar, J. & Agarwal, L. (2010). Microbial production and applications of 1,2-propanediol. *Indian J. Microbiol.* **50**, 2–11

148. Frey, P.A. (2001). Radical Mechanisms of Enzymatic Catalysis1. *Annu. Rev. Biochem.* **70**, 121–148
149. Cheng, S. & Bobik, T.A. (2010). Characterization of the PduS Cobalamin Reductase of Salmonella enterica and Its Role in the Pdu Microcompartment. *J. Bacteriol.* **192**, 5071–5080

150. Sampson, E.M., Johnson, C.L.V. & Bobik, T.A. (2005). Biochemical evidence that the pduS gene encodes a bifunctional cobalamin reductase. *Microbiol. Read. Engl.* **151**, 1169–1177

151. Dung, J.K.S., Johnson, D.A. & Schroeder, B.K. (2012). First Report of *Pectobacterium wasabiae* Causing Aerial Stem Rot of Potato in Washington State. *Plant Dis.* **96**, 1819–1819

152. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410

153. Pitts, A.C., Tuck, L.R., Faulds-Pain, A., Lewis, R.J. & Marles-Wright, J. (2012). Structural Insight into the Clostridium difficile Ethanolamine Utilisation Microcompartment. *PLoS ONE* **7**, e48360

154. Cai, F., Sutter, M., Cameron, J.C., Stanley, D.N., Kinney, J.N. & Kerfeld, C.A. (2013). The structure of CcmP, a tandem bacterial microcompartment domain protein from the β-carboxysome, forms a subcompartment within a microcompartment. *J. Biol. Chem.* **288**, 16055–16063

155. Frey, M., Rothe, M., Wagner, A.F. & Knappe, J. (1994). Adenosylmethionine-dependent synthesis of the glycyl radical in pyruvate formate-lyase by abstraction of the glycine C-2 pro-S hydrogen atom. Studies of [2H]glycine-substituted enzyme and peptides homologous to the glycine 734 site. *J. Biol. Chem.* **269**, 12432–12437

156. Hoover, D.M. & Lubkowski, J. (2002). DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **30**, e43

157. Prodromou, C. & Pearl, L.H. (1992). Recursive PCR: a novel technique for total gene synthesis. *Protein Eng.* **5**, 827–829

158. Terwilliger, T.C., DiMaio, F., Read, R.J., Baker, D., Bunkoczi, G., Adams, P.D., Grosse-Kunstleve, R.W., Afonine, P.V. & Echols, N. (2012). phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. *J. Struct. Funct. Genomics* **13**, 81–90

159. Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S.M. & Bricogne, G. (2004). Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2210–2221

160. Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F. & Vagin, A.A. (2011). *REFMAC* 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367

161. Holm, L. & Rosenström, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549

162. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797

163. Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.-Y., Pieper, U. & Sali, A. (2007). Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci. Editor. Board John E Coligan Al* **Chapter 2**, Unit 2.9

164. Simons, K.T., Ruczinski, I., Kooperberg, C., Fox, B.A., Bystroff, C. & Baker, D. (1999). Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins* **34**, 82–95

165. Brunger, A.T. (2007). Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* **2**, 2728–2733

166. Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. & Warren, G.L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921

167. Wu, S. & Zhang, Y. (2008). MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins* **72**, 547–556

168. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, D277–D280

169. Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321

170. Feynman, R.P., Leighton, R.B. & Sands, M.L. (Addison-Wesley Pub. Co.: Reading, Mass., 1963). *The Feynman lectures on physics*.

171. Hilser, V.J., García-Moreno E., B., Oas, T.G., Kapp, G. & Whitten, S.T. (2006). A Statistical Thermodynamic Model of the Protein Ensemble. *Chem. Rev.* **106**, 1545–1558

172. Frauenfelder, H., Sligar, S.G. & Wolynes, P.G. (1991). The energy landscapes and motions of proteins. *Science* **254**, 1598–1603

173. Fraser, J.S. & Jackson, C.J. (2011). Mining electron density for functionally relevant protein polysterism in crystal structures. *Cell. Mol. Life Sci. CMLS* **68**, 1829–1841

174. Koshland, D.E. (1998). Conformational changes: How small is big enough? *Nat. Med.* **4**, 1112–1114

175. Fersht, A. (W. H. Freeman: New York, 1998). *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*.

176. Karplus, M. & Kuriyan, J. (2005). Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6679–6685

177. Eisenmesser, E.Z., Millet, O., Labeikovsky, W., Korzhnev, D.M., Wolf-Watz, M., Bosco, D.A., Skalicky, J.J., Kay, L.E. & Kern, D. (2005). Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117–121

178. Fraser, J.S., Clarkson, M.W., Degnan, S.C., Erion, R., Kern, D. & Alber, T. (2009). Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669–673

179. Boehr, D.D., McElheny, D., Dyson, H.J. & Wright, P.E. (2006). The Dynamic Energy Landscape of Dihydrofolate Reductase Catalysis. *Science* **313**, 1638–1642

180. Hilser, V.J., Wrabl, J.O. & Motlagh, H.N. (2012). Structural and Energetic Basis of Allostery. *Annu. Rev. Biophys.* **41**, 585–609

181. Gunasekaran, K., Ma, B. & Nussinov, R. (2004). Is allostery an intrinsic property of all dynamic proteins? *Proteins* **57**, 433–443

182. Li, P., Martins, I.R.S., Amarasinghe, G.K. & Rosen, M.K. (2008). Internal dynamics control activation and activity of the autoinhibited Vav DH domain. *Nat. Struct. Mol. Biol.* **15**, 613–618

183. Huse, M. & Kuriyan, J. (2002). The conformational plasticity of protein kinases. *Cell* **109**, 275–282

184. Warshaw, D.M., Hayes, E., Gaffney, D., Lauzon, A.-M., Wu, J., Kennedy, G., Trybus, K., Lowey, S. & Berger, C. (1998). Myosin conformational states determined by single fluorophore polarization. *Proc. Natl. Acad. Sci.* **95**, 8034–8039

185. Tokuriki, N. & Tawfik, D.S. (2009). Protein Dynamism and Evolvability. *Science* **324**, 203–207

186. Fenwick, R.B., van den Bedem, H., Fraser, J.S. & Wright, P.E. (2014). Integrated description of protein dynamics from room-temperature X-ray crystallography and NMR. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E445–454

187. Lang, P.T., Ng, H.-L., Fraser, J.S., Corn, J.E., Echols, N., Sales, M., Holton, J.M. & Alber, T. (2010). Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Sci. Publ. Protein Soc.* **19**, 1420–1431

188. Levin, E.J., Kondrashov, D.A., Wesenberg, G.E. & Phillips, G.N., Jr (2007). Ensemble refinement of protein crystal structures: validation and application. *Struct. Lond. Engl. 1993* **15**, 1040–1052

189. Van den Bedem, H., Dhanik, A., Latombe, J.-C. & Deacon, A.M. (2009). Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1107–1117

190. Ren, Z., Chan, P.W.Y., Moffat, K., Pai, E.F., Royer, W.E., Jr, Šrajer, V. & Yang, X. (2013). Resolution of structural heterogeneity in dynamic crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 946–959

191. Artiukh, R.I., Kachalova, G.S., Lanina, N.F., Nikol'skiĭ, D.O., Timofeev, V.P. & Bartunik, K.D. (2002). [Effect of various humidity on local dynamic structure of lysozyme in a spin-labeled tetragonal crystal]. *Biofizika* **47**, 795–805

192. Harata, K. & Akiba, T. (2004). Phase transition of triclinic hen egg-white lysozyme crystal associated with sodium binding. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 630–637

193. Rupp, B. (Garland Science: New York, 2009). *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*.

194. Evans, G., Axford, D., Waterman, D. & Owen, R.L. (2011). Macromolecular microcrystallography. *Crystallogr. Rev.* **17**, 105–142