Title
Selective Transport and Targeted Assembly in the 1,2-Propanediol Bacterial Microcompartment

Author
Chang, Sunny Chun

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Selective Transport and Targeted Assembly in the 1,2-Propanediol Bacterial Microcompartment

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Molecular Biology

by

Sunny Chun Chang

2016

ABSTRACT OF THE DISSERTATION

Selective Transport and Targeted Assembly in the
1,2-Propanediol Bacterial Microcompartment

by

Sunny Chun Chang
Doctor of Philosophy in Molecular Biology
University of California, Los Angeles, 2016
Professor Todd O. Yeates, Chair

This dissertation is the culmination of my graduate studies in the laboratory of Todd O. Yeates at UCLA. The research presented here is a study of 1,2-propanediol utilization (Pdu), a scavenger pathway used by common gut bacteria to thrive in the human gut environment. Encapsulating the Pdu pathway is a novel non-membrane, proteinaceous shell (approximately 100-200 nm in diameter) also known as a bacterial microcompartment (BMC) and the focus of investigation in the present work. BMCs are a conserved mechanism for housing metabolic processes that involve volatile or toxic intermediates. They are found in approximately 20% of sequenced bacterial genomes. However, little is known about BMC properties for small molecule transport and assembly. My dissertation work revealed important aspects of selective transport and shell protein organization for the Pdu BMC and other BMC shell proteins through hypothesis-driven research.

As an introduction to this dissertation, chapter 1 summarizes the history of research on Pdu BMCs and recent applications in biotechnology. Chapter 2 is a comprehensive review, reprinted with permission from Microbiology and Molecular Biology Reviews (see Acknowledgments), of diverse bacterial microcompartments of known function and their possible applications in bioengineering of fuel and drug biosynthesis. Chapter 3 is an exposition on biochemical and structural characterization on selective transport of small molecules in the shell protein PduA, testing my first hypothesis about substrate entry and toxic intermediate encapsulation. This article is reprinted with permission from Proceedings of the National Academy of Sciences (see Acknowledgments). To follow up on the results of Chowdhury, Chun, et al. (2015), Chapter 4 presents a molecular dynamics approach to study free energy barriers to small molecules through the shell protein PduA, which supported our previous conclusions. This manuscript is in submission for journal peer review.

Another type of BMC shell protein, called EutL, is a promising candidate for pore-conducting small molecule transport. In Chapter 5, I describe molecular dynamics studies on EutL, previously reported by several groups in open and closed pore conformations by X-ray crystallography, in order to observe the large structural rearrangements required for conformational transition. Chapter 6 reports on the study of homologous shell protein, PduB, that I hypothesized can also have an open pore structure. Here, I used Tryptophan emission spectroscopy and X-ray crystallography to test this hypothesis. I outline future work for the continuation of this project.

Lastly, the latter part of my dissertation focuses on questions of BMC shell assembly, a difficult topic of study due to non-uniform distributions of size and shape among BMCs of a

particular system and highly redundant motifs in the BMC shell. Chapter 7 details the structural and *in vivo* studies of the shell protein PduJ that has 80% amino acid sequence identity to PduA. However, PduJ is found to not be functionally synonymous with PduA and its genic location in the Pdu operon may affect its post-translational assembly. This research was published electronically ahead of print in Molecular Microbiology (June 2016) and is reprinted here with permission (see Acknowledgments).

Finally, Chapter 8 chronicles the study of Pdu enzyme N-terminal peptides binding Pdu BMC shell proteins for two reasons. First, the literature on this subject contributed by many research groups is sometimes inconsistent, which may be attributed to the difficulty of studying amphipathic peptides in a biochemical setting. A thorough study of the Pdu enzyme N-terminal peptides using biophysical chemistry has not been carried out prior to this work and would benefit the research community. Second, a more quantitative analysis could be used to mathematically model Pdu BMC assembly and, in combination with data on pore permeability (described in chapter 4) and enzyme kinetics, accurately simulate production efficiency of the Pdu BMC. This information is highly valuable for the industrial scale use of Pdu BMCs, the bioengineering and synthetic biology of which is already an active area of research. I outline the future work for the continuation of this project, with notes in the Appendix, and offer advice for using different techniques.

In conclusion, this dissertation work contributes significant findings to the expanding knowledge of the Pdu BMC and details further studies of interest for posterity in the BMC research community.

The dissertation of Sunny Chun Chang is approved.

Linda G. Baum

Kendall N. Houk

Feng Guo

Todd O. Yeates, Committee Chair

To my loving family.

Table of Contents

Acknowledgements

I would like to make the following acknowledgements to recognize the people and organizations that made the work presented in this dissertation possible. Please note that my maiden name Sunny Chun is used in the publications below.

Chapter 2 contains the reprinted article from Microbiology and Molecular Biology Reviews: Chiranjit Chowdhury, Sharmistha Sinha, Sunny Chun, Todd O. Yeates, Thomas A. Bobik. Diverse Bacterial Microcompartment Organelles. *Microbiol. Mol. Biol. Rev*. 2014, 78(3):438–468. DOI: 10.1128/MMBR.00009-14. All authors contributed content to the article.

Chapter 3 contains the reprinted article from the Proceedings of the National Academy of Sciences: Chiranjit Chowdhury*, Sunny Chun*, Allan Pang, Michael R. Sawaya, Sharmistha Sinha, Todd O. Yeates, Thomas A. Bobik. Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proc. Natl. Acad. Sci*. 2015, 112(10):2990–2995. DOI: 10.1073/pnas.1423672112. Chiranjit Chowdhury, Sharmistha Sinha, and Thomas A. Bobik contributed *in vivo* growth assays of chromosomal mutants and *in vitro* characterizations of purified microcompartments. Sunny Chun, Allan Pang, Michael R. Sawaya and Todd O. Yeates contributed structural and biochemical analysis of the mutant proteins studied in this work.

Chapter 4 contains the unpublished manuscript: Sunny Chun, Jiyong Park, Kendall N. Houk, Todd O. Yeates. Molecular Dynamics Simulations of Selective Metabolite Transport Across the Propaediol Bacterial Microcompartment Shell. 2016 (in submission). Sunny Chun, Jiyong Park, Kendall N. Houk, and Todd O. Yeates contributed molecular dynamic simulations and free energy calculations to this work.

Chapter 5 studies were conducted by Sunny Chun in the laboratory of J. Andrew McCammon at University of California, San Diego under the guidance of postdoctoral fellow Changsun Eun.

Chapter 6 studies were conducted by Sunny Chun, Qiaojun Chen, and Jeannette Bowler. Qiaojun Chen was an undergraduate research fellow who cloned and prepared recombinant fusion proteins for X-ray crystallography. Jeannette Bowler was a graduate rotation student who cloned and prepared mutant recombinant fusion proteins and carried out Tryptophan emission spectroscopy studies. Thomas A. Bobik and graduate student Brent Lehman contributed plasmid constructs for this work. Sunny Chun solved the crystal structure of one mutation construct.

Chapter 7 contains the e-published article reprinted with permission from Molecular Microbiology: Chiranjit Chowdhury, Sunny Chun, Michael R. Sawaya, Todd O. Yeates, Thomas A. Bobik. The function of the PduJ microcompartment shell protein is determined by the genomic position of its encoding gene. *Mol. Microbiol*. 2016 (epublished ahead of print). DOI: 10.1111/mmi.13423. Chiranjit Chowdhury and Thomas A. Bobik contributed *in vivo* growth assays of chromosomal mutants and *in vitro* characterizations of purified microcompartments. Sunny Chun, Michael R. Sawaya and Todd O. Yeates contributed structural and biochemical analysis of the mutant proteins studied in this work.

Chapter 8 studies were carried out by Sunny Chun. I would like to acknowledge Andrew Ah Young Pati and Christina Vizcarra for initial discussions; Lisa Johnson, Dan Anderson, and Xin Cong for guidance on peptide labeling, reverse HPLC purification, and MALDI-MS analysis; Martin Phillips and Mathew Graf at the Biochemistry Instrumentation Facility at UCLA for help with binding assays and instruments.

I would also like to acknowledge Department of Energy Crystallization Core and X-ray Core Facility scientists Michael Collazo, Duilio Cascio, Michael Sawaya; Northeastern Collaborative Access Team Beamline 24-ID-C at the Advanced Photon Source, Argonne National Laboratory scientists M. Capel, K. Rajashankar, N. Sukumar, J. Schuermann, I. Kourinov, and F. Murphy; Department of Energy Cassini Cluster staff members Duilio Cascio, Alex Lisker, Derek Pitman, Luki Goldshmidt; the Hoffman2 cluster at the Institute for Digital Research and Education, UCLA; the Stampede Supercomputer at the Texas Advanced Computing Center. The laboratory of Todd O. Yeates is supported by the Biological and Environmental Research program of the Department of Energy Office of Science.

I would like to thank my committee members, Drs. Margot E. Quinlan, Linda G. Baum, Kendall N. Houk, Feng Guo, Todd O.Yeates for supervising my research progress and troubleshooting my experiments. In addition to my committee members, the following faculty were influential in my doctoral studies through discussions at laboratory group meetings: David Eisenberg, Pascal Egea, and James Bowie. I would like to thank the director of the National Institutes of Health (NIH) Chemistry-Biology Interface (CBI) Training Program, Heather Maynard, for research and professional advice and J. Andrew McCammon at University of California, San Diego for allowing me the resources to carry out my CBI internship in his laboratory.

I would like to acknowledge the Molecular Biology Interdepartmental-PhD Program, the NIH Chemistry-Biology Interface Training Program (Grant T32-GM008496), and the University of California, Los Angeles Graduate Division for their support of my doctoral studies.

*These authors contributed equally to this work.

Biographical Sketch

California Institute of Technology	B.S. Chemistry, 2010
City of Hope, Beckman Research Institute	Research Associate I, 2011
University of California, Los Angeles	Chemistry-Biology Interface NIH T32 Predoctoral Training Fellowship, 2012-2015
University of California, Los Angeles	Paul D. Boyer Outstanding Teaching Award, 2013
The Protein Society	Graduate Student Ambassador for the Protein Society Annual Symposium, 2014
Review Article	Chiranjit Chowdhury, Sharmistha Sinha, **Sunny Chun**, Todd O. Yeates, Thomas A. Bobik. Diverse Bacterial Microcompartment Organelles. *Microbiol. Mol. Biol. Rev*. 2014, 78(3):438–468. DOI: 10.1128/MMBR.00009-14.
Research Article	Chiranjit Chowdhury*, **Sunny Chun***, Allan Pang, Michael R. Sawaya, Sharmistha Sinha, Todd O. Yeates, Thomas A. Bobik. Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proc. Natl. Acad. Sci*. 2015, 112(10):2990–2995. DOI: 10.1073/pnas.1423672112.

Research Article	Chiranjit Chowdhury, **Sunny Chun**, Michael R. Sawaya, Todd O. Yeates, Thomas A. Bobik. The function of the PduJ microcompartment shell protein is determined by the genomic position of its encoding gene. *Mol. Microbiol*. 2016 (epublished ahead of print). DOI: 10.1111/mmi.13423.
Research Article	**Sunny Chun**, Jiyong Park, Kendall N. Houk, Todd O. Yeates. Molecular Dynamics Simulations of Selective Metabolite Transport Across the Propaediol Bacterial Microcompartment Shell. 2016 (in submission)

*These authors contributed equally to this work.

Chapter 1

Introduction to 1,2-Propanediol Utilizing Microcompartments

Anaerobic metabolism of diols by animals, yeast, and bacteria, has been a growing research interest since the 1960s. Particularly, 1,2-propanediol and 1,2-ethanediol are fascinating because they are reduced substrates, compared to carbohydrates, and can be used as sole carbon and energy sources by certain enterobacteria[1].

The 1,2-propanediol utilization (*pdu*) operon and the associated *cob* operon (cobalamin biosynthesis) are induced in *trans* by 1,2-propanediol via the transcriptional activator PocR[2,3]. 1,2-propanediol is a fermentation product of common plant sugars, such as *L*-rhamnose[4,5] and *L*-fucose[6,7]. Fucose is also one of the 4 main substituents of intestinal mucins. The catabolism of 1,2-propanediol produces half-molar equivalents of *n*-propanol and propionate and only occurs under anaerobic conditions[4], by the induction of B12-dependent diol dehydratase[1]. The yield of this anaerobic metabolism of 1,2-propanediol is limited to a gain of 1 mol of ATP per 2 mol of substrate consumed[1].

In 1999, Bobik et al. completed sequencing of the *pdu* operon in *Salmonella enterica* serovar Typhimurium LT2 and showed that it includes sequences homologous to bacterial microcompartments (BMC) by conserved BMC-fold domains[8]. This was confirmed by the observation of polyhedral bodies in *S. enterica* during growth only on 1,2-propanediol and by the localization of B_{12}-dependent diol dehydratase to the polyhedra. This was a surprising finding that demonstrated the complexity of bacterial systems that use this vitamin B_{12}-dependent pathway. It gave rise to the obvious question, why would 1,2-propanediol utilization be compartmentalized in BMCs?

BMCs are proteinaceous organelles of prokaryotic origin. They demonstrate a step towards organizational complexity, and thus, evolution to higher orders of life. Approximately 20% of bacteria carry BMC operons of at least seven metabolic types, including carbon fixation, propanediol and ethanolamine utilization[9]. They act as natural bioreactors by providing compartmentalization which results in one or more benefits for the system: (1) a concentrating mechanism for enzymes, cofactors, substrates, (2) the ability to channel intermediates to subsequent reaction sites, (3) protection for cytoplasmic components against toxic or volatile intermediates, (4) increase in enzymatic efficiencies, and (5) release of desired products. Following the initial evidence of *pdu*-associated bacterial microcompartments (Pdu BMCs), biochemical and structural characterization of the Pdu BMC has confirmed some of these aspects of BMCs.

The Pdu BMC are proteinaceous polyhedral bodies that are 100 to 150 nm in diameter with a shell thickness of 3 to 4 nm[8,10] . The major protein contents from purified BMCs were determined to be PduABB'CDEGHJKOPTU where PduB' is a truncation of PduB by 37 N-terminal amino acids. Isoelectric focusing of dodecyl-maltoside treated BMCs indicated that PduJKP may have post-translational modifications. The relative abundances of Pdu BMC shell proteins PduABB'JKTU was determined to be J:A:B:B':U:K:T = 15:10:7:6:2:1:1[10].

In the metabolic pathway of 1,2-propanediol utilization, first, 1,2-propanediol is converted to propionaldehyde by PduCDE, the Ado-B_{12}-dependent diol dehydratase[10]. Propionaldehyde can then be reduced to propanol by PduQ for the regeneration of NAD or oxidized to propionyl-coA by PduP, propionaldehyde dehydrogenase, which is further converted to propionyl-phosphate and then to propionic acid for entry in the citric acid cycle, which aids in

bacterial Krebs (or tricarboxylic acid) cycle, for the generation of ATP and cell carbon. PduO, adenosyltransferase, and PduGH, diol dehydratase reactivating factor, reactivate diol dehydratase. Propionaldehyde accumulates to toxic levels in BMC-minus mutants of *S. enterica* grown on 1,2-PD, resulting in 20h growth arrest[11]. The Ames test showed that propionaldehyde is a mutagen[11]. Thus, the Pdu BMC primarily serves to sequester propionaldehyde and protect cells from cytotoxicity by DNA damage.

Several groups have described the importance of the shell protein PduA to propanediol metabolism in the Pdu BMC[12]. *In vivo*, deletion of PduA disrupts BMC formation.[13] The structure of PduA features a pore in the hexamer assembly of radius 2.8Å[14]. Thus, PduA is predicted to be a major route of entry and exit into the Pdu BMC. PduJ is the most abundant protein but the function is yet unknown. It shares 77.7% sequence identity to PduA.[10]

The Pdu BMC can be categorized as a "metabolosome," a body housing several related metabolic processes of a pathway. A metabolosome can be engineered, for example, for biofuel and green chemical production, drug biosynthesis, and a variety of other functional pathways that would benefit in improved overall kinetics from compartmentalization and optimized substrate channeling via high density enzyme scaffolds (reviewed[15,16]). To this end, several groups have envisioned and demonstrated the first steps of engineering the Pdu BMC. Parsons et al. (2008) transformed a laboratory K12 *E. coli* strain with the 21 gene Pdu operon from *Citrobacter freundii*, which resulted in the expression of fully functional Pdu BMCs in a non-native host[12]. Their next proof of concept (2010) was to express the empty BMC shell by the induced overexpression of only the shell components PduABJKNTU[17]. They showed that PduAJKN were the essential structural components of fully formed BMCs, where gene order

affects proper assembly, and that PduV may be required for axial distribution of BMCs via interactions with the bacterial cytoskleton.

More recently, Sargent et al. (2013) engineered empty BMC shell expression in *E. coli* from Pdu BMC shell components PduABJKNTU from *Salmonella* enterica. Puncta corresponding to GFP-tagged PduD, a Pdu BMC cargo protein, were observed by fluorescence confocal microscopy only in the presence of the shell components indicating the formation of Pdu BMCs. However, the stoichiometry of shell components did not match that of the wildtype system and PduK barely expressed, so the quality of the synthetic BMCs could not be guaranteed[18].

To control the uniformity of the sample, a different approach was taken by Lassila et al. (2014) taking BMCs from an operon of unknown function in *Haliangium ochraceum*; the purified BMCs were homogenous and small (approximately 80 nm in diameter)[19]. The uniformity of these BMCs was ideal for engineering a controllable system. Lassila et al. designed a synthetic operon from the minimal components of the BMC of unknown function from *Haliangium ochraceum* and demonstrated that they could purify homogenous minimal BMCs.

In another example of engineering pre-existing BMC genes, Cai et al. (2014) engineered hybrid carboxysomes by interchanging pore residues of single-BMC-fold domain proteins, for example, CcmK2 and CcmK4, and producing carboxysomes that incorporated both α- and β-type BMC shell proteins[20].

Some obstacles to engineering BMCs that have been identified include targeting proteins of interest to the BMC lumen; a lack of information on how substrates enter, products leave, and intermediates are retained; and the lack of identification of the minimal components of a

functional BMC[21]. Our laboratory's research explores these aspects of BMCs. The research presented here focuses on the elucidation of the route of entry and exit of small molecules through the Pdu BMC shell and the investigation of a selective diffusion bias between substrate v. intermediate metabolites through the Pdu BMC shell. As a follow-up of the results of this research, I also present on-going work and future projects related to the study of binding between Pdu enzyme Ntail peptides and Pdu shell proteins.

Chapter 2

Diverse Bacterial Microcompartment Organelles

Introduction

The following article is a comprehensive review on the diversity of BMCs, titled ‘Diverse Bacterial Microcompartment Organelles’ and published in Microbiology and Molecular Biology Reviews. The high prevalence of BMCs in bacteria in the environment and in the human microbiome underscores their importance and significant relationship with global and human health. However, our limited knowledge of BMCs and how they function confines our understanding of their dynamic role and ability to thrive in unique and highly competitive microenvironments. Here, we emphasize the need for further studies on the characterization of unknown BMC systems and on the assembly and function of BMC shells, a topic that requires structural determination and biochemical investigation.

MMBR
Journals.ASM.org

Diverse Bacterial Microcompartment Organelles

Chiranjit Chowdhury,[a] Sharmistha Sinha,[a] Sunny Chun,[b] Todd O. Yeates,[b,c,d] Thomas A. Bobik[a]

Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, Iowa, USA[a]; Molecular Biology Institute,[b] Department of Energy Institute for Genomics and Proteomics,[c] and Department of Chemistry and Biochemistry,[d] University of California, Los Angeles, Los Angeles, California, USA

(continued)

SUMMARY

Bacterial microcompartments (MCPs) are sophisticated protein-based organelles used to optimize metabolic pathways. They consist of metabolic enzymes encapsulated within a protein shell, which creates an ideal environment for catalysis and facilitates the channeling of toxic/volatile intermediates to downstream enzymes. The metabolic processes that require MCPs are diverse and widely distributed and play important roles in global carbon fixation and bacterial pathogenesis. The protein shells of MCPs are thought to selectively control the movement of enzyme cofactors, substrates, and products (including toxic or volatile intermediates) between the MCP interior and the cytoplasm of the cell using both passive electrostatic/steric and dynamic gated mechanisms. Evidence suggests that specialized shell proteins conduct electrons between the cytoplasm and the lumen of the MCP and/or help rebuild damaged iron-sulfur centers in the encapsulated enzymes. The MCP shell is elaborated through a family of small proteins whose structural core is known as a bacterial microcompartment (BMC) domain. BMC domain proteins oligomerize into flat, hexagonally shaped tiles, which assemble into extended protein sheets that form the facets of the shell. Shape complementarity along the edges allows different types of BMC domain proteins to form mixed sheets, while sequence variation provides functional diversification. Recent studies have also revealed targeting sequences that mediate protein encapsulation within MCPs, scaffolding proteins that organize lumen enzymes and the use of private cofactor pools (NAD/H and coenzyme A [HS-CoA]) to facilitate cofactor homeostasis. Although much remains to be learned, our growing understanding of MCPs is providing a basis for bioengineering of protein-based containers for the production of chemicals/pharmaceuticals and for use as molecular delivery vehicles.

INTRODUCTION

Bacterial microcompartments (MCPs) were first observed in the 1950s by electron microscopy (EM) as polyhedral inclusion bodies in the cytoplasm of cyanobacteria (1). The structures seen were about 100 to 200 nm in diameter with a thin protein shell and an electron-dense lumen. Initially, they were thought to be viruses, but later studies with *Thiobacillus* showed that the protein shell was filled with RuBisCO and that these structures play an important role in CO_2 fixation by the Calvin cycle (2). Because of their role in CO_2 fixation, this class of MCPs (which is the archetype of the family) was named the carboxysome (2). Further studies of carboxysome composition paved the way for the identification of diverse MCPs (3). Homologs of carboxysome shell proteins were found in diverse contexts, including operons used for 1,2-propanediol (1,2-PD) and ethanolamine degradation by *Salmonella* (4, 5). Subsequent studies showed that *Salmonella* forms MCPs for the utilization of these compounds as carbon and energy sources (5–8). Today, MCPs are thought to be involved in seven or more metabolic processes (9). All MCPs have related protein shells that are composed primarily of bacterial microcompartment (BMC) domain proteins (10–14). However, different types of MCPs encapsulate different enzymes and have various physiological roles (9, 10, 13–19). The MCPs that have been best studied are similar in overall structure. They are polyhedral, usually about 100 to 150 nm in diameter, and are assembled from about 5 to 20,000 polypeptides of 10 to 20 different types. Studies indicate that MCPs are made completely out of protein subunits, and to date, there is no evidence for lipid components. MCP shells have no detectable structural or sequence relationship to viral capsids. Functionally diverse MCPs are thought to have analogous physiological roles (20).

ESSENTIAL FEATURES OF BACTERIAL MICROCOMPARTMENTS

MCPs Are Used To Optimize Pathways That Have Toxic or Volatile Intermediates

Studies on carboxysomes, ethanolamine utilization (Eut) MCPs, and 1,2-PD utilization (Pdu) MCPs indicate that the general function of bacterial MCPs is to optimize metabolic pathways having

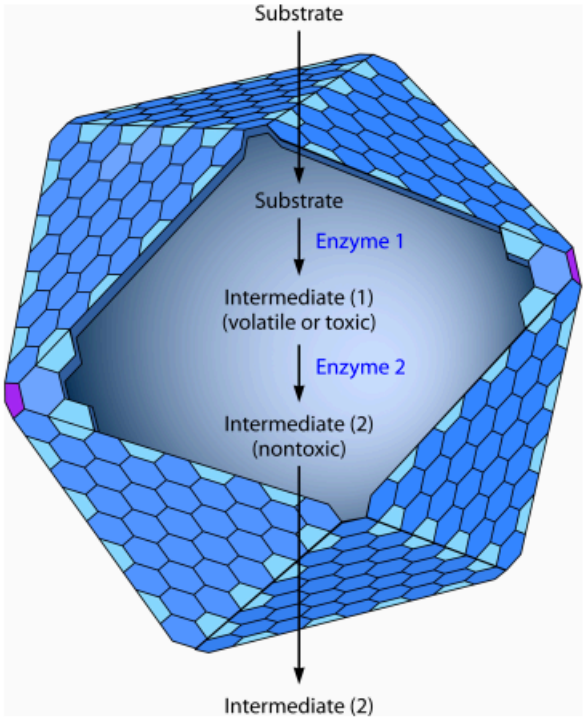


FIG 1 Bacterial microcompartments are used to optimize pathways with toxic or volatile intermediates. The protein shell of the microcompartment acts as a diffusion barrier that helps channel a toxic or volatile intermediate to the next pathway enzyme. In the absence of the microcompartment, such intermediates would either accumulate in the cytoplasm, causing cellular toxicity, or diffuse across the cell envelope and into the environment, resulting in the loss of valuable carbon. The carboxysome and the Pdu and Eut microcompartments help to channel CO_2, propionaldehyde, and acetaldehyde, respectively. All three of these compounds pass easily through the cell envelope, and propionaldehyde and acetaldehyde are cytotoxic. In addition, MCPs may enhance catalysis by bringing enzymes, substrates, and cofactors together at high concentrations.

intermediates that are toxic to the cell or that pass readily through the cell envelope (particularly certain volatile compounds) (2, 21–24). A key feature of MCPs is the use of a protein shell as a diffusion barrier to help channel toxic/volatile metabolic intermediates to downstream enzymes (21–23, 25, 26). A model that outlines how this is achieved is shown in Fig. 1. A metabolic substrate enters a bacterial cell and then diffuses from the cytoplasm into the lumen of the MCP. Within the MCP, the substrate is converted to a toxic or volatile intermediate. The MCP shell acts as a diffusion barrier that helps channel this intermediate to a downstream enzyme, which then converts it to a compound that is nontoxic and/or well retained by the cell envelope. The latter compound moves from the MCP to the cytoplasm, where it enters central metabolism, providing energy and/or carbon for cell growth. In the case of the carboxysome, sequestration allows the concentration of CO_2 to be elevated in the immediate vicinity of RuBisCO, improving its catalytic efficiency (13, 16, 18). For other MCPs, sequestration helps to balance the production and consumption of the toxic/volatile catabolic intermediate. This helps to reduce cytotoxicity, DNA damage, and/or the loss of valuable carbon to the environment (21–23, 26). In addition, the MCP may be used to provide an optimal environment for catalysis by bringing enzymes, substrates, and cofactors together at high concentrations in a precise orientation and possibly by tuning their individual activities.

MCP Distribution and Ecology

Genome sequence analyses indicate that bacterial MCPs are widely distributed and functionally diverse. Current estimates indicate that about 16% of bacteria in 11 different phyla produce MCPs of seven main types, with several subtypes that have extended functionality (9, 20). In addition, recent studies identified a possible eighth type of MCP exclusive to the *Verrucomicrobia* and *Planctomycetes* (27). Multiple MCP types are often found within a single genome, except for carboxysomes, where only one type (α or β) is found per genome (20). Several studies have suggested that MCP loci are subject to frequent horizontal gene transfer, including recent comprehensive analyses of sequenced genomes (6, 13, 20, 28, 29). In many cases, MCPs are encoded by large operons that include multiple MCP shell protein paralogs along with MCP-associated enzymes. The associated enzymes define the MCP type and allow predictions about new MCP functions (9, 15). A number of MCPs have also been investigated experimentally. Studies to date have established that MCPs are used to enhance carbon fixation by the Calvin cycle (the carboxysome) and to optimize the catabolism of 1,2-PD and ethanolamine (the Pdu and Eut MCPs, respectively). In addition, based on sequence analyses, MCPs are predicted to be involved in the metabolism of ethanol, fucose, rhamnose, and an unspecified amino alcohol (9, 20, 27). These compounds are widely found in nature, and their metabolism is important for the natural cycling of matter. Their use provides metabolically capable organisms with a growth advantage in a variety of environments. Of special note is the fact that MCP-dependent processes have a major impact on global carbon fixation and play an important role in the dissemination of enteric pathogens (16, 30).

Carboxysomes Play a Major Role in Global Carbon Fixation

Carboxysomes are found in virtually all bacteria that use the Calvin cycle for CO_2 fixation, including cyanobacteria (a diverse group of photosynthetic bacteria) and many if not all chemoautotrophs (13, 16). Cyanobacteria are widely distributed in freshwater and marine environments, and it has been estimated that as much as 25% of all carbon fixation on Earth occurs within carboxysomes (16). The carboxysome is an essential part of a carbon dioxide-concentrating mechanism (CCM) used to enhance the efficiency of RuBisCO, the rate-limiting enzyme of the Calvin cycle (31). The inefficiency of RuBisCO is due to its relatively low turnover rate (about 1 to 13 s^{-1} depending on the organism) and the fact that it reacts with O_2 in a nonproductive process known as photorespiration, which can drain away up to 50% of the carbon fixed by the Calvin cycle. The bacterial CCM is used to ameliorate these inefficiencies. It begins with active transport processes that elevate the cytoplasmic levels of bicarbonate. Subsequently, bicarbonate is believed to diffuse into the carboxysome, where a carboxysomal carbonic anhydrase (CA) converts HCO_3^- to CO_2 within the carboxysome lumen. The protein shell of the carboxysome restricts the outward diffusion of CO_2, causing it to accumulate proximal to RuBisCO. Elevated CO_2 levels in the immediate vicinity of RuBisCO enhance CO_2 fixation and reduce

competing photorespiration. A key finding that demonstrates the importance of the carboxysome in CO_2 fixation is that cyanobacteria expressing wild-type levels of RuBisCO, but unable to form carboxysomes, are incapable of growth at ambient CO_2 levels. This finding as well as a great deal of additional evidence demonstrating the importance of the CCM in global carbon fixation are described in excellent recent reviews (16, 18); hence, further discussion of the carboxysome here is limited mainly to comparative aspects of MCP function.

The Eut and Pdu MCPs Play Important Roles in Growth and Dissemination of Enteric Pathogens

Recent studies have indicated that ethanolamine degradation, which occurs within the Eut MCP, promotes the growth of *Salmonella* in the inflamed intestine (30). This, in turn, leads to increased transmission of *Salmonella* to new hosts (30). When *Salmonella* colonizes the intestine, it deploys effector molecules that induce an inflammatory response (32). This results in the production of reactive oxygen species that mediate the formation of tetrathionate from thiosulfate (an endogenous sulfur compound) (33). *Salmonella* is then able to respire ethanolamine with tetrathionate as the terminal electron acceptor, which gives it a pronounced growth advantage compared to competing flora (30). Ethanolamine is a particularly important electron donor for tetrathionate respiration because of its prevalence in the intestinal environment (30). Phosphatidylethanolamine is the most abundant lipid in enterocytes, a cell type that is continually shed from the intestinal epithelium.

A number of studies have also suggested that 1,2-PD degradation, which is mediated by the Pdu MCP, is important for enteric pathogenesis. *In vivo* expression studies showed that *pdu* genes are induced in host tissues (34, 35), and competitive index studies with mice showed that *pdu* mutations confer a virulence defect (35). In *Listeria*, comparative genomics indicated a link between 1,2-PD degradation and pathogenesis, and several *pdu* genes are induced during growth within host cells (36, 37). In addition, as described above, recent studies have shown that tetrathionate is an important terminal electron acceptor in the inflamed gut (33), and *Salmonella* is known to respire 1,2-PD with tetrathionate (38). Furthermore, 1,2-PD is expected to be present in the intestinal environment since it is a major product of the anaerobic breakdown of the common plant sugars rhamnose and fucose.

Targeting Systems for Encapsulation of MCP Proteins

For MCPs to function, specific enzymes must be encapsulated within the protein shells. Recent studies have shown that short N-terminal sequences are used to target proteins to diverse MCPs (39–42). The initial hint that targeting sequences are important for the encapsulation of enzymes within MCPs was the observations that the first enzyme of 1,2-PD degradation (the PduCDE diol dehydratase [DDH]) has N-terminal extensions on two of three subunits that were absent from homologs unassociated with MCPs (43) and that these extensions were not required for enzymatic activity (44). Later studies showed that a short C-terminal sequence is used to encapsulate proteins within a compartment, unrelated to MCPs, known as a bacterial nanocompartment (45), and it was suggested that terminal sequences might also be used to target proteins to MCPs, citing prior observations of N-terminal extensions that were nonessential for enzymatic activity (43, 44). Conclusive evidence for MCP targeting sequences was first reported by Fan et al., who showed that a short N-terminal sequence (18 amino acids) was responsible for encapsulation of enzymes into the Pdu MCP (40). A key finding was that deletion of 18 N-terminal amino acids had little effect on the enzymatic activity of PduP but largely prevented its encapsulation into the Pdu MCP (40). In addition, when this 18-amino-acid sequence was fused to green fluorescent protein (GFP), glutathione *S*-transferase (GST), or maltose-binding protein (MBP), these proteins were targeted to the Pdu MCP, proving that this N-terminal sequence was both necessary and sufficient for encapsulation of enzymes into the Pdu MCP (40). Fan et al. also conducted bioinformatic studies that showed that putative N-terminal targeting sequences are widely distributed among diverse MCPs (40). Later studies showed that the predicted targeting sequence of PduD was indeed functional and that it mediated the encapsulation of the PduCDE diol dehydratase into the Pdu MCP (39). In other studies, Choudhary et al. showed that the predicted EutC signal sequence could target proteins to the Eut MCP as well as to recombinant MCP shells (42). In the case of the β-carboxysome, recent studies showed that a short C-terminal sequence of a presumptive scaffolding protein, CcmN, is necessary for MCP assembly (46). Bioinformatic analysis predicted that this sequence (like previously identified N-terminal targeting sequences) forms an amphipathic helix (40, 41). Hence, it was suggested (46) that the C-terminal sequence of CcmN might function in the β-carboxysome similarly to the N-terminal targeting sequences elucidated in previous studies on Pdu and other types of MCPs (40, 46). Additional amphipathic helixes were identified at the C terminus or between domains of several other MCP proteins, suggesting functional conservation (46). With respect to targeting, studies have also shown that heterologous RuBisCO can be localized to the carboxysome (47), that fusions to the entire PduC protein (which lack an identifiable targeting sequence) are directed to the Pdu MCP (48), and that the N-terminal portion of the PduV protein (42 amino acids) mediates the association of heterologous proteins with the exterior of the Pdu MCP (48). It is also of note that a number of MCP-associated enzymes do not have obvious targeting sequences, suggesting that other mechanisms are also used for localization of MCP proteins.

The Mechanism of Protein Targeting to MCPs

Although MCP targeting sequences are reminiscent of plastid signal sequences, their mechanisms and origins are distinctly different. Studies by Fan et al. showed that the N-terminal targeting sequence of the PduP enzyme mediates encapsulation by binding a short C-terminal helix of the PduA or PduJ shell protein (40, 41). This possibility was proposed when N-terminal targeting sequences were first identified (40) and by an analysis of shell protein crystal structures (12). Yeates and colleagues pointed out that the N termini and C termini of shell proteins, which typically reside on the concave side of BMC hexamers, tend to diverge in length and structure between MCP paralogs and often appear flexible or disordered in crystal structures, which are ideal properties for binding sites with divergent specificity. Hence, the finding that lumen enzymes are targeted to MCPs by binding to shell proteins, in conjunction with diverse potential binding sites, led to the proposal that specific lumen enzymes bind partner shell proteins and that this mediates both targeting as well as the internal order of the Pdu MCP (40).

Further studies on the mechanism of enzyme targeting to MCPs were done by using alanine scanning mutagenesis to investigate

the PduP targeting sequence (41). These studies indicated that E7, I10, and L14 of PduP (which are on one face of an α-helix) are the key resides involved in binding the C-terminal helix of the PduA shell protein. Subsequent studies that looked at encapsulation of GFP fused to the PduP targeting sequence found that residues L6, L9, I10, I13, and L14 were most important for encapsulation (49). The differences here may be due to the different methods by which encapsulation was measured: fluorescence microscopy versus MCP purification followed by enzyme assays. More recently, a series of *in vitro* studies were conducted to investigate the targeting of PduP to the Pdu MCP of *Citrobacter freundii* (50). The solution structure of a synthetic peptide corresponding to the PduP targeting sequence was found to be a helix that likely exists in a coiled-coil conformation (50). Further studies indicated that this helix mediated encapsulation by interacting with the PduK shell protein. In contrast, studies conducted with *Salmonella* (described above) indicated that the PduP targeting sequence binds to the PduA and PduJ shell proteins to mediate encapsulation (41). This apparent difference could reflect divergence between the *C. freundii* and *Salmonella* systems, or the PduP targeting sequence might interact with multiple shell proteins. Recent studies of *C. freundii* also indicated that a synthetic 18-amino-acid targeting peptide was capable of diffusing into the lumen of purified empty MCP shells, although surface binding was not ruled out (50). This raises some intriguing questions about the permeability of MCP shells to peptides. Nonetheless, overall, available data support the idea that binding of MCP enzymes to shell proteins results in encapsulation during assembly.

Scaffolding Proteins Are Used To Organize the Interior of the β-Carboxysome

With respect to internal organization, the best-understood MCP is the β-carboxysome (16). In this system, binding studies indicate that the CcmM protein (which has two isoforms) cross-links multiple RuBisCO and carbonic anhydrase (CA) molecules together into an extended network that promotes CO_2 fixation (46, 51–54). The N terminus of CcmM has sequence similarity to γ-CA, and the C terminus typically has three to five tandem repeats with similarity to the small subunit of RuBisCO (SSU-like repeats) (53). In some versions of CcmM, its N-terminal region has redox-sensitive CA activity, while in other versions, CA activity is absent, and an alternative CA (CcaA) is required for carboxysome function. Current models propose that a long form of CcmM (CcmM-58) cross-links RuBisCO, recruits (or provides) carbonic anhydrase activity, and tethers these enzymes to the carboxysome shell (16). Binding to the shell occurs via an intermediary protein (CcmN), which binds to the CcmK2 shell protein via a short C-terminal sequence (46). The short form of CcmM (CcmM-35) consists only of SSU-like repeats and is proposed to be restricted to the central portion of the carboxysome (away from the shell), where it further cross-links RuBisCO. Currently, only the β-carboxysome has been shown to have extensive internal organization. Further studies are needed to understand the primary factors that control the internal organization of other MCPs.

MCP Assembly

MCP assembly has been studied primarily for the carboxysome. Recent work with the β-carboxysome of *Synechococcus* indicates that the cargo proteins (RuBisCO and CA) assemble into a procarboxysome by a process that requires CcmM (55, 56). Subsequently or concomitantly, the shell forms around the cargo, pinching off excess cargo, which serves as a nucleus for biogenesis of the next carboxysome (55, 56). The formation of the shell around the procarboxysome is thought to start with the binding of CcmN, after which a C-terminal peptide of CcmN recruits the additional shell components CcmK2, CcmO, and CcmL, leading to the enclosure process (55). In the case of the α-carboxysome of *Halothiobacillus*, a cryo-electron tomography study suggested that the cargo and the shell assemble simultaneously (57). The α-carboxysome lacks the CcmM protein, which is thought to organize the β-procarboxysome, and the order of shell protein assembly has not been investigated. The assembly of MCPs other than the carboxysome has not yet been directly investigated. A process in which the shell pinches off excess cargo is reasonable, since mutants unable to form the shell of the Pdu MCP accumulate cargo proteins as enlarged polar bodies (22, 58). It is also of note that empty carboxysome and Pdu MCP shells can be produced by recombinant strains in the absence of cargo proteins (59, 60). In wild-type cells, simultaneous assembly of the shell and the cargo would help to minimize such events.

BMC Domain Proteins Are the Basic Building Blocks of the MCP Shell

The shells of bacterial MCPs are made primarily of a family of proteins whose structural core is the BMC domain, and variations upon this core provide functional diversity (11, 61–63). The first crystallographic structures of BMC domains were determined for the carboxysome shell proteins CcmK2 and CcmK4 (63). These proteins form flat hexamers, where one side is concave and the other is relatively flat. In some crystal forms, these hexamers tile edge-to-edge to form extended protein sheets. Thus, it was proposed that sheets of BMC domain proteins form the facets of the carboxysome shell (63). Later crystal structures showed that sheet formation was a conserved feature of many BMC domain proteins, and shell proteins were shown to form true two-dimensional (2D) layers *in vitro* by electron microscopy (64), thereby substantiating their role in shell formation. The initially characterized MCP shell proteins were about 100 amino acids in length and contained a single BMC domain. Later structures identified shell proteins whose monomers have two fused BMC domains. The double-BMC-domain proteins form trimers composed of six BMC domains in a pseudohexameric arrangement that is similar overall to that of the hexameric single-BMC-domain proteins (65–68). These trimers likely arose by gene duplication, and this variation may allow protein conformational changes by eliminating the strict 6-fold symmetry of the BMC domain hexamer (11). As described below, such structural rearrangements may allow BMC domain trimers to act as gated pores for metabolite transport across the MCP shell. Crystallographic studies have also found variants of hexameric and pseudohexameric BMC domain proteins where the BMC domain is circularly permuted (69). The tertiary structure of permuted BMC domain proteins is very close to that of the canonical BMC domain proteins, but the order of the structural elements is different. First observed for the PduU protein, domain permutations result in protein termini at different locations, allowing new functional elaborations at the ends of the polypeptide chain (69). Interestingly, for the structures determined so far, BMC domain proteins have a conserved flat hexagonal shape. Hence, it is thought that shape complementarity along the edges allows different types of shell proteins to tessellate into

mixed molecular sheets, with functional diversification arising from the variations of the individual BMC domain proteins.

Although a single-layer MCP shell is currently the favored model, some studies have suggested that the shell may be a double layer of BMC domain proteins (70, 71). A few crystal structures have found BMC domain proteins associated with their concave surfaces face-to-face, consistent with a double-layered shell (66, 70, 71). In addition, a recent solution study with the CcmK2 protein of *Thermosynechococcus* supports the idea of a double layer (71). However, it is difficult to explain how lumen enzymes would specifically associate with the inner face of a double-layered shell. In addition, an extended double-layer sheet has never been observed in crystals. Hence, the possibility of the MCP shell with a double layer or a partial double layer requires further study.

Pentamers Occupy the Vertices of the MCP Shell

Pentameric non-BMC-domain shell proteins are thought to form the vertices of MCPs and help impart the curvature needed to form a closed structure from the flat protein sheets of hexameric BMC domain proteins (62). Studies of large viruses have shown that pentagonal proteins at the vertices are used to form icosahedral capsids. Crystallography identified a class of pentameric MCP proteins that are encoded by most or all MCP operons (62). Genetic studies in several systems showed that a known pentamer or a pentamer homolog is required for MCP formation *in vivo* (48, 58, 72). In another study, the pentameric protein, CsoS4, of the α-carboxysome was not absolutely required for shell assembly but was required for shell closure and integrity (73). In another case, a conflicting result was initially obtained with regard to a pentamer homolog encoded by the *eut* operon. The EutN protein was found to be a hexamer in crystals (74). However, later studies showed that EutN was pentameric in solution, suggesting that the hexameric form may have been a minor species that selectively crystallized (75). Thus, pentameric vertex proteins are thought to be a widely conserved feature of MCP shells, and because of their key role in MCP formation, this class of proteins has been named bacterial microcompartment vertex (BMV) proteins (75).

The Central Pores of Some BMC Domain Proteins Are Thought To Act as Selectively Permeable Metabolite Conduits

Current models of MCP function require that the shell acts as a permeability barrier to a toxic or volatile intermediate while allowing the substrates, products, and cofactors of lumen enzymes to traverse the shell. This suggests that the shell is selectively permeable. The first crystal structures of BMC domain proteins revealed that they pack tightly together into extended two-dimensional layers with only a few small openings (63). The most notable openings were central pores that were proposed to function in selective metabolite transport (63, 68). Subsequent crystallography revealed that the pores of different BMC domain proteins diverge in their properties, consistent with a role in the selective transport of various metabolites (11, 12, 61). In the case of the carboxysome shell proteins CcmK1, CcmK2, and Ccmk4, the central pores are lined with positively charged residues that could reasonably promote the uptake of a negatively charged compound like bicarbonate, which is a primary substrate of the carboxysome (63). For the PduA protein, which is a component of an MCP used for 1,2-PD degradation, its central pore is lined with numerous H-bond donors and acceptors. This property is proposed to allow preferential movement of 1,2-PD compared to propionaldehyde, which must be held inside the MCP to prevent cellular toxicity (68). In addition, modeling suggests that the PduJ and PduK shell proteins have central pores similar to those of PduA and might also transport 1,2-PD. Thus, overall, MCP pores could control metabolite movement based on their steric/electrostatic properties, and within a single MCP, more than one shell protein could provide similar pores. We point out, however, that crystallographic studies have identified functionally diversified BMC domain proteins that lack small central pores, as described above. In some cases, the central pore is blocked; in other cases, it is replaced with a metal cluster (12, 14). Perhaps even more intriguing, some BMC domain proteins may have gated pores that are used to control metabolite movement across the MCP shell (66, 76).

The MCP Shell May Have Gated Pores

Recent studies have indicated that the MCP shells might contain gated pores that control the movement of metabolites between the MCP lumen and the cytoplasm of the cell. The first evidence of a gated pore was crystal structures that showed that a double-BMC-domain protein in the α-carboxysome (CsoS1D) can adopt two conformations: one where the central pore is open and another where it is closed (66, 76). This led to the proposal of a gated pore that opens and closes in response to a signal to allow metabolite movement across the MCP shell. Presumably, an incoming metabolite would bind the pore, which would open to allow passage and then close behind. This might allow relatively large compounds to cross the shell without compromising the ability of the MCP to minimize the escape of toxic/volatile metabolic intermediates. An alternative hypothesis is that gated pores in the MCP shell might work like an airlock. This idea is based on a recently reported crystal structure in which the CcmP protein formed a dimer of pseudohexamers with a central cavity formed where two concave surfaces come together (70). This would place a gated pore on either side of a central container in an airlock-like arrangement. It is also apparent that the open form of the gated pore is larger than the pores found in hexameric single-BMC-domain proteins, suggesting that it is used to allow the transport of bulkier metabolites. In the case of the carboxysome, larger pores could accommodate ribulose bisphosphate, the substrate of RuBisCO, or 3-phosphoglycerate, the product of the RuBisCO reaction. In addition, homologs of double-BMC-domain proteins having open and closed conformations have been found in operons for the Pdu and Eut MCPs as well as virtually all MCP operons (6, 67, 76, 77). In these cases, a large gated pore might mediate the movement of cofactors required by lumen enzymes, such as ATP, NAD/H, coenzyme A (HS-CoA), and coenzyme B_{12}. It is of note, however, that the *eut* and *pdu* operons each have only a single homolog with a predicted gated pore (EutL and PduB), and although not all of the shell proteins in these systems have been structurally characterized, the current body of structural data suggests that gated pores are only possible in these proteins. On the other hand, current models suggest that multiple enzymatic cofactors must pass through the shell, implying that either the shell proteins with gated pores allow the passage of distinct molecules or additional transport mechanisms may be involved, which are not yet understood. In addition, genomic studies have suggested that gated pores might be used to transport other classes of molecules, since CsoS1D and homologs (which are in the gated-pore

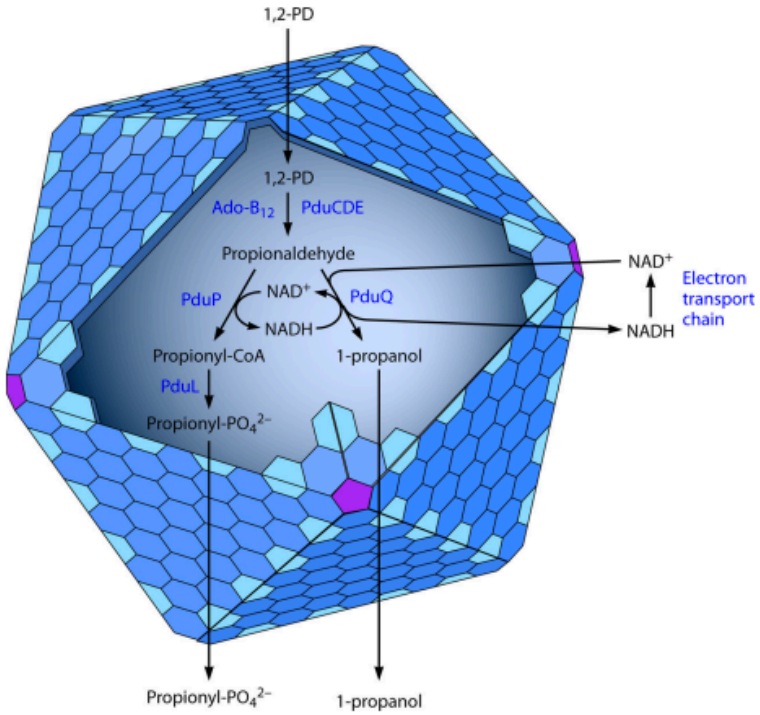


FIG 2 The enzymes encapsulated within MCPs are supplied with cofactors by internal and external recycling. Recent studies showed that the PduQ enzyme recycles NADH to NAD^+ internally within the Pdu MCP. NAD^+ is required for the activity of the PduP enzyme, which is essential for growth on 1,2-propanediol. However, because a PduQ mutant still grows at a rate about 50% of that of the wild type on 1,2-propanediol, external recycling is also indicated (79). In addition, studies of the Eut MCP have shown internal recycling of coenzyme A (81).

class of shell proteins) are often found proximal to enzymes and proteins of unknown function (11). A final observation is that the closing of the large central pores opens up three smaller pores that might promote the movement of enzyme substrates, suggesting a possible dual role of gated-pore-type shell proteins (65, 77). However, to date, no genetic or biochemical studies have directly substantiated the various structure-based models for molecular transport through BMC pores. Hence, the idea of selectively permeable pores as well as gated pores across the shell of MCPs requires further study.

Diverse BMC Domain Proteins May Function as Electron Conduits or Bind Nucleic Acids

Crystallographic studies have revealed a great deal of diversity in BMC domain proteins. The PduT protein has an iron-sulfur cluster where the central pore is typically found (60, 68, 78). It was proposed that PduT might act as an electron conduit between the cytoplasm of the cell and the MCP lumen. Alternatively, it might function as an iron-sulfur insertase that rebuilds damaged Fe-S centers of lumen enzymes or in redox regulation. However, as yet, no genetic or biochemical evidence supporting these hypotheses has been reported. Another MCP protein, EutK, has a BMC domain fused to a helix-turn-helix DNA/RNA binding domain. Presumably, EutK might have a role in regulating the expression of BMC domain proteins, but potential regulatory signals are not clear at this time. In addition, a number of BMC domains are found fused to domains of unknown function, some of which are apparently disordered and might participate in protein-protein binding interactions (11, 12). Thus, overall, a great deal still remains to be learned about the functional diversity of BMC domain proteins.

Enzymatic Cofactors Are Internally Recycled within MCPs, at Least in Part

The first studies to show that cofactors are recycled internally within bacterial MCPs were recently reported by Cheng et al., who showed that the primary role of the PduQ enzyme is to recycle NAD/H within the lumen of the Pdu MCP (79). However, these studies indicated that a portion of NAD/H recycling was also mediated by the electron transport chain, which would require cofactor transport across the MCP shell (Fig. 2) (79). In the Pdu MCP, the role of the PduQ 1-propanol dehydrogenase is to regenerate NAD^+ from NADH during fermentative growth. However, Cheng et al. found that a *pduQ* deletion mutant of *Salmonella* grew at a reduced rate on 1,2-PD under aerobic conditions (79). This was somewhat unexpected since the electron transport chain is available to regenerate NAD^+ during aerobic growth. A possible explanation is that the PduQ enzyme is needed to provide extra capacity for NAD^+ regeneration beyond that available through the electron transport chain or that the role of the PduQ enzyme is to recycle NAD^+ internally within the Pdu MCP. Further studies showed that ectopic expression of a non-MCP-associated alcohol dehydrogenase could not correct the *pduQ* growth defect even when its total enzymatic activity exceeded that of the PduQ en-

zyme. This finding was contrary to the idea that the role of the PduQ enzyme was to provide extra NAD/H recycling capacity. Further studies showed that genetic disruption of the Pdu MCP eliminated the slow-growth phenotype of the *pduQ* deletion mutant, indicating that PduQ has an MCP-specific function. The latter finding was supported by *in vitro* studies with purified MCPs which showed that the PduQ enzyme could regenerate NAD^+ to support the activity of the PduP propionaldehyde dehydrogenase (PDH). Thus, *in vitro* and *in vivo* evidence indicates that the primary role of the PduQ enzyme is to support the activity of the PduP aldehyde dehydrogenase by recycling NADH to NAD^+ internally within the Pdu MCP. However, the studies by Cheng et al. also showed that the *pduQ* deletion mutant was still able to grow on 1,2-PD at a reduced rate. This would be possible only if NAD^+ can be regenerated independently of the PduQ enzyme. It was proposed that some NADH exits the MCP via specific pores, is oxidized back to NAD^+ by the electron transport chain, and then reenters the MCP. Thus, supplying the enzymes of the Pdu MCP with NAD^+/NADH is thought to require both internal recycling and transport across the protein shell via specific pores. Studies with the Pdu system also showed that all of the enzymes needed for adenosylcobalamin (Ado-B_{12}) recycling were components of purified MCPs, and it was therefore proposed that cobalamin recycling can also occur internally within the Pdu MCP (80).

Recent studies of the Eut MCP indicated internal recycling of coenzyme A (81). For the Eut MCP, the EutD enzyme recycles HS-CoA, and growth on ethanolamine is essentially abolished by a *eutD* deletion mutant. Consequently, it was proposed that CoA is unable to cross the shell of the Eut MCP and must be internally recycled. Extrapolation of this finding led to the proposal that no large cofactors cross the shell of the Eut MCP. It was suggested that NAD/H and CoA are internally recycled, that the EutBC ethanolamine ammonia lyase (EAL) and its reactivation factor (which require B_{12} and ATP) are associated with the outer surface of the MCP shell (rather than being encapsulated), and that the acetaldehyde produced by EutBC is injected into the lumen of the Eut MCP. In this model, specific pores are not needed for cofactor transport. In its current form, this model does not apply to the Pdu MCP, where the enzyme that converts 1,2-PD to propionaldehyde (diol dehydratase) has been localized to the interior of the Pdu MCP (22). Further work will be needed to understand the biochemical details of the Eut MCP.

1,2-PROPANEDIOL DEGRADATION AND THE Pdu MCP

For many years, the carboxysome was the only known MCP, and it seemed to be unique, a specialized organelle used to enhance the catalytic activity and specificity of RuBisCO, a slow, relatively nonspecific enzyme. Then, in 1994, genetic studies and sequence analyses showed that the PduA protein (which is encoded by an operon involved in B_{12}-dependent 1,2-PD degradation by *Salmonella*) had relatively high sequence identity to the carboxysome shell protein CcmK (4). Consequently, it was proposed that a carboxysome-like structure might be involved in 1,2-PD degradation (4). A few years later, Bobik and coworkers showed that *Salmonella* conditionally forms polyhedral bodies similar in size and shape to carboxysomes during growth in the presence of 1,2-PD but not during growth on a variety of other carbon sources (6, 22). The Pdu MCPs are polyhedral and about 100 to 150 nm in diameter. They consist of a protein shell that encapsulates enzymes and cofactors used for 1,2-PD degradation. The mass of a Pdu MCP is about 600 MDa, and it is estimated to include 18,000 individual polypeptides of 18 to 20 different types (17). It is thought that the Pdu MCPs are made completely of protein subunits, and there is no evidence that they contain nucleic acid, carbohydrate, or lipid components. The shells of the Pdu MCP are built primarily from BMC domain proteins and also include a BMV pentamer that forms the vertices (6). A variety of studies indicate that the Pdu MCPs are used to optimize growth on 1,2-PD by sequestering a toxic/volatile pathway intermediate (propionaldehyde) (6, 22, 58, 79, 80, 82–84).

Ecology of 1,2-PD Degradation

1,2-PD is a major product of the fermentation of two common plant sugars, rhamnose and fucose (85, 86). These sugars are common in plant cell walls, bacterial capsules, and the glycoconjugates of eukaryotic cells. Hence, 1,2-PD metabolism is thought to be especially important in anaerobic environments such as the large intestines of higher animals, sediments, and the depths of soils. Genes for B_{12}-dependent 1,2-PD degradation are found in a number of soil-dwelling and enteric bacteria, including *Salmonella*, *Klebsiella*, *Shigella*, *Yersinia*, *Listeria*, *Lactobacillus*, *Lactococcus*, *Clostridium*, and at least one strain of *Escherichia coli* (E24377A) (9, 15, 17). In all cases examined, *pdu* gene clusters are highly conserved, indicating that MCPs are essential for 1,2-PD degradation (17). Analyses of *pdu* gene clusters also suggest frequent horizontal gene transfer and operon reorganization as well as conditional selective pressure (6, 20, 29). In addition, a number of studies have indicated that 1,2-PD degradation plays a role in enteric pathogenesis and is important for growth in the intestinal environment, as described above (30, 34).

Pathway of 1,2-PD Degradation

The pathway of 1,2-PD degradation begins with the conversion of 1,2-PD to propionaldehyde by coenzyme B_{12}-dependent diol dehydratase (28, 85–87) (Fig. 3). Propionaldehyde is oxidized to propionyl-CoA by propionaldehyde dehydrogenase, or alternatively, it is reduced to form 1-propanol by 1-propanol dehydrogenase (6, 85–87). Under fermentative conditions, propionyl-CoA is converted by a phosphotransacylase (PTAC) to propionyl-phosphate and then by a reversible propionate kinase to propionate, producing one molecule of ATP by substrate-level phosphorylation (88). However, 1,2-PD fermentation does not produce a source of carbon for biosynthetic reactions, since 1-propanol is excreted in order to eliminate excess reducing equivalents (38). Thus, *Salmonella* does not grow on 1,2-PD as a sole carbon and energy source under fermentative conditions. 1,2-PD metabolism does, however, enhance cell yield by providing additional ATP when a small amount of yeast extract is added to growth medium to provide biosynthetic building blocks (89).

Under aerobic conditions, *Salmonella enterica* is able to grow on 1,2-PD as a sole carbon and energy source (87). In this case, some propionyl-CoA is converted to propionate, but a portion also feeds into the methyl citrate pathway (90, 91). The methyl citrate pathway is an aerobic pathway used for metabolism of propionate/propionyl-CoA derived from diverse carbon sources. It does not function under anaerobic conditions. The genes for the methyl citrate pathway are found outside the *pdu* operon at the *prp* locus (90, 91). In this pathway, propionyl-CoA is joined to oxaloacetate to form methyl citrate, which is subsequently converted to succinate and pyruvate. Pyruvate is metabolized further

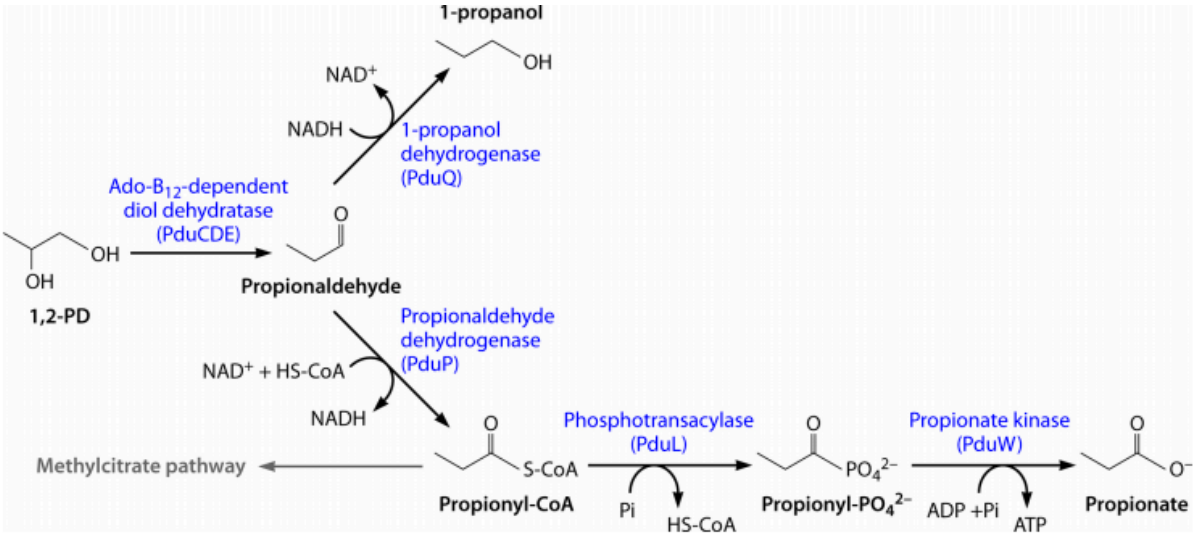


FIG 3 Pathway of 1,2-propanediol degradation. Under fermentative conditions, the pathway of 1,2-propanediol degradation provides a source of ATP and reducing power (NADH) but no source of cell carbon. Under aerobic conditions or during anaerobic respiration with tetrathionate, propionyl-CoA feeds into the methyl citrate pathway to provide additional ATP as well as biosynthetic building blocks.

to acetyl-CoA, which enters the tricarboxylic acid (TCA) cycle. *Salmonella* strains carrying *prp* mutations are unable to grow on 1,2-PD as a sole carbon and energy source (92).

Salmonella can also use 1,2-PD also as a sole carbon and energy source during anaerobic respiration with tetrathionate as the terminal electron acceptor (38). Interestingly, tetrathionate (a somewhat obscure chemical) is found in the inflamed gut, where it provides a substantial growth advantage to *Salmonella*, as described above (33). More common terminal electron acceptors for anaerobic respiration, such as nitrate, fumarate, dimethyl sulfoxide (DMSO), and trimethylamine N-oxide (TMAO), have not been shown to support the anaerobic respiration of 1,2-PD, and the reason why they are unable to do so is currently unclear.

Growth on 1,2-PD Requires Coenzyme B_{12}

Growth of *Salmonella* on 1,2-PD requires coenzyme B_{12} (adenosylcobalamin [Ado-B_{12}]), which is an essential cofactor for diol dehydratase, the first enzyme in the pathway (87). *Salmonella* synthesizes B_{12} *de novo* only under strict anaerobic conditions (93). Consequently, growth of *Salmonella* on 1,2-PD in the presence of oxygen requires supplementation of the growth medium with complex precursors such as vitamin B_{12} (cyanocobalamin [CN-B_{12}]), which *Salmonella* can convert to Ado-B_{12} even when O_2 is present. The paradox that *Salmonella* synthesizes B_{12} *de novo* only under anaerobic conditions but cannot grow by the fermentation of 1,2-PD was previously pointed out (88). Subsequently, however, it was shown that the respiration of 1,2-PD with tetrathionate supported growth with the *de novo* synthesis of B_{12} (38). This apparently resolved the paradox that *Salmonella* synthesizes B_{12} *de novo* only in the absence of oxygen but is unable to grow on 1,2-PD (and ethanolamine) by fermentation (38).

Genes for 1,2-PD Degradation

The genes required for 1,2-PD degradation form a contiguous cluster at the *pdu* locus on centisome 44 of the *S. enterica* chromosome (4, 6, 87). This locus appears to include all the structural proteins and enzymes that comprise the Pdu MCP as well as the regulatory genes that control its production (6, 82). The *pocR* and *pduF* genes are adjacent to and divergently transcribed from the main *pdu* operon (Fig. 4) (4, 94). The *pocR* and *pduF* genes encode a positive transcriptional regulatory protein and a 1,2-PD diffusion facilitator, respectively (4, 95, 96). The *pdu* operon includes 22 *pdu* genes, *pduABB'CDEGHJKLMNOPQSTUVWX*, that encode 1,2-PD degradative enzymes (PduCDELPQW), B_{12} recycling proteins (PduGHOSX), as well as BMC domain proteins and other MCP-related proteins (PduMV) (4, 6).

Regulation of the Genes for 1,2-PD Degradation

The genes for 1,2-PD degradation are contiguous and coregulated with 20 genes (of about 30 total) used for the *de novo* synthesis of cobalamin (*cob*) (Ado-B_{12}) (95–97). Both the *cob* and *pdu* genes are induced in response to 1,2-PD during growth on poor carbon sources (87, 95, 96). This scheme makes sense, as it allows the coproduction of the 1,2-PD degradative pathway together with its required cofactor. Furthermore, coregulation suggests that 1,2-PD catabolism is a primary reason for *de novo* B_{12} synthesis in *S. enterica* (88). If one counts both the *cob* and *pdu* genes, *Salmonella* devotes over 1% of its genome to 1,2-PD degradation, suggesting that these genes are under strong selection in natural environments (88).

Both local and global systems coregulate the *cob* and *pdu* genes (95, 96, 98). Local control requires the PocR transcription factor, which activates transcription when bound by its coeffector 1,2-PD (99). Global control is mediated by the cyclic AMP receptor protein-cyclic AMP complex (CRP-cAMP) and the two-component ArcA/ArcB system, which sense carbon availability and environmental redox potential, respectively (94, 98, 100, 101). Induction occurs both aerobically and anaerobically when 1,2-PD is present and cAMP-CRP levels are sufficiently high. Induction does not

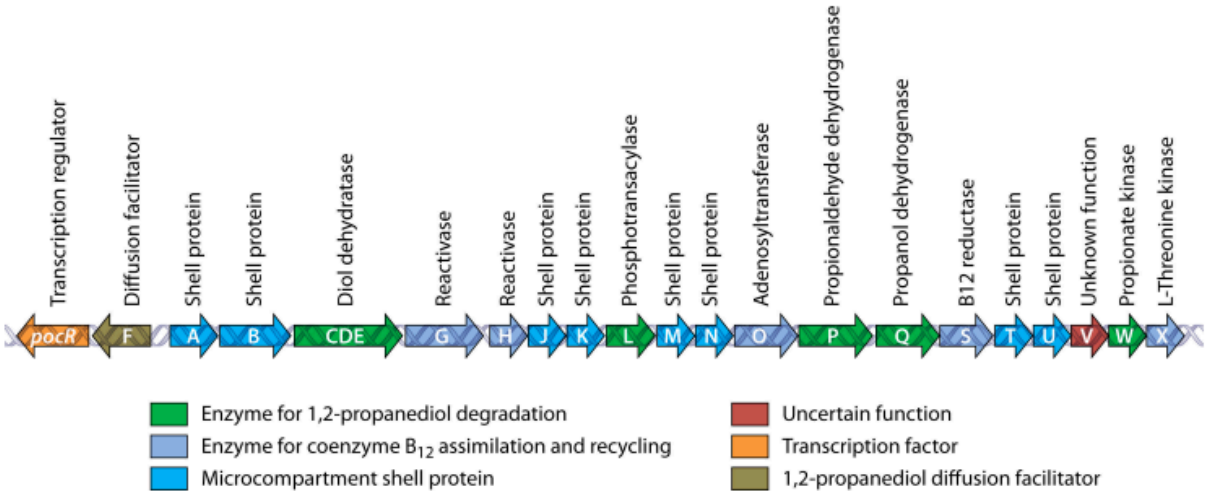


FIG 4 The *pdu* locus of *Salmonella enterica*. The genes involved in 1,2-PD degradation, including those for MCP formation, cluster at a single locus. The genes for MCP formation are interspersed with those encoding pathway enzymes. This general arrangement is found for many MCP operons, although in some cases, the genes may be dispersed.

require the ArcA/ArcB system, but under anaerobic conditions, ArcA/ArcB and CRP-cAMP have an additive effect that increases the maximal expression level. Overall, the *cob* and *pdu* operons are induced by 1,2-PD both aerobically and anaerobically in the absence of preferred carbon sources such as glucose.

The Pdu MCP Functions To Minimize Propionaldehyde Toxicity

The proposed function of the Pdu MCP is to sequester propionaldehyde to prevent cytotoxicity, DNA damage, and carbon loss to the environment; propionaldehyde is poorly retained by the bacterial cell envelope (22, 24). The initial clue for this idea is based on the observation that both 1,2-PD and ethanolamine degradation genes are found in operons that contain homologs of carboxysome shell proteins (4, 5). Both of these processes proceed by analogous pathways with aldehyde intermediates (propionaldehyde and acetaldehyde, respectively). Hence, it was suggested that MCPs might sequester toxic aldehydes formed during 1,2-PD and ethanolamine degradation (5). Later studies showed that shell mutants unable to form the Pdu MCP accumulate propionaldehyde to toxic levels during growth on 1,2-PD (22, 24). These mutants undergo a 20-h period of growth arrest due to propionaldehyde toxicity and are subject to increased mutation rates during growth on 1,2-PD. The explanation for why *Salmonella* overcomes propionaldehyde toxicity and resumes growth after about 20 h is that 1,2-PD metabolism continues during this period even though cells are not growing (22, 24). Over time, 1,2-PD is completely consumed, halting the formation of propionaldehyde. Subsequently, propionaldehyde is depleted by conversion to propionate and 1-propanol, at which time growth of *Salmonella* resumes, with propionate and 1-propanol serving as the carbon and energy sources (22, 24). The aldehyde toxicity model is further supported by the findings that *polA* (DNA repair polymerase) and *gsh* (glutathione biosynthesis) mutants are unable to grow on 1,2-PD or ethanolamine (23, 102). Thus, several lines of evidence support the idea that the Pdu MCP functions to mitigate propionaldehyde toxicity (Fig. 5).

Besides propionaldehyde toxicity, an additional phenotype of mutants unable to form the Pdu MCP is that they grow faster than the wild type at limiting levels of B_{12} (22). This phenotype is seen for a variety of shell mutants, providing solid support for the idea that faster growth results from a shell defect (58, 103). At limiting B_{12} levels, propionaldehyde toxicity is not observed due to decreased activity of B_{12}-dependent diol dehydratase, the enzyme that produces propionaldehyde from 1,2-PD. The current explanation for why MCP mutants grow faster than the wild type at limiting B_{12} levels is that the shell acts as a barrier that regulates/impedes the movement of substrates and cofactors into and out of the MCP. In the absence of the shell, enzymatic activities increase due to improved substrate availability. However, when the B_{12} supply is sufficient, a shell defect leads to propionaldehyde accumulation, which results in toxicity and DNA damage, as described above.

Purification, Composition, and Enzymology of the Pdu MCP

The Pdu MCP from *S. enterica* serovar Typhimurium LT2 was purified intact (Fig. 6) (82). 2D electrophoresis followed by proteomics analyses identified 14 different polypeptide components (PduABB′CDEGHJKOPTU). Subsequent studies determined that 4 additional proteins (PduMNQS) are tightly associated with the Pdu MCP and that the PduV protein might be associated with the outer surface of the shell (48, 58, 79, 80, 84). Of the 19 polypeptides that are components of the Pdu MCP, 8 (PduABB′JKNTU) are likely shell components, PduM and PduV are of uncertain function, and 9 polypeptides are subunits of six enzymes: B_{12}-dependent diol dehydratase (DDH) (PduCDE), DDH reactivase (PduGH), propionaldehyde dehydrogenase (PduP), 1-propanol dehydrogenase (PduQ), cobalamin reductase (PduS), and adenosyltransferase (ATR) (PduO) (6, 28, 79, 80, 83, 89, 104–107). DDH, propionaldehyde dehydrogenase, and 1-propanol dehy-

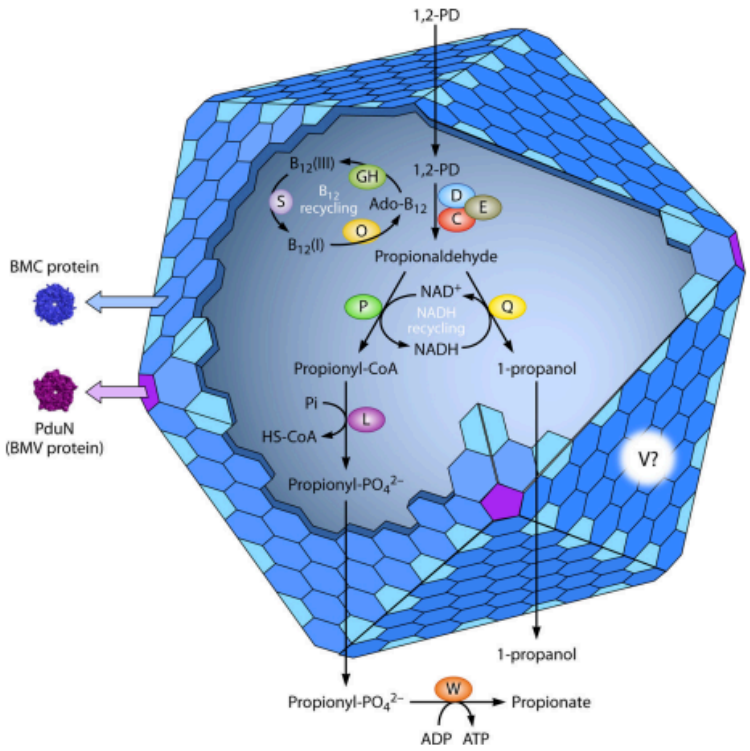


FIG 5 Model for 1,2-propanediol degradation by *S. enterica*. 1,2-PD degradation occurs within the Pdu MCP. The function of the MCP is to sequester propionaldehyde to prevent toxicity and carbon loss. The first step of 1,2-PD degradation is its conversion to propionaldehyde. The protein shell of the Pdu MCP acts as a diffusion barrier and helps to channel propionaldehyde to propionaldehyde dehydrogenase and 1-propanol dehydrogenase. It is thought that propionyl-phosphate leaves the MCP and feeds into the methyl citrate pathway, which provides additional energy and biosynthetic building blocks.

drogenase are highly active in purified MCPs (82, 83). The activities of the other associated enzymes have not been measured by using purified MCPs. The presence of highly active enzymes indicates that the Pdu MCP is a metabolically active organelle. Results have also shown that measurement of PduCDE, PduP, and PduQ activities requires the addition of appropriate substrates and cofactors to enzymatic assay mixtures, showing that the Pdu MCPs as purified are permeable to the metabolites and cofactors of 1,2-PD degradation (79, 82, 83). The detailed mechanisms by which substrates, products, and cofactors enter and exit the Pdu MCP have not been worked out fully, but specific pores centrally located in BMC domain shell proteins provide the most plausible routes for transit, as described above.

Enzymes of the 1,2-PD Degradative Pathway

The B_{12}-dependent PduCDE diol dehydratase. In the first step of 1,2-PD degradation, coenzyme B_{12}-dependent DDH catalyzes the conversion of 1,2-PD to propionaldehyde (28, 108). DDH has three nonidentical subunits, encoded by the *pduCDE* genes, in a stoichiometry of $\alpha_2\beta_2\gamma_2$ (28). The PduCDE DDH is the major enzymatic component of the Pdu MCP, comprising about one-third of its total mass (82). Mutations that inactivate the PduCDE enzyme prevent growth on 1,2-PD, showing that it is essential for this process (28, 109). DDH enzymes are closely related in sequence to B_{12}-dependent glycerol dehydratase (GDH) used by many genera of bacteria for glycerol fermentation (43, 85). DDH enzymes have been purified and characterized from several sources (110–114). No structure is currently available for the PduCDE enzyme from *Salmonella enterica*. However, the crystal structure of a homologous protein, PddABC from *Klebsiella oxytoca*, has been reported (115). The PduCDE subunits of *Salmonella* share 94, 92, and 93% identities with the α, β, and γ subunits, respectively, of the PddABC enzymes in *Klebsiella*. The structure of the *Klebsiella* enzyme reveals its composition as a dimer of heterotrimers, $(\alpha\beta\gamma)_2$ (115). The three subunits interact to maintain the overall structure. By sequence analysis, it was observed that 20 and 16 amino acids at the N termini of the β and γ subunits, respectively, are missing from GDH enzymes (43, 44). Studies with the *Klebsiella* enzyme showed that these extensions were not essential for enzymatic activity but did substantially affect the solubility of the protein, with truncated proteins becoming much more soluble (44). More recent studies, described above, determined that the N-terminal sequence of the β subunit (PduD) of DDH from *Salmonella* targets itself and the PduCDE subunits into the Pdu MCP (39). However, no evidence that the N-terminal extension on the γ subunit has a role in MCP localization could be obtained.

The PduL phosphotransacylase. The PduL phosphotransacylase (PTAC) catalyzes the interconversion of propionyl-PO_4^{2-} and propionyl-CoA. Liu et al. reported biochemical and genetic studies establishing that PduL is a new class of PTAC and demon-

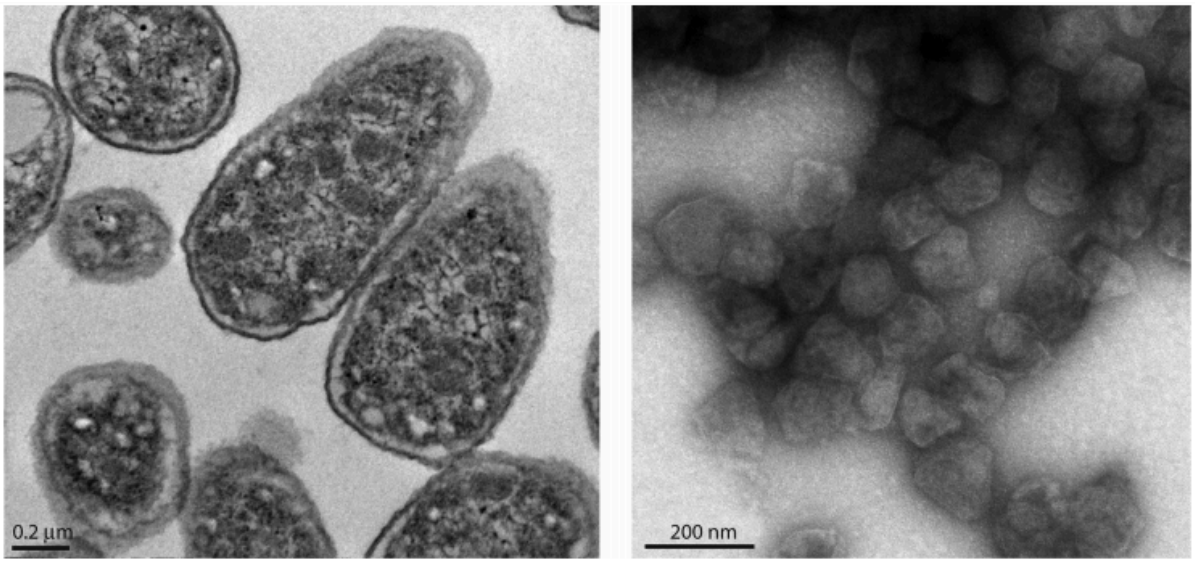


FIG 6 Electron micrographs of Pdu microcompartments from *S. enterica*. (Left) Thin sections of whole cells. (Right) Purified microcompartments.

strating its role in 1,2-PD degradation (89). The PduL enzyme has relatively high PTAC activity and reasonably low K_m values (89). Genetic analyses have shown that *pduL* mutants are unable to ferment 1,2-PD and grow slowly under aerobic conditions, demonstrating the role of PduL in 1,2-PD catabolism (89). BLAST and PSI-BLAST analyses have shown that PduL lacks significant similarity to known PTAC enzymes (89). PduL consists of 210 amino acids, while other PTAC representatives in *Salmonella enterica*, Pta and EutD, are 714 and 338 amino acids, respectively. Recent studies have shown that a PTAC enzyme encoded by the *eut* operon (EutD) is used to recycle CoA internally within the Eut MCP (81). PduL could have a similar role in the Pdu MCP (81).

The PduP propionaldehyde dehydrogenase. The PduP propionaldehyde dehydrogenase (PDH) catalyzes the conversion of propionaldehyde plus HS-CoA and NAD^+ to propionyl-CoA plus NADH (83). Null mutations in the *pduP* gene prevent growth on 1,2-PD, establishing its essential role in this process (83). The PduP enzyme of *Salmonella* has been purified with an N-terminal His tag and partially characterized (83). An enzyme assay showed that it has high levels of PDH activity, and mass spectrometry established propionyl-CoA as a reaction product. The PduP enzyme of *Salmonella* was also shown to be an MCP component by immuno-EM and by analysis of purified MCPs, which were found to be about 8% PduP (82, 83). Recent studies showed that a short N-terminal sequence targets PduP to the interior of the MCP, as described above (40). Homologs of PduP from other organisms have been purified and studied biochemically (116, 117). Some PduP homologs show broad enzymatic activity on various aliphatic aldehydes, including 3-hydroxypropionaldehyde (HPA), as evident from studies with the *Klebsiella pneumoniae* and *Lactobacillus reuteri* enzymes (116, 117). These homologs, which share ~86% amino acid sequence identity with the *Salmonella* PduP enzyme, are used for production of 3-hydroxypropionic acid during glycerol degradation (116, 117).

The PduQ 1-propanol dehydrogenase. Cheng et al. demonstrated in 2012 that PduQ is a part of the Pdu MCP and that it has 1-propanol dehydrogenase activity (79). Results also showed that this enzyme is oxygen sensitive and helps recycle NAD^+ internally within the Pdu microcompartment in order to supply the oxidized cofactor required for propionaldehyde dehydrogenase (PduP). However, studies indicated that NAD^+ regeneration also occurs via the electron transport chain, which requires movement of NAD^+/NADH across the MCP shell (79), as more fully described above.

The PduW enzyme is a propionate kinase. The PduW enzyme is a propionate kinase that catalyzes the conversion of propionyl-phosphate to propionate in a reaction that is coupled to the production of ATP from ADP plus inorganic phosphate (92). This reaction provides a source of ATP during the fermentation of 1,2-PD by *Salmonella* (86, 92). However, PduW also works in the reverse direction to allow the reuptake of propionate derived from 1,2-PD (92). During growth on 1,2-PD, a significant percentage of the initial carbon available to the cell as 1,2-PD is excreted into the medium as propionate (90, 92). Subsequently, PduW works in concert with the housekeeping phosphotransacetylase (Pta) as well as acetyl-CoA and propionyl-CoA synthases (Acs and PrpE, respectively) to recapture exogenous propionate via conversion to propionyl-CoA, which feeds into the methyl citrate pathway, supporting growth (91, 118). The use of PduW and Pta is preferred at high propionate concentrations, whereas Acs and PrpE are required for efficient growth when propionate levels are low (92).

Enzymes for Reactivation of Diol Dehydratase and B_{12} Recycling

The first enzyme in 1,2-PD degradation, B_{12}-dependent DDH, is subject to mechanism-based inactivation (119, 120). Consequently, 1,2-PD degradation requires systems for DDH reactivation and B_{12} recycling (121). During DDH catalysis, the adenosyl group of Ado-B_{12} is periodically lost due to catalytic by-reactions and replaced with an alternative upper ligand, usually a hydroxyl

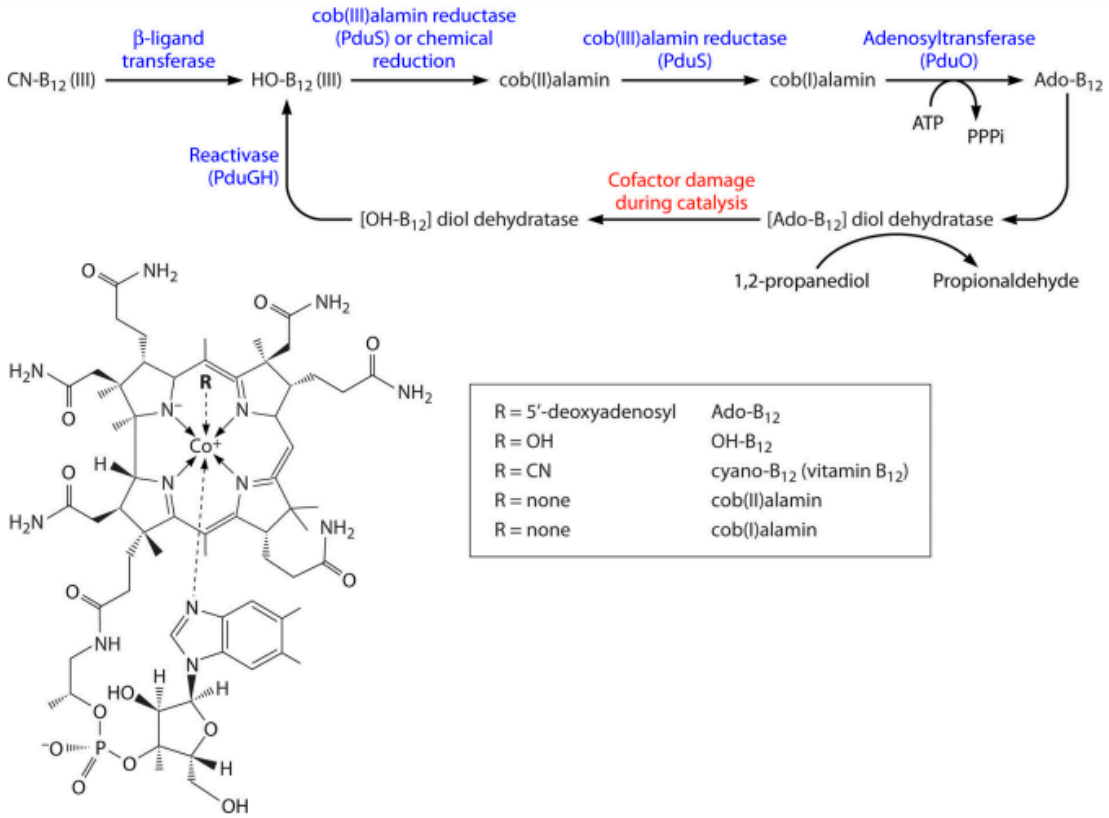


FIG 7 Vitamin B_{12} assimilation and recycling. The first step of 1,2-propanediol degradation is catalyzed by coenzyme B_{12}-dependent diol dehydratase. To supply diol dehydratase with its required cofactor, vitamin B_{12} (CN-B_{12}) can be taken up from the environment and converted to the active coenzyme Ado-B_{12}. During substrate turnover, catalytic by-reactions sometimes convert Ado-B_{12} to HO-B_{12}, which remains enzyme bound. HO-B_{12} is removed from diol dehydratase by the PduGH reactivase and then converted back to Ado-B_{12} by a pathway similar to that used for assimilation of coenzyme B_{12} from the environment.

group, resulting in the formation of an inactive [OH-B_{12}-DDH] complex (120, 122). Reactivation begins with the PduGH reactivase, which converts the inactive [OH-B_{12}-DDH] to free OH-B_{12} and apoenzyme (120, 122–127). Subsequently, Ado-B_{12} spontaneously reassociates with apo-DDH to form an active holoenzyme, and the OH-B_{12} is recycled back to Ado-B_{12} by cobalamin reductase (PduS) and adenosyltransferase (PduO) (80, 104, 105) (Fig. 7). The PduGH, PduS, and PduO enzymes are encoded by the *pdu* operon, and all these enzymes are associated with the Pdu MCP, suggesting that DDH reactivation and OH-B_{12} recycling occur within the MCP lumen (80, 104, 105). The PduS and PduO enzymes are also used for the assimilation of exogenous OH-B_{12} and are redundant with a set of constitutively expressed enzymes for this process (128–130).

The PduS B_{12} reductase. Sampson et al. first reported that the PduS enzyme of *Salmonella* is a B_{12} reductase that catalyzes the reduction of OH-B_{12} to cob(I)alamin in two one-electron steps (80, 106, 107) (Fig. 7). PduS is part of a B_{12} recycling pathway used to supply Ado-B_{12} to DDH (see above), and recent studies have shown that PduS is a component of the Pdu MCP (80). Strains with *pduS* mutations are only marginally impaired in 1,2-PD degradation due to the fact that *Salmonella* has redundant systems for B_{12} reduction (80, 128–130). Although PduS has both cob(III)alamin and cob(II)alamin reductase activities *in vitro*, it is thought that only the cob(II)alamin reductase activity is relevant *in vivo*, since cob(III)alamin reduction occurs as a spontaneous reaction under reducing conditions in the cytoplasm. On the other hand, however, it cannot be ruled out that the environment inside the Pdu MCP is more oxidizing than the general cytoplasm, as has been suggested for carboxysomes (56, 131).

The PduS enzyme of *Salmonella* is 451 amino acids in length. It contains two canonical 4Fe-4S motifs (^{264}CX2CX2CX3C^{274} and ^{309}CX2CX2CX4C^{320}) and a nonclassical Rossmann fold proposed to bind flavin mononucleotide (FMN) (^{120}YX2G[D/E]E^{125}) (80). It also has an NADH binding motif (^{28}GXGXG^{33}) and an N-terminal glycine-rich loop (^{25}GX2GXGGAG[F/L]P[A/T]X2K^{39}) proposed to bind the ADP moiety of NADH. Lastly, PduS has a soluble ligand binding β-grasp (SLBB) fold proposed to bind B_{12}, and this is supported by *in vitro* binding studies (106). Biochemical studies have also shown that PduS is an oxygen-sensitive iron-sulfur flavoprotein and that each monomer contains one molecule of noncovalently bound FMN as well as two coupled 4Fe-4S centers (80, 106). The reduction of cob(II)alamin to cob(I)alamin by PduS is measured by a linked assay with an adenosyltransferase [which converts cob(I)alamin to Ado-B_{12}] or by trapping cob(I)alamin with iodoacetate (107). Direct measurement of cob(I)alamin is difficult due to its high reactivity. *In vitro* studies indicate

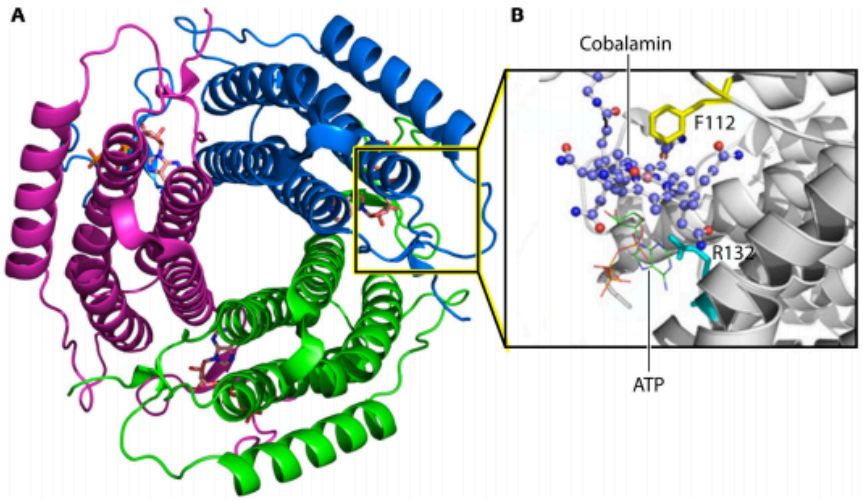


FIG 8 Structure of the PduO adenosyltransferase of *Lactobacillus reuteri*. The structure shown corresponds to the N-terminal ATR domain of PduO from *Salmonella*. (A) PduO in complex with cobalamin and ATP. (B) Residues interacting with ATP and cobalamin are labeled.

that the substrate of PduS is cob(II)alamin bound to an ATR. When cob(II)alamin binds ATR, it undergoes a transition to the 4-coordinate base-off conformer (132–135). Transition to this state raises the midpoint potential of the cob(II)alamin/cob(I)alamin couple by about 250 mV, facilitating reduction.

The PduO adenosyltransferase. Johnson et al. determined that the PduO enzyme of *Salmonella* is an ATR that catalyzes the terminal step of B_{12} assimilation, the conversion of cob(I)alamin and ATP to the coenzyme Ado-B_{12} and triphosphate (105). PduO was the founding member of a new class of ATR enzymes. PduO-type ATRs are the most widespread in nature and are found in higher animals (105, 136–138). As mentioned above, PduO is a component of the Pdu MCP and is thought to be required for recycling B_{12} within this organelle (82). Strains with a *pduO* deletion mutation are only slightly impaired for growth on 1,2-PD due to functional redundancy with the CobA adenosyltransferase (105). Double mutants lacking both *pduO* and *cobA* are unable to synthesize Ado-Cbl and do not grow on 1,2-PD (105, 128). As expected, growth of these double mutants on 1,2-PD is partially corrected by supplementation of media with Ado-B_{12} (105).

Sequence analyses show that the PduO enzyme of *Salmonella* consists of two domains (104). The N-terminal domain, which is about 185 amino acids in length, catalyzes the adenosyltransferase reaction (104). The C-terminal domain (about 150 amino acids in length) is a widely conserved domain of unknown function (DUF336). Although the three-dimensional structure of PduO of *Salmonella* has not been characterized, crystal structures for the PduO N-terminal ATR domains from *Lactobacillus reuteri* and several other organism have been reported (133, 139, 140). For *Lactobacillus*, cocrystallization of the PduO protein with ATP and cob(II)alamin reveals the molecular features of four-coordinate cob(II)alamin stabilization at its active site, where the corrin ring is found in a hydrophobic environment (133, 139) (Fig. 8). Residue Phe112 is critical for the displacement of 5,6-dimethylbenzimidazole from its coordination bond with the cobalt ion of the ring; mutagenesis studies show an 80% loss of activity (141) (Fig. 8).

The PduGH diol dehydratase reactivase. The PduGH protein is thought to be a DDH reactivase because of its high sequence identity to GDH reactivase (DdrAB) (6). The function of DdrAB is to remove inactive forms of B_{12} from GDH, as described above. Early studies of GDH reported that it is inactivated by its substrate (142, 143). Later, Honda et al. showed that GDH was reactivated by ATP and Mn^{2+} in permeabilized cells (*in situ*) (119), and Mori et al. showed that GDH of *Klebsiella* could be reactivated in permeabilized cells of *E. coli* expressing GDH and two adjacent downstream open reading frames (ORFs) (123). These two ORFs were required for reactivation and were named *ddrA* and *ddrB* (123). In further studies, purified DdrAB was shown to reactivate GDH in the presence of ATP and Mg^{2+}. More recent mechanistic studies have shown that DdrAB reactivates GDH by removal of the damaged cofactor, perhaps by a subunit exchange mechanism (124, 144).

Shell Proteins of the Pdu MCP

The current view is that the shell of the Pdu MCP is composed of eight different proteins, PduA, PduB, PduB′, PduJ, PduK, PduN, PduT, and PduU, and possibly a ninth, PduM (58, 84). PduA, PduB, PduB′, PduJ, PduK, PduT, and PduU have one or more BMC domains and are thought to tile together to form the bulk of the MCP shell, which can be modeled as a mixed protein sheet (11, 58, 82). Based on sequence similarity, the PduN protein is thought to be a pentamer that forms the vertices of the Pdu MCP (75). Evidence indicates PduM is an essential MCP structural protein, but its exact role is uncertain (84). Pores found in the center of BMC domain shell proteins are thought to function in the selective transport of enzyme substrates, products, and cofactors via passive and dynamic mechanisms (63). Proteomics analysis has shown that the major shell proteins of the Pdu MCP are PduA, PduB, PduB′, and PduJ and that the minor shell components are PduK, PduN, PduT, and PduU (82). Deletion of the *pduBB′*, *pduJ*, or *pduN* gene severely impairs MCP formation and results in two phenotypes characteristic of a nonfunctional MCP: (i) propionaldehyde toxicity during growth on 1,2-PD and (ii) faster growth on

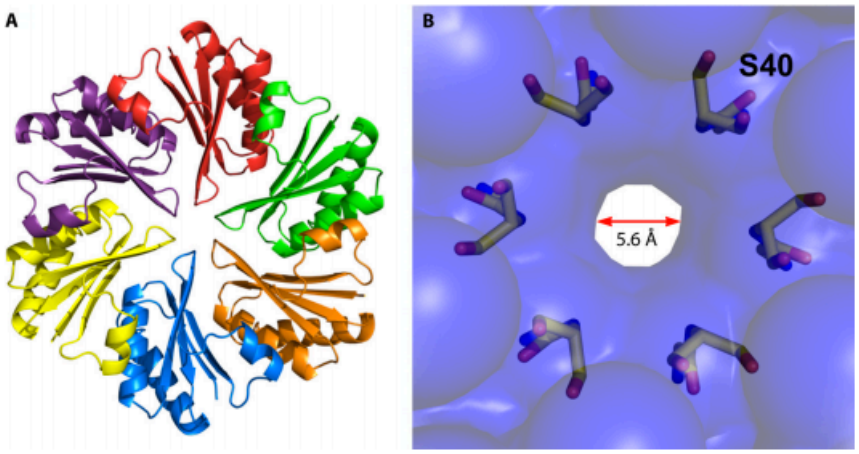


FIG 9 Structure of the PduA shell protein. (A) Hexameric structure of PduA. (B) Surface structure of PduA with pore-lining residues highlighted.

1,2-PD when the B_{12} level is limiting (22, 58). Deletion of the *pduK*, *pduT*, or *pduU* gene did not greatly influence MCP structure or growth on 1,2-PD (58). Strains with a *pduU* or *pduT* deletion do, however, grow slightly slower than the wild type, suggesting that these genes are required for maximal activity of the 1,2-PD degradative enzymes encased within the MCP shell (58). Strains with a *pduA* deletion have intermediate phenotypes. They form larger-than-normal MCPs and are subject to an intermediate level of propionaldehyde toxicity (58). A number of genetic and crystallographic studies of MCP shell proteins have been reported, and these investigations yielded numerous insights into their assembly and function in the Pdu MCP.

PduA is a canonical BMC domain protein. The *pduA* gene encodes a small 94-amino-acid protein which serves as one of the major shell components of the Pdu MCP; it is estimated to comprise 16% of the total shell protein (82). PduA has a single BMC domain and is related to the major carboxysome shell proteins. As evident from immunoelectron microscopy, the PduA protein is a part of the MCP shell (22). Initial studies indicated that a *pduA* deletion mutant of *Salmonella* was unable to produce polyhedral organelles (MCPs) (22). However, more recent studies showed that the initial mutant investigated was in fact a *pduA pduB* double-deletion mutant and that a *pduA* single mutant forms enlarged MCPs with an appearance similar to that of the wild type (58). Nonetheless, *pduA* deletion mutants exhibit phenotypes indicative of damaged MCPs, including growth arrest due to propionaldehyde toxicity and an increased growth rate at limiting B_{12} levels, although these phenotypes are somewhat less severe for *pduA* mutants than for mutants unable to form MCPs (58).

The crystal structure shows that PduA is a symmetric homohexamer (Fig. 9) shaped like a hexagonal disc (68). Adjacent PduA hexamers interact closely to facilitate close packing and thus form a continuous two-dimensional layer without any patencies. This type of molecular layer is formed by a number of MCP shell proteins and is thought to be the basis of MCP shell formation (63).

PduA hexamers have distinct openings along their central axes that are believed to serve as pores for molecular transport across the shell (68). The properties of the PduA pore is suggestive of 1,2-PD transport (68). The pore surface is mostly polar because of the exposed backbone amide nitrogens and carbonyl oxygen atoms of G39 and S40 as well as the side-chain hydroxyl of S40. The diameter of the pore is 5.6 Å, which should occupy 1,2-PD without requiring substantial conformational changes (Fig. 9). The numerous hydrogen bond donors and acceptors in the PduA pore could bind the two hydroxyl groups of 1,2-PD, facilitating transport. On the other hand, propionaldehyde is more hydrophobic than 1,2-PD and presents only a single hydrogen bond acceptor and no donors, potentially limiting the speed of its efflux from the MCP (68).

The PduJ shell protein is closely related to PduA in sequence but is functionally divergent. The PduJ protein shares 80% amino acid sequence identity with PduA, has similar pore-lining residues, and is presumed to have a similar architecture, although its structure has not yet been reported (68). PduJ comprises about 22% of the MCP shell, and in contrast to *pduA*, a *pduJ* deletion mutant forms elongated MCPs (58). This elongated phenotype suggests improper closure of the MCP; hence, PduJ may form the edges that join the MCP facets at the proper angle for closure (58). Strains with a deletion of the *pduJ* gene have phenotypes associated with a nonfunctional MCP, as described above (58).

The PduBB′ proteins may form a gated pore. Havemann et al. identified the *pduB* gene of *Salmonella* by sequence analysis (22). Subsequent proteomics showed that *pduB* is in fact two overlapping genes that are expressed in the same reading frame to produce the PduB and PduB′ proteins (82). PduB and PduB′ are identical in amino acid sequence except that PduB is 37 amino acids longer than PduB′. Further studies in *Citrobacter* showed that PduB and PduB′ are produced from alternative start sites (60). Genetic tests showed that *pduBB′* deletion mutants are unable to form MCPs. EM found that most mutant cells have polar inclusion bodies consisting of aggregated MCP components (Fig. 10) (58). As expected, *pduB* mutants are subject to aldehyde toxicity during growth on 1,2-PD, indicating that the MCP is no longer functional (58).

PduB and PduB′ together comprise over 50% of the total shell protein of the Pdu MCP, 27% and 25%, respectively. Both PduB and PduB′ have two tandem BMC domains and are related to BMC domain proteins proposed to form gated pores in the MCP

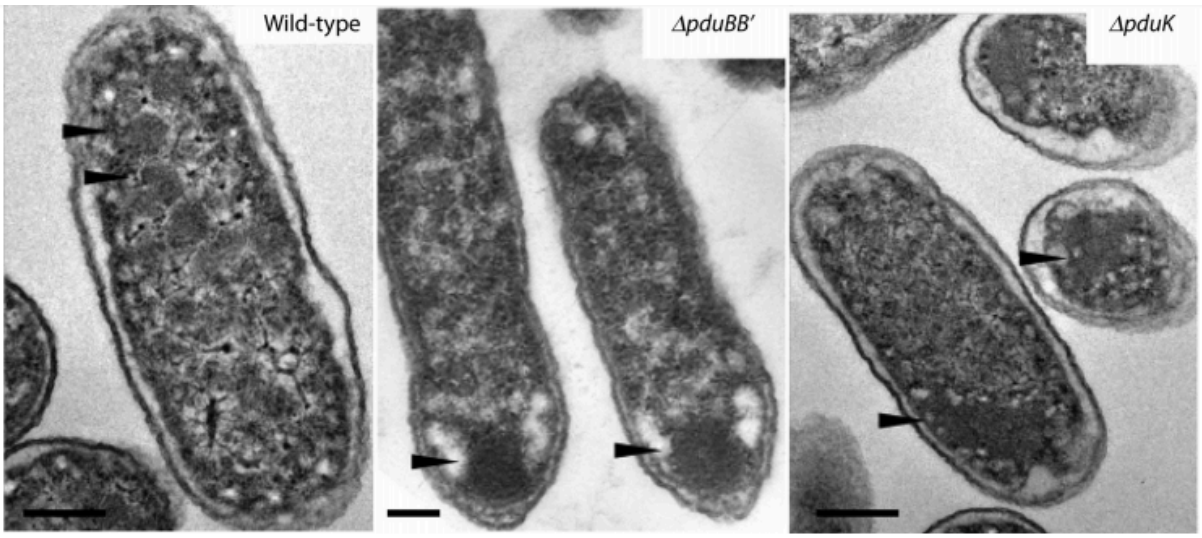


FIG 10 Electron micrographs of mutants impaired for formation of the Pdu microcompartment. The arrows mark normal microcompartments in wild-type *S. enterica*, polar bodies in the Δ*pduBB′* mutant, and aggregated microcompartments in the Δ*pduK* mutant. (Reprinted from reference 58.)

shell (66, 76). The structure of PduB or PduB′ has not been reported for *Salmonella*. However, a crystal structure of the *Lactococcus* PduB protein forms trimers shaped like flat hexagons similar to hexameric BMC domain proteins (77). Like a number of other BMC domain proteins, PduB forms extended two-dimensional sheets in crystals. However, the edge contacts between PduB monomers are slightly out of register compared to other BMC domain proteins, which might lower sheet stability (77). Modeling suggests that PduB and PduA can tile together into a mixed sheet where the edges precisely align, resulting in greater stability (Fig. 11), and this arrangement was suggested to be physiologically important, although it is not yet supported by experimental evidence (77).

The PduK shell protein has a C-terminal domain of unknown function. The PduK protein is 160 amino acids in length. It consists of an N-terminal BMC domain that has a fairly long C-terminal extension (~70 amino acids). It is evident that PduK contains a recognizable [Fe-S] cluster binding motif in its C-terminal domain (CNLCLDPKCPRQKGEPRTLC; cysteine residues that presumably form an Fe-S cluster are underlined) from residue positions 132 to 151. In addition, purified PduK was found to bind iron, supporting the presence of an iron-sulfur center (68). This raises the possibility that PduK functions in electron transfer reactions (68). However, further studies are needed to test this possibility. The structure of PduK has not yet been determined. Genetic analysis showed that a PduK deletion mutation does not impair MCP function but does cause its aggregation (58) (Fig. 10). On this basis, it was proposed that PduK has a role in segregation at cell division (58).

The PduT shell protein has an Fe-S cluster in place of a central pore. PduT is a 184-residue-long tandem-BMC-domain shell protein that assembles as a pseudohexameric homotrimer (68). PduT accounts for about 3% of the total shell protein of the Pdu MCP (82). In crystals, PduT does not form the extended two-dimensional layers seen for several other BMC domain proteins (68). Interactions between the hexamer edges bury only about half as much surface area compared to PduA; this may serve to control the incorporation of PduT, which is a minor shell component. Interestingly, the crystal structure of PduT also revealed a site at the central pore for binding a [4Fe-4S] cluster (68, 78). The ligands for the metal cluster include three cysteine residues (C38), one from each subunit of the trimer, that are located on the con-

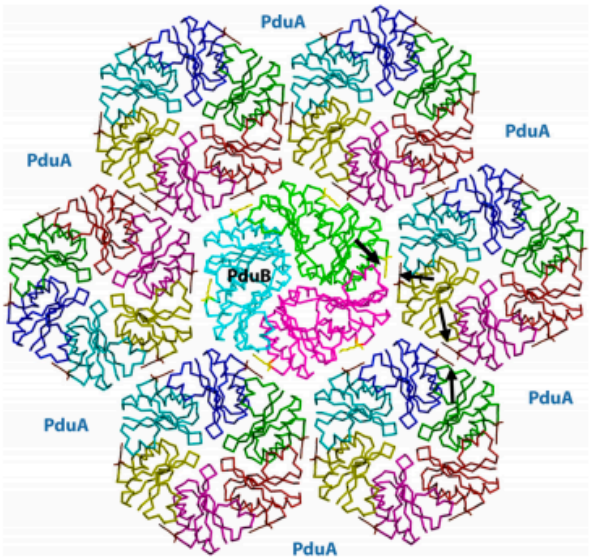


FIG 11 Structure of PduB from *Lactobacillus reuteri*. Modeling indicates that coassembly of PduB and PduA mosaics helps promote stable edge interactions. In this model, a PduB trimer (center) is surrounded by six PduA hexamers. This arrangement brings lysine residues at the edges into register for favorable contacts (arrows).

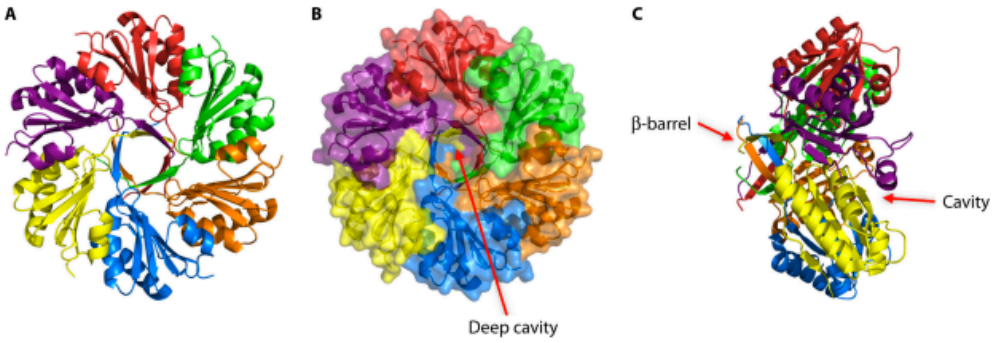


FIG 12 The PduU shell protein is a circularly permutated hexamer whose central pore is capped by a β-barrel. (A) Ribbon model of PduU; (B) surface structure showing a deep cavity on one face of PduU; (C) side view showing a β-barrel that caps the central pore.

served loop at the centermost position of the pore, but the fourth ligand to the iron was not identified (68). It is expected to be positioned transverse to the pore and could extend into the cytoplasm or luminal side of the MCP. It is notable that the [4Fe-4S] cluster is readily accessible from both sides, suggesting a role in electron transfer across the shell or in rebuilding Fe-S centers within the MCP lumen (68). Moreover, PduT contains two more cysteine residues (C108 and C136) in domain II in close proximity, which may form a disulfide linkage. The formation of this disulfide bond might alter the conformation of PduT, resulting in a redox-sensitive change in its pore function (68). In this regard, recent studies by Cheng and Bobik showed that the PduS cobalamin reductase is a component of the Pdu MCP (80). Hence, the putative electron transport function of PduT could be used to support the PduS B_{12} reductase, which mediates a one-electron reduction of the central cobalt atom of B_{12} (80, 106). Alternatively, PduT might be used to rebuild the Fe-S centers of PduS following oxidative damage (80). Genetic tests showed that mutations in PduT only marginally impair growth on 1,2-PD and do not significantly affect MCP structure or function (80). This suggests that the function of PduT might be most important under specific growth conditions not yet identified.

The PduU shell protein has a pore capped by a β-barrel. PduU is a single-BMC-domain protein with a circularly permuted BMC fold (69). It does not form well-packed hexagonal layers in crystals. Instead, PduU has a tendency to form strips of side-by-side hexamers, which may reflect that it is a minor MCP component (Fig. 12). PduU is unique among shell proteins in having an unusual, six-stranded, parallel β-barrel capping the central pore region (Fig. 12C) (69). Typical β-barrels are comprised of larger numbers of β-strands, although a similarly small β-barrel was recently reported in an amyloid peptide (145). When they appear in transmembrane proteins, they are usually comprised of 12 strands or more, thereby creating pores for molecular transport (146). The center of the PduU β-barrel is occluded and unlikely to play a role in transport. It could instead present a specific enzyme binding site. PduU is also unusual in having a deep cavity or depression on the face of the hexamer opposite the β-barrel (69). The surface of this cavity presents several hydrophobic residues, hinting at unknown molecular interactions.

The PduN protein is homologous to pentamers that form MCP vertices. PduN is homologous to pentameric BMV proteins such as CcmL (62). BMV proteins are proposed to form the vertices of bacterial MCPs and aid in their closure (62). Mutants lacking PduN form grossly abnormal MCPs and are subject to propionaldehyde toxicity during growth on 1,2-PD (58). PduN was detected in purified Pdu MCPs by Western blotting. It was also found to be associated with the Pdu MCP of *Citrobacter* by GFP labeling experiments (48). PduN was not detected by Coomassie staining of protein gels due to low abundance (82). This is expected, since there are only 12 vertices in an icosahedron.

The PduM protein is required for MCP formation, but its exact role is uncertain. The PduM protein is required for the formation of the Pdu MCP, but it has no BMC or BMV domain and lacks recognizable sequence similarity to any protein of known function (84). PduM is 163 amino acids in length, and its structure has not yet been determined. Electron microscopy of *pduM* deletion mutants found MCPs that are grossly abnormal in shape along with amorphous cytoplasmic aggregates (84) (Fig. 13). Strains with *pduM* deletion mutations have phenotypes characteristic of a nonfunctional MCP (84).

Other Pdu Enzymes

The PduV GTPase may be involved in MCP segregation at cell division. PduV is a protein of unknown function related to the AAA family of ATPases (17). Indeed, recent observations indicate that purified PduV has weak GTPase activity (48). In addition, GFP fusion analyses suggest that PduV is localized at the outer surface of the microcompartment (48). Interestingly, PduV is proposed to play a role in regulating the spatial distribution of microcompartments within the growing bacterial cell (48). Live-imaging techniques indicate that the movement of the Pdu MCP within the cytoplasm requires the presence of PduV (48). Thus, it can be speculated that PduV might interact with cytoskeletal filaments. However, further studies are required to verify the function of PduV.

PduX is an L-threonine kinase used for B_{12} synthesis. Fan and Bobik showed that the PduX enzyme of *Salmonella enterica* is an L-threonine kinase used for the *de novo* synthesis of Ado-B_{12} and the assimilation of cobyric acid (Cby). PduX converts L-threonine plus ATP to L-threonine-phosphate plus ADP (147). L-Threonine-phosphate is a precursor for the nucleotide loop of Ado-B_{12}. Genetic tests showed that *pduX* mutants were impaired in the synthesis of Ado-B_{12} from cobyric acid and in the *de novo* synthesis of Ado-B_{12} (147). A recombinant PduX enzyme was purified and shown to have L-threonine kinase activity (148). PduX was the

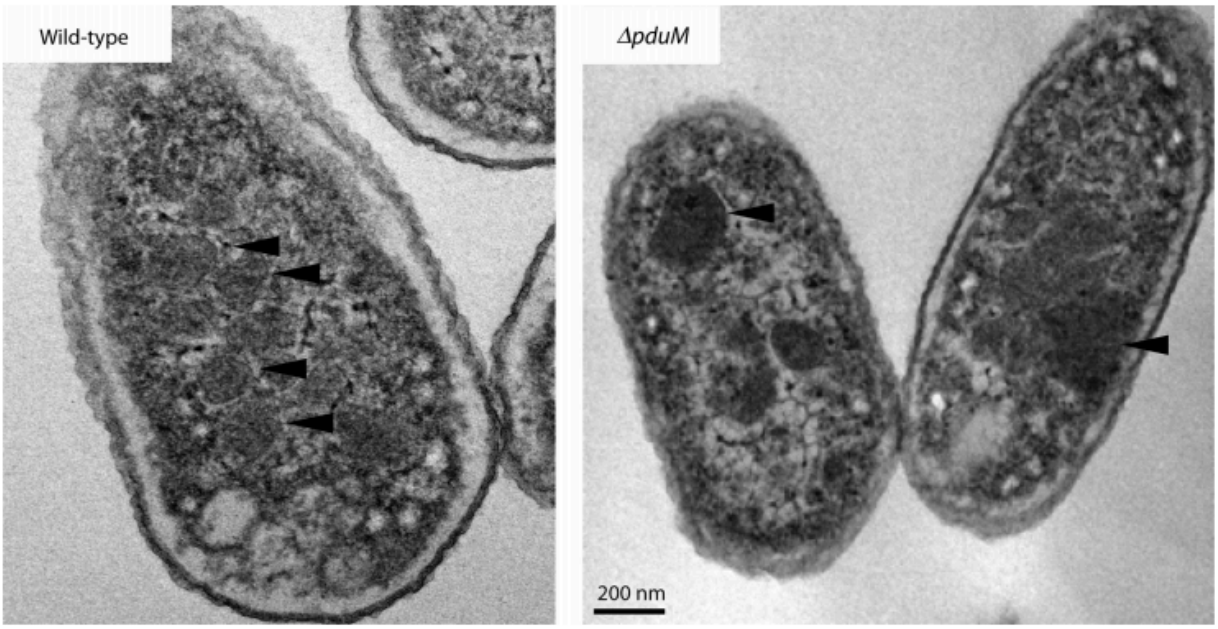


FIG 13 Comparison of wild-type microcompartments of *S. enterica* to those of a Δ*pduM* mutant. (A) Wild-type *S. enterica*; (B) Δ*pduM* deletion mutants. (Reprinted from reference 84.)

first enzyme identified that phosphorylates free L-threonine (147, 148). Further studies of PduX indicate an ordered mechanism in which ATP is the first substrate to bind, and mutation analysis identified several key catalytic residues (148). The role of the PduX enzyme in 1,2-PD degradation is to help provide Ado-B_{12} for DDH (148).

In *Lactobacillus*, the Pdu MCP Is Likely Used for 1,2-PD and Glycerol Degradation as Well as Reuterin Production

Lactobacillus reuteri has a gene cluster closely related to the *pdu* operon of *Salmonella*, which includes genes for 1,2-PD degradation and MCP formation (149). Studies have shown that *L. reuteri* forms a Pdu MCP used for the catabolism of 1,2-PD, as described above, as well as for the metabolism of glycerol (149). Glycerol is metabolized to 3-hydroxypropionaldehyde (HPA) and then to 1,3-propanediol. Although *L. reuteri* does not grow on glycerol as a sole carbon source, added glycerol improves the growth rate and yield during the fermentation of glucose and a variety of other carbohydrates (150). *L. reuteri* is a heterofermentative organism that normally produces lactate, ethanol, and a small amount of acetate during glucose fermentation. Addition of glycerol leaves lactate production essentially unchanged but increases acetate production and decreases ethanol production (150). Furthermore, glycerol supplementation leads to the production of substantial levels of 1,3-propanediol (the product of glycerol dehydration and reduction). These results indicate that HPA (derived from glycerol) serves as an electron acceptor in place of acetyl-phosphate during glucose fermentation, allowing the use of acetyl-phosphate for additional energy generation and thereby increasing the growth rate and yield (150, 151). This is quite different from the situation in *Salmonella*, where glucose and other carbohydrates strongly repress the expression of the *pdu* genes, preventing cofermentation of 1,2-PD (or glycerol) with simple carbohydrates (95, 100). Another interesting aspect of glycerol degradation by *L. reuteri* is that it leads to the production of HPA (also known as reuterin), a broad-spectrum antibiotic associated with the probiotic effects of *L. reuteri* (152, 153). Whole-genome sequence analysis revealed that genes responsible for glycerol degradation are part of the *L. reuteri pdu* operon (154). This operon includes genes with high sequence similarity to the PduCDE diol dehydratase of *Salmonella* (154). Gene knockout studies as well as biochemical investigations confirmed the involvement of the *L. reuteri* PduCDE enzymes in reuterin production (154). The fairly extensive literature on *L. reuteri* as a probiotic and the role of reuterin will not be covered in this review.

THE Eut MCP

A variety of bacteria catabolize ethanolamine as a sole source of carbon and nitrogen by a B_{12}-dependent pathway (155–157). The initial step of ethanolamine degradation is its conversion to acetaldehyde and ammonia by ethanolamine ammonia lyase (EutBC). Subsequent to aldehyde formation, the ethanolamine degradative pathway is analogous to that for 1,2-PD degradation. Like 1,2-PD degradation, ethanolamine utilization involves a bacterial MCP (Eut MCP). The initial evidence for a Eut MCP was the finding that the genes for ethanolamine degradation are clustered with genes encoding MCP shell proteins in *Salmonella* (5, 7). Subsequent studies by Shively et al. showed that *Salmonella* forms MCPs during growth on ethanolamine (158). Similar to the Pdu MCP, the Eut MCP is polyhedral, and its shell is composed primarily of BMC domain proteins (EutK, EutM, EutS, and EutL) and a pentamer (EutN) that forms the vertices of the shell (7, 76). The Eut MCP has been proposed to improve process efficiency while protecting cells from acetaldehyde toxicity (21, 23). However, recent

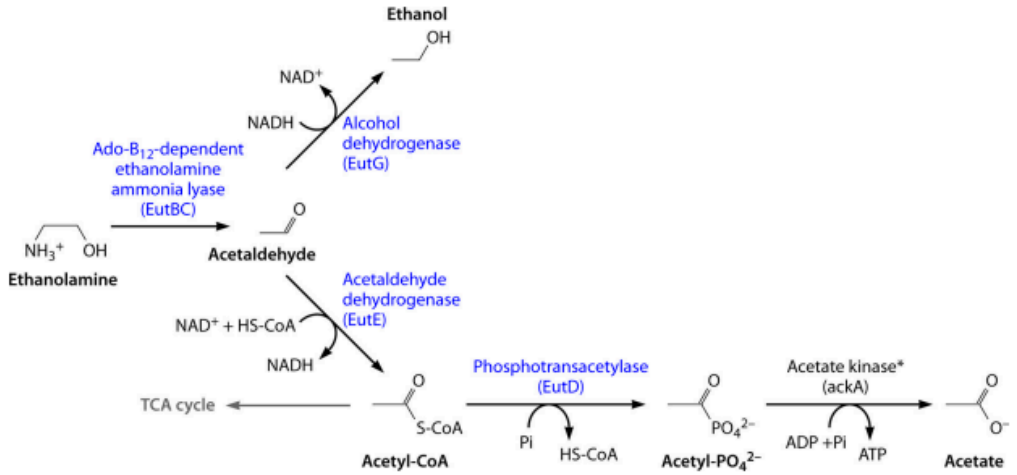


FIG 14 Pathway of coenzyme B_{12}-dependent ethanolamine utilization. Ethanolamine degradation proceeds by a pathway analogous to that of 1,2-PD degradation (Fig. 3), the main difference being C_2 intermediates rather than C_3. As is the case for 1,2-PD degradation, the fermentation of ethanolamine provides a source of ATP and reducing power (NADH) but no source of cell carbon. Under aerobic conditions or during anaerobic respiration with tetrathionate, acetyl-CoA feeds into the TCA cycle to provide additional ATP and biosynthetic building blocks. *, the acetate kinase used for ethanolamine degradation is a housekeeping enzyme encoded outside the *eut* operon.

studies indicate that the main role of the Eut MCP is to prevent the loss of a volatile metabolite (acetaldehyde) through the cell membrane (21). The latter model is supported by extensive genetic tests showing that *Salmonella* mutants unable to make the Eut MCP cannot grow due to carbon loss (21).

Ecology of Ethanolamine Degradation

Genomic analyses indicate that a number of bacteria, including *Actinobacteria*, *Proteobacteria*, *Enterococcus*, *Erwinia*, *Flavobacterium*, *Klebsiella*, *Mycobacterium*, *Pseudomonas*, *Achromobacter*, *Corynebacterium*, *Clostridium*, *Arthrobacter*, *Escherichia*, as well as several other pathogens that cause food poisoning, use an MCP for the B_{12}-dependent degradation of ethanolamine, which can serve as a sole source of both carbon and nitrogen (157). Ethanolamine is a by-product of phosphodiesterase-mediated breakdown of phosphatidylethanolamine, which is a major component of cell membranes in both bacteria and mammals. Hence, ethanolamine is thought to be an important carbon and nitrogen source in diverse environments, including the mammalian gut, where ethanolamine is derived from the host diet as well as bacterial and intestinal epithelial cells (159). Indeed, recent studies indicate that the degradation of ethanolamine within the gut provides a substantial growth advantage to *Salmonella* compared to competing flora and thereby promotes its dissemination to new hosts (see above) (30).

Pathway of Ethanolamine Degradation

The pathway of ethanolamine degradation begins with its conversion to ammonia and acetaldehyde by B_{12}-dependent ethanolamine ammonia lyase (Fig. 14) (EutBC) (155). Some acetaldehyde is converted to acetyl-CoA, then to acetyl-phosphate, and then to acetate. This produces one molecule of ATP by substrate-level phosphorylation (7, 160, 161). In addition, some acetaldehyde is converted to ethanol in order to maintain NADH/NAD^+ balance. The enzymes involved in these reactions include acetaldehyde dehydrogenase (EutE), phosphotransacetylase (EutD), ethanol dehydrogenase (EutG), and the housekeeping acetate kinase (AckA), whose gene maps outside the *eut* gene cluster. Under fermentative conditions, *Salmonella* can use ethanolamine as a sole nitrogen source but not as a sole source of carbon and energy, since none of the ethanolamine pathway intermediates can be diverted into biosynthetic reactions under these conditions (38). Ethanolamine degradation does, however, enhance fermentative cell yield (by providing additional ATP) when yeast extract is available to provide biosynthetic intermediates (38). Under aerobic conditions, *Salmonella* uses ethanolamine as a sole carbon and nitrogen source. When oxygen is available, acetyl-CoA is metabolized via the TCA cycle and the glyoxylate shunt (160). Ethanolamine also functions as a sole carbon and energy source for *Salmonella* during anaerobic respiration when tetrathionate is available as a terminal electron acceptor (38). More common terminal electron acceptors for anaerobic respiration, such as nitrate, fumarate, DMSO, and TMAO, have not been shown to support the anaerobic respiration of ethanolamine, but the reason why they are unable to do so remains uncertain. The pathway of ethanolamine degradation in genera other than *Salmonella* is likely to be similar in terms of the reactions involved. However, regulation is known to be variable, and the types of terminal electron acceptors that support ethanolamine respiration might also vary (157, 159).

Genes for Ethanolamine Degradation

In *Salmonella*, the best-studied case, the genes for ethanolamine utilization (*eut*) are located in a single operon located between the *cysA* and *purC* genes (Fig. 15). This operon contains 17 genes: *eutS*, *eutP*, *eutQ*, *eutT*, *eutD*, *eutM*, *eutN*, *eutE*, *eutJ*, *eutG*, *eutH*, *eutA*, *eutB*, *eutC*, *eutL*, *eutK*, and *eutR* (5). Out of these 17 genes, *eutK*, *eutL*, *eutM*, *eutN*, and *eutS* encode structural proteins that make up the outer shell of the Eut MCP (7, 76). The *eutB* and *eutC* genes encode B_{12}-dependent ethanolamine ammonia lyase (EAL), which breaks down ethanolamine into acetaldehyde and ammo-

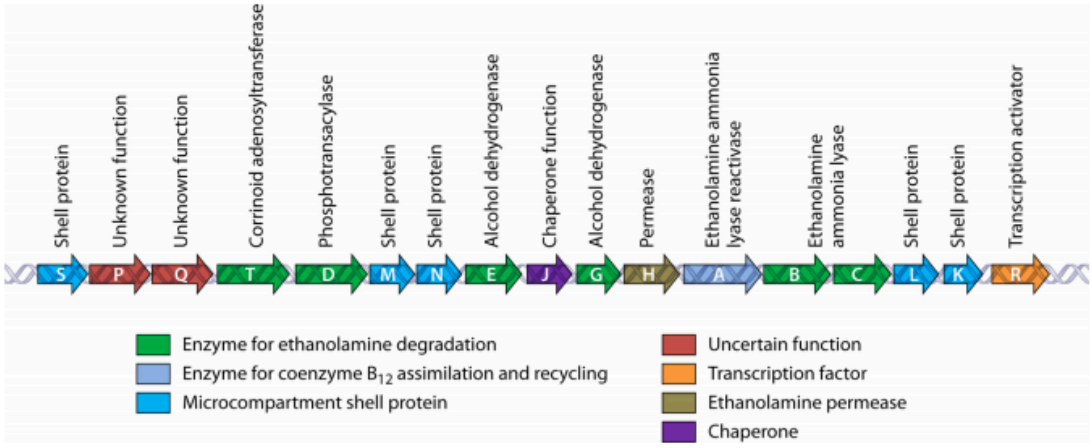


FIG 15 The ethanolamine utilization (*eut*) operon of *S. enterica*. The genes for ethanolamine degradation are found at a single locus that encodes both pathway enzymes and the proteins used to build the Eut MCP.

nia (161). The *eutA* gene encodes an EAL reactivase (162), and the *eutT* gene encodes a corrinoid cobalamin adenosyltransferase that converts cob(I)alamin to Ado-B_{12}, the required cofactor of EAL (163–165). The genes used for the metabolism of acetaldehyde include *eutE*, which encodes an aldehyde dehydrogenase that converts acetaldehyde plus HS-CoA plus NAD^+ to acetyl-CoA plus NADH (7, 161). Subsequently, acetyl-CoA is converted into acetate by the sequential action of a phosphotransacetylase encoded by *eutD* and a housekeeping acetate kinase (*ackA*) encoded outside the *eut* operon (7, 161). The *eutH* gene encodes a transport protein for ethanolamine that is most important at low pH (166). The *eutJ* gene is proposed to encode a chaperone whose specific function is currently unknown (7). The functions of *eutP* and *eutQ* are also unknown, although EutP shares sequence similarity to GTPases (7). Lastly, the *eutR* gene encodes a positive regulatory protein required for induction of transcription of the *eut* operon (167, 168). Genomic studies indicate that there is some variation in the structures of Eut operons among different phyla (157, 159). In most cases, the genes that encode B_{12}-dependent ethanolamine lyase (*eutBC*), the key enzyme of ethanolamine degradation, are associated with proteins for MCP formation (159). Within such operons, there is some variation in the number of types of MCP genes present, likely reflecting divergence in the details of MCP construction. In some cases, however, *eutBC* genes are found in genomes that lack MCP-specific proteins. For these organisms, an alternative method for mitigating acetaldehyde toxicity/volatility is implied (159).

Regulation of the *eut* Operon

As is the case for other aspects of ethanolamine utilization, the regulation of the *eut* operon is best understood in *Salmonella*. In *Salmonella*, regulation of the *eut* operon is unusual in two respects: (i) a positive transcriptional regulator (EutR) is encoded within the *eut* operon at its downstream end, and (ii) induction of the *eut* operon by the EutR protein requires the simultaneous presence of two effectors, ethanolamine and Ado-B_{12} (167, 168). Because the *eutR* gene is located within the *eut* operon, autoinduction occurs and is required for maximal operon expression. Placement of the *eutR* gene within the operon is thought to help balance competition between EAL and the EutR regulatory protein for a limited amount of Ado-B_{12} (168). At the global level, the *eut* operon is regulated by the cAMP-CRP system; hence, induction requires growth on poor carbon sources (167, 168).

In organisms other than *Salmonella*, alternative strategies are used for *eut* operon regulation (159). *Enterococcus faecalis* uses a two-component sensor kinase (EutW) in conjunction with a response regulator (EutV) that has an RNA binding motif suggested to control expression by antitermination (169). In addition, *Enterococcus* uses an intercistronic Ado-B_{12}-sensitive riboswitch as a second layer of control (169). Based on comparative genomics, this multilayer circuit is also present in *Clostridium* and *Listeria* species.

The Eut MCP Functions To Minimize the Loss of a Volatile Metabolic Intermediate, Acetaldehyde

The Eut MCP includes at least five MCP shell proteins, EutK, EutL, EutM, EutN, and EutS (7). Genetic studies have shown that EutK, EutL, EutM, and EutN are indispensable for the growth of *Salmonella* on ethanolamine under conditions that promote the volatilization of acetaldehyde (21). Mutations in the *eutK*, *eutL*, *eutM*, and *eutN* genes disrupt the shell of the MCP, resulting in acetaldehyde leakage to an extent that prevents growth due to carbon loss (21). Other studies have suggested that acetaldehyde toxicity may also be important. Genetic tests showed that *polA* (DNA repair polymerase) and *gsh* (glutathione biosynthesis) mutants are unable to grow on ethanolamine, and aldehyde toxicity was proposed to account for these phenotypes (23, 102). Potential toxic effects of acetaldehyde have also been proposed based on analogy with the Pdu MCP, whose primary role is mitigation of propionaldehyde toxicity (5). Given the much higher vapor pressure of acetaldehyde than of propionaldehyde, carbon loss due to volatility is likely to be more significant in the *eut* system than in the *pdu* system.

Shell Proteins of the Eut MCP

The EutM shell protein is a canonical BMC domain hexamer. EutM is a flat cyclic hexameric protein with the α/β fold characteristic of the MCP shell proteins that form extended sheets. It is a

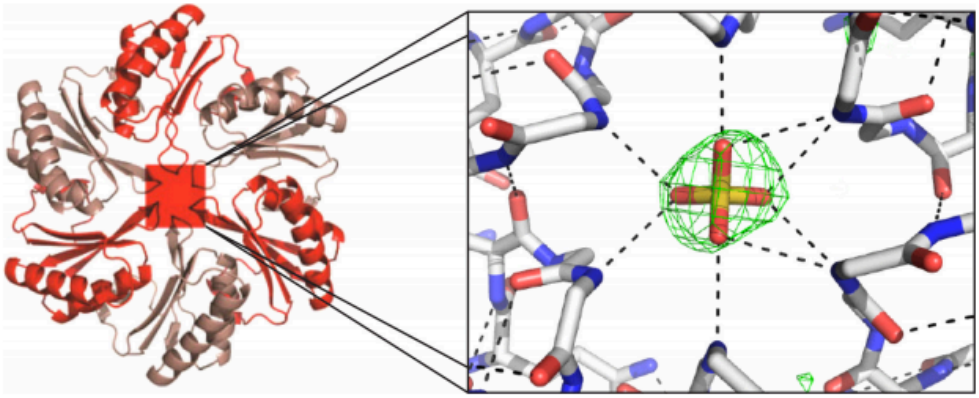


FIG 16 Hexameric structure of EutM. EutM is a good candidate for a major shell protein of the Eut MCP. Its central pore is occupied by a sulfate ion, consistent with a role as a molecular conduit for small molecules. (Reprinted from reference 76 with permission from AAAS.)

97-amino-acid-long protein whose structure has been determined at a resolution of 2.1 Å (76). In the crystal structure, the central pore in the EutM hexamer was occupied with a sulfate ion (from the crystallization buffer), which is consistent with the idea that such pores can be used as conduits for ions and small molecules (Fig. 16). Because EutM has structural features typical of MCP shell proteins, it is a good candidate for a major building block of the Eut MCP (76). In contrast, other BMC domain proteins encoded by the *eut* operon were found to have unusual structural features, suggesting specialized functional roles, as further described below (76).

EutL is a pseudohexameric trimer that may serve as a gated pore for metabolite transport. EutL is the largest BMC domain protein in the Eut MCP system (76). An interesting feature of this protein is that each monomer is a combination of two tandem, circularly permuted BMC domains (65, 76). Consequently, the ultimate tertiary structure of the trimeric protein is a pseudohexamer similar in overall shape to canonical BMC domain hexamers. Another unique feature of EutL is that it appears to have a gated pore at the center of the trimer (Fig. 17) (76). The idea of a gated pore is based on crystal structures that revealed pore-open and pore-closed conformers (76). Comparison of these conformers suggests that there are substantial movements (~15 Å) in the loop residues (residues 65 to 83 and 174 to 185) during interconversion of the two forms. The open pore is triangular, with edges of ~11 Å in length, and the narrowest part of the pore has a diameter of ~8 Å, which is large enough for the transport of enzymatic cofactors. Presumably, a cofactor could bind the closed pore, which would open, allow the factor into the MCP, and then close. In addition, EutL was also found to crystallize in molecular layers, suggesting that it is present as a major structural component of the Eut MCP shell (65, 76).

The EutS shell protein is a bent hexamer. EutS is a 111-amino-acid-long shell protein that adopts the typical α/β fold of the BMC domain proteins (76). The crystal structure of the EutS hexamer has been reported at a 1.65-Å resolution (76). The primary amino acid sequence of EutS is circularly permuted with reference to the sequences of other BMC domain proteins. The structure of EutS is also quite different from that of the canonical BMC domain proteins. Typical BMC domain proteins form flat hexameric struc-

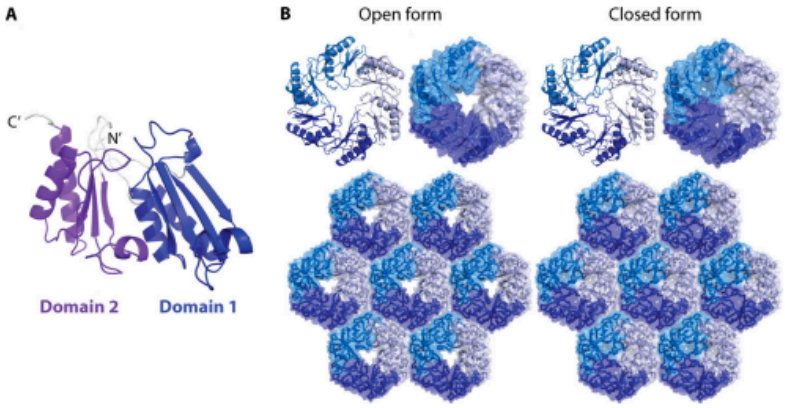


FIG 17 The structure of the EutL shell protein suggests a gated pore to control the movement of metabolites from the cytoplasm to the MCP lumen. (A) EutL monomer in its closed form. (B) Open and closed configurations of EutL trimers in both ribbon diagram and surface representations. EutL forms extended protein sheets, as do a number of other MCP shell proteins (bottom). (Reprinted from reference 76 with permission from AAAS.)

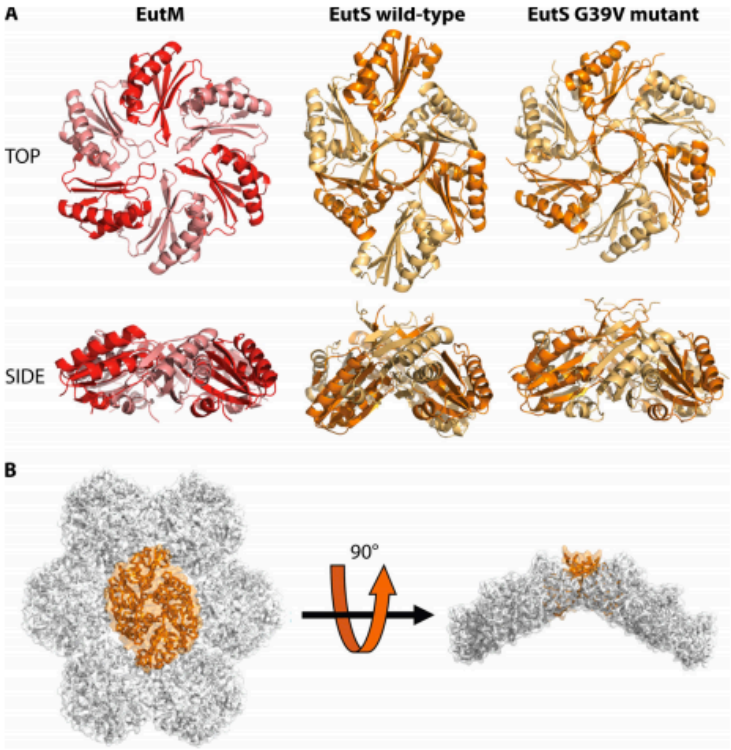


FIG 18 Comparison of the structures of the EutS and EutM shell protein hexamers. (A) EutM, wild-type EutS, and the EutS G39V mutant shown in two views. The wild-type EutS hexamer is bent away from a flat configuration by approximately 40°. The EutS G39V mutant is flat. (B) Hypothetical model showing how EutS (orange) might introduce curvature in an otherwise flat hexameric sheet of shell protein hexamers. (Reprinted from reference 76 with permission from AAAS.)

tures. In contrast, the EutS hexamer has a bend of ~40° (Fig. 18). Due to this bend, the EutS protein is proposed to form the edges that join the facets of the Eut MCP (76). Interestingly, based on sequence analysis, a EutS G39V mutant was constructed and was found to form a flattened symmetrical hexamer similar to that of the canonical BMC domain proteins when crystallized (Fig. 18). Moreover, further investigation revealed that G39 is conserved among EutS proteins (encoded in the *eut* operons of other bacteria) but is absent from other BMC proteins. Thus, a subtle change may have allowed functional diversification of the EutS protein.

The EutK shell protein has a C-terminal nucleic acid binding domain. EutK is an exceptional protein in the Eut MCP system. While all other shell proteins are reported to be oligomeric (trimeric or hexameric) in nature, EutK was found to be monomeric in solution (76). Moreover, in addition to the conserved BMC domain, EutK has a 60-amino-acid C-terminal extension following the BMC domain. Although the crystal structure of the full-length EutK protein has not been reported, that of the 60-amino-acid extension was solved and shown to contain a helix-turn-helix motif typical of nucleic acid binding proteins (76). Moreover, this structure has a localized high positive charge, further supporting its role in DNA/RNA binding. However, as yet, its exact functional role is unknown.

The EutN shell protein is a pentamer. EutN is not a BMC domain protein. Rather, it has sequence similarity to CcmL and CsoS4A, which are pentameric vertex proteins in carboxysomes. Initially, the structure of EutN from *E. coli* was reported as a hexamer at a resolution of 2.7 Å (74). It was then suggested that EutN might distort to fit in the vertices of the Eut MCP and that this distortion could account for the irregular polyhedral shape of Eut MCPs compared to the nearly icosahedral structure of carboxysomes (74). However, a recent report unequivocally demonstrated that EutN is primarily a pentamer in solution by using a technique called oligomeric characterization by addition of charge (9). A comparison of conditions used to obtain crystals of EutN and those of the solution state experiments offered no clear explanation for the observed difference between crystalline and solution oligomer states. The hexameric form of EutN may have been a minor species that selectively crystallized, although the possibility that EutN polymorphism is important in the Eut MCP cannot be discounted.

DIVERSE MICROCOMPARTMENTS

Protein database searches indicate that about 16% of bacterial species have MCP gene clusters that consist of enzyme-encoding genes interspersed with genes for MCP shell proteins (13, 15, 17, 18, 20, 170). The α- and β-carboxysomes (16) as well as the Pdu and Eut MCPs described above represent the best-characterized MCP types. In addition, recent studies have identified an MCP for degradation of plant saccharides that may be limited to the *Verrucomicrobia* and *Planctomycetes* (27). Beyond these MCPs, three

additional types of MCPs have been identified computationally, and several variants representing apparent functional embellishments on the well-characterized MCP types have been highlighted as well (9, 15). Based on bioinformatic analyses, these three lesser-studied MCPs are proposed to function in ethanol metabolism during ethanol/acetate fermentation (Etu MCP), 1,2-PD degradation by a B_{12}-independent pathway that involves a glycyl radical enzyme (Grp MCP), and amino alcohol utilization (9). These three types as well as the recently identified MCP involved in plant saccharide breakdown are described further below. For more details on the carboxysome MCP, the reader is referred to excellent recent reviews (13, 16).

The Grp MCP for B_{12}-Independent 1,2-Propanediol Degradation

The *grp* operon is proposed to mediate the degradation of 1,2-PD (or possibly glycerol) by using a glycyl radical enzyme that is B_{12} independent, rather than the more common B_{12}-dependent enzyme (9). Operons of the *grp* type characteristically contain four enzymes that are interspersed with MCP shell genes (9). The four core enzymes common to *grp* operons include a glycyl radical diol dehydratase that has sequence similarity to pyruvate formate lyase, glycyl radical activase (*S*-adenosylmethionine synthase), aldehyde dehydrogenase, and a putative chaperone. Based on the presence of these enzymes, the Grp MCP is proposed to mediate 1,2-PD degradation by a pathway similar to that used by the Pdu MCP, except that the first reaction (the conversion of 1,2-PD to propionaldehyde) is mediated by a glycyl radical enzyme rather than the B_{12}-dependent diol dehydratase (9). A role for *grp* in 1,2-PD degradation is further supported by microarray studies that showed *grp* operon induction in response to fucose, a common methylpentose that is broken down via 1,2-PD (171). Nonetheless, *grp* operons might also function in glycerol degradation, since the glycyl radical enzyme found in these systems is also active with glycerol (172, 173). Most likely, a Grp MCP would be involved in the use of glycerol as a hydrogen acceptor, as described above for *L. reuteri*, since the enzymes associated with the fermentation of glycerol via dihydroxyacetone are absent from *grp* gene clusters. However, as yet, a role for *grp* genes in glycerol degradation is not supported by physiological studies.

An interesting aspect of the Grp MCPs is that they appear to be elaborated as several diversified subclasses (9). Grp operons that contain only the common enzymatic core (the four enzymes described above) together with MCP proteins are conserved across >20 species, indicating a broad distribution in nature (9). A second type of Grp MCP gene cluster encodes two proteins in addition to this core: one has homology to the C terminus of PduO (a domain of unknown function that is part of the PduO adenosyltransferase), and another has similarity to a macrophage inflammatory channel protein known to transport small neutral metabolites across the membrane. The latter protein might be involved in the transport of metabolites across the shell in the Grp MCPs, but as yet, there is no experimental evidence to support this idea. A third subclass of Grp MCP operon encodes two drug resistance proteins, two regulatory proteins, and a phosphotransacylase that is a homolog of PduL from the Pdu MCP system (89). The PduL-type phosphotransacylase is expected to function in the predicted pathway of 1,2-PD degradation and may be a substitute for an analogous housekeeping enzyme used in other Grp systems. The roles of the drug resistance proteins within the context of the Grp MCP are unclear, and the distinctive regulatory proteins suggest a divergent control system. Another *grp* gene cluster encodes two regulatory proteins similar to a histidine kinase sensor and a response regulator receiver present in the *eut* operon. These clusters likely represent a third type of regulatory system for *grp* gene expression that may be similar to that found in some Eut systems (see above) (159, 169). Yet another variation on the *grp* MCP might include enzymes for later steps of rhamnose and fucose degradation. Both rhamnose and fucose are fermented via 1,2-PD, and some *grp* gene clusters encode homologs of fuculose and rhamnulose aldolase that are induced by these methylpentoses (171, 174).

Shell Proteins of the Grp MCP

The gene clusters for the Grp MCPs have at least four paralogs of genes that encode hexameric/pseudohexameric BMC domain shell proteins (9). In addition, they also contain a gene encoding a pentameric vertex protein (75). A distinctive feature of the Grp MCPs in *Desulfovibrio salexigens* and *Desulfovibrio desulfuricans* is they have up to three different tandem-BMC-domain shell proteins, suggesting divergent functional requirements. Also of interest is an unusual Grp shell protein from *Pectobacterium wasabiae* (Pecwa_4094). UV-visible (UV-Vis) spectroscopy of purified Pecwa_4094 indicates an iron-sulfur cluster, which would be the first such cluster present in a single-BMC-domain protein (9). Overall, the shell of the Grp MCP appears to have a number of intriguing functional variations that may provide insight into MCP evolution and diversification.

Amino Alcohol Degradation May Occur within an MCP in Some Species

Bioinformatic studies identified a presumptive MCP in at least four organisms, *viz.*, *Mycobacterium smegmatis*, *Mycobacterium* sp. strain MCS, *Mycobacterium gilvum*, and *Mycobacterium vanbaalenii* (but not *Mycobacterium tuberculosis*), that may be involved in the metabolism of amino alcohols (9). The operons identified in these organisms encode two BMC shell proteins as well as a putative vertex protein (9). Two enzymes found in these operons are similar to enzymes used for 1-amino-2-propanol metabolism, and others include a class III aminotransferase, an amino acid permease-associated protein, and an aminoglycoside phosphotransferase (9). Hence, an MCP involved in amino alcohol metabolism was proposed.

MCP for Ethanol/Acetate Fermentation (Etu MCP)

Seedorf and coworkers proposed that an MCP is involved in the fermentation of ethanol/acetate by *Clostridium kluyveri* (175). In *C. kluyveri*, MCP shell genes are associated with two genes for ethanol metabolism, NAD-dependent ethanol dehydrogenase and NAD(P)-dependent acetaldehyde dehydrogenase (175). In ethanol/acetate fermentation, ethanol is converted to acetyl-CoA via acetaldehyde (reverse direction from the Eut system). Presumably, the MCP would serve to trap acetaldehyde to prevent carbon loss and cellular toxicity. The Etu MCP has been characterized to only a very limited extent. The crystal structure of the EtuB shell proteins was determined at a 3-Å resolution—it is a trimer with pseudohexagonal symmetry (67)—and initial genomic studies suggest that the Etu MCP may be one of the simplest MCPs found in nature, consisting of only two shell proteins and encapsulated enzymes (176, 177).

MCP for Degradation of Plant Saccharides in *Planctomyces*

Recent studies of *Planctomyces limnophilus* identified an MCP used for the degradation of L-fucose, L-rhamnose, and fucoidan (27). This MCP was proposed to function in the detoxification of pathway intermediates formed during the degradation of these sugars. In the case of L-fucose or L-rhamnose, the *P. limnophilus* MCP is proposed to mediate the disproportionation of lactaldehyde (produced from either L-fucose or L-rhamnose) to 1,2-propanediol and lactate (27). Subsequently, lactate would feed into central metabolism to provide carbon and energy for growth. In *P. limnophilus*, there is no evidence for the further degradation of 1,2-propanediol, which is thought to be excreted as a metabolic end product (27).

BIOTECHNOLOGY APPLICATIONS OF BACTERIAL MCPs

MCPs as Bionanoreactors for *In Vivo* Pathway Optimization

A primary goal of metabolic engineering is the design of reaction pathways for the high-level production of chemicals and pharmaceuticals. There are several factors, such as diffusion restrictions, alternate metabolic routes, accumulation of toxic intermediates, and inhibitory products, that frequently reduce the efficiency of engineered pathways. In nature, cells may bypass such issues by colocalizing metabolic enzymes in a compartment (176–178). A hallmark example of this strategy can be seen in organelles of eukaryotes such as the peroxisome (179, 180). Using bacteria, the workhouses for industrial chemical production, synthetic biologists have begun to address these difficulties by using scaffolds to organize enzymes. By using scaffolds built from protein or DNA, product yields have been substantially increased for a number of chemicals, including glucaric acid, mevalonate, and resveratrol (181–183). An alternative to scaffolds with several inherent advantages is compartmentalization. Pathway encapsulation would not only increase flux through pathways with poor kinetic properties but also allow control of molecules that enter or exit, sequestration of intermediates, and optimization of the local chemical microenvironment. Several possible compartment platforms have been suggested for both *in vivo* and *in vitro* applications, including viral capsids, lumazine synthase capsids, polymeric cages, liposomes, and bacterial MCPs (184–187). Bacterial MCPs seem particularly well suited for the development of intracellular bioreactors, since evolution has designed MCPs to isolate biochemical reactions with the purpose of regulating enzyme activity, enhancing pathway flux, and protecting cells from toxic intermediates. Indeed, with this in mind, studies aimed at developing the tools to repurpose MCPs are being carried out in an increasing number of laboratories.

Heterologous Expression of Bacterial MCPs

Several groups have conducted studies toward the development of MCPs for biotechnological applications by expressing MCP proteins in *E. coli*. An early advance in recombinant engineering of MCPs was reported by Parsons et al., who expressed several *pdu* genes (*pduA*, *pduB*, *pduJ*, *pduK*, *pduN*, *pduT*, and *pduU*) from *Citrobacter freundii* in *E. coli* to successfully produce empty Pdu MCP shells (60). Here it was found that the minimum number of Pdu MCP shell components required to form a microcompartment was only six proteins (PduA, -B, -B′, -J, -K, and -N). The absence of the remaining Pdu shell proteins PduU and PduT did not prevent formation of the recombinant MCP shells. However, in *Salmonella*, the production of Pdu MCPs requires the PduM protein, and PduA was found to be dispensable (84). Hence, the possibility that there may be some variation in the assembly requirements, even of closely related MCP systems, should be considered. In more recent studies, the *Salmonella pdu* genes were also used to produce synthetic MCPs and compartmentalize specific proteins (188). In these studies, vectors for expression of nonnative cargo proteins were constructed, and Western blots as well as an *in vivo* protease accessibility assay were used to evaluate encapsulation. Recently, a proof-of-concept was done with the Pdu system, where targeting sequences were used to coencapsulate pyruvate decarboxylase and alcohol dehydrogenase within a recombinant shell, leading to increased ethanol production compared to production by strains with unencapsulated enzymes (50). Studies aimed at the production of synthetic MCPs have also been carried out with the Eut system, where empty Eut MCPs were produced in *E. coli* by the expression of the *S. enterica* Eut shell genes *eutK*, *eutL*, *eutM*, *eutN*, and *eutS* (42). Surprisingly, a minimal Eut MCP that was morphologically similar to wild-type Eut MCPs was formed by expression of the EutS protein alone, offering the possibility of a minimal system for biotechnology applications (42). In a very recent study, *E. coli* was used to produce recombinant MCP shells from *Haliangium ochraceum* as well as to target enzymes to these shells (189). A potential advantage of this system for biotechnology applications was the high yield of shells and their uniformity and small size. Also of note is a recently reported flow cytometry system that enables rapid quantitation of protein encapsulation within MCPs (49). This system, which is based on encapsulation conferring proteolytic protection to a fluorescent reporter with a small stable RNA (SSRA) tag, should substantially facilitate the development of optimized targeting systems (49). In another line of investigation, the recombinant production of carboxysomes in *E. coli* was achieved by the expression of no more than 10 genes from *Halothiobacillus neapolitanus*, and the structures formed were very similar to those of the native host (59). This advance provides a synthetic system to drive enhanced CO_2 fixation as well as for fundamental research on carboxysomes (59).

Targeting Heterologous Proteins to Bacterial MCPs

Targeting heterologous proteins for encapsulation is a key step toward utilizing MCPs as customizable nanobioreactors. As described above in this review, N-terminal targeting sequences (of about 18 amino acids in length) have been used to target heterologous proteins, such as GFP, to bacterial MCPs in several instances (40, 42). Mechanistically, MCP targeting sequences are thought to form α-helices that bind to short C-terminal helices of shell proteins, leading to encapsulation during assembly (41). While more work is needed to provide a detailed understanding of this mechanism, results to date suggest that engineering interactions between the tails of shell proteins and lumen enzymes might allow the encapsulation of multiple heterologous proteins within a synthetic shell at defined stoichiometries.

Cargo Delivery by Bacterial MCPs

The application of MCPs as cargo carriers for molecular delivery was discussed recently elsewhere (190). Such cargos could include metal nanoparticles, cytotoxic chemicals, and cytotoxic proteins that are surrounded by a self-assembling protein shell. This strategy would have potential applications for medicinal uses such as tumor imaging and cancer therapy. In cancer therapy, reactive

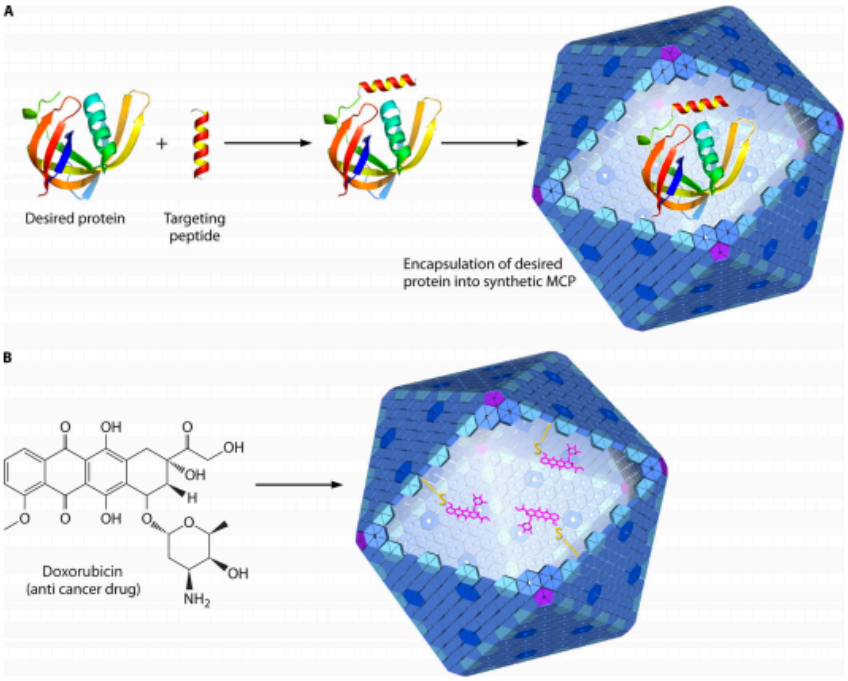


FIG 19 Possible biotechnology applications of bacterial microcompartments. (A) Directing a protein of interest into a synthetic microcompartment might allow engineering of nanobioreactors for chemical production. (B) Bacterial microcompartments as drug delivery vehicles. Doxorubicin, an anticancer drug, might be encapsulated in the lumen of the synthetic microcompartment and targeted to diseased cells.

analogs of anticancer drugs could be chemically attached to thiols on the interior of the MCP surface by engineering cysteine residues at desired positions. In this way, cytotoxic drugs could be sequestered from normal cells while being transported to targeted cancer cells, thereby potentially limiting associated side effects (Fig. 19) (190). A number of potential targeting systems to direct MCPs to disease sites can also be envisioned (190). Chemical tethers such as cysteine residues could be engineered to the MCP surface, followed by chemical linking of a desired small molecule. Alternatively, monoclonal antibodies and recombinant antibody fragments conjugated to the MCP surface are also possible choices (190).

Controlling the Microenvironment within Bacterial MCPs

The shells of bacterial MCPs are made of protein subunits punctuated by pores that are thought to selectively control the entry and egress of small molecules (63). This system, when understood at a fundamental level, could allow precise control over the MCP microenvironment. Pores of desired selectivity could be engineered to allow the entry of desired substrates, sequester key metabolic intermediates, block metabolic branching, improve flux past inefficient enzymes, and selectively allow product exit. However, as yet, the principles of pore function are obscure, and more foundational work is needed to support engineering efforts aimed at building pores with desired selectivity.

CONCLUSIONS

Bacterial MCPs are widespread in nature and functionally diverse. They evolved in bacteria as a means of optimizing metabolic pathways that have toxic or volatile intermediates. MCPs allow metabolic reactions to occur in a unique microenvironment that is bounded by a protein shell rather than a lipid bilayer. The protein shell restricts the diffusion of small polar molecules (CO_2, aldehydes, and perhaps other small polar molecules) that are not effectively retained by a traditional cell envelope based on a lipid bilayer. Although this purpose is simple conceptually, implementation requires an organelle that is 1,000 times larger than a ribosome and that is composed of 10 to 20 different functionally diverse proteins. Works by a number of laboratories employing the varied methods of crystallography, genetics, biochemistry, and cell biology have yielded insights into the underlying principles of bacterial MCPs and provided a basis for the development of engineering tools needed to design purpose-built MCPs. Indeed, many potential applications have been suggested, as described above. However, much of the MCP story still remains to be told before such applications can be practically realized. Key unanswered questions include how pore selectivity is controlled, how multiple shell proteins assemble into an ordered structure with a defined stoichiometry and overall size, the factors that control MCP stability/lifetime, as well as the principles that control the internal organization and activity of encapsulated enzymes. Thus, while we have learned a great deal about bacterial MCPs, many intriguing questions remain to be answered.

ACKNOWLEDGMENTS

This work was supported by grant AI081146 from the National Institutes of Health to T.O.Y. and T.A.B. and by NSF grant MCB0616008 to

T.A.B. S.C. was supported by NRSA GM-0846 and the UCLA Graduate Division.

We thank members of the Bobik and Yeates laboratories for helpful discussions.

REFERENCES

1. **Drews G, Niklowitz W.** 1956. Beiträge zur Cytologie der Blaualgen. II. Zentroplasma und granulare Einschlüsse von *Phormidium uncinatum*. Arch. Mikrobiol. **24**:147–162.
2. **Shively JM, Ball F, Brown DH, Saunders RE.** 1973. Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of *Thiobacillus neapolitanus*. Science **182**:584–586. http://dx.doi.org/10.1126/science.182.4112.584.
3. **Cannon GC, Shively JM.** 1983. Characterization of a homogenous preparation of carboxysomes from *Thiobacillus neapolitanus*. Arch. Microbiol. **134**:52–59. http://dx.doi.org/10.1007/BF00429407.
4. **Chen P, Andersson DI, Roth JR.** 1994. The control region of the *pdu/cob* regulon in *Salmonella typhimurium*. J. Bacteriol. **176**:5474–5482.
5. **Stojiljkovic I, Baeumler AJ, 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.
6. **Bobik TA, Havemann GD, Busch RJ, Williams DS, Aldrich HC.** 1999. The propanediol utilization (*pdu*) operon of *Salmonella enterica* serovar Typhimurium LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B_{12}-dependent 1,2-propanediol degradation. J. Bacteriol. **181**:5967–5975.
7. **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.
8. **Brinsmade SR, Paldon T, Escalante-Semerena JC.** 2005. Minimal functions and physiological conditions required for growth of *Salmonella enterica* on ethanolamine in the absence of the metabolosome. J. Bacteriol. **187**:8039–8046. http://dx.doi.org/10.1128/JB.187.23.8039-8046.2005.
9. **Jorda J, Lopez D, Wheatley NM, Yeates TO.** 2013. Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. Protein Sci. **22**:179–195. http://dx.doi.org/10.1002/pro.2196.
10. **Yeates TO, Crowley CS, Tanaka S.** 2010. Bacterial microcompartment organelles: protein shell structure and evolution. Annu. Rev. Biophys. **39**: 185–205. http://dx.doi.org/10.1146/annurev.biophys.093008.131418.
11. **Yeates TO, Jorda J, Bobik TA.** 2013. The shells of BMC-type microcompartment organelles in bacteria. J. Mol. Microbiol. Biotechnol. **23**: 290–299. http://dx.doi.org/10.1159/000351347.
12. **Yeates TO, Thompson MC, Bobik TA.** 2011. The protein shells of bacterial microcompartment organelles. Curr. Opin. Struct. Biol. **21**: 223–231. http://dx.doi.org/10.1016/j.sbi.2011.01.006.
13. **Kerfeld CA, Heinhorst S, Cannon GC.** 2010. Bacterial microcompartments. Annu. Rev. Microbiol. **64**:391–408. http://dx.doi.org/10.1146/annurev.micro.112408.134211.
14. **Kinney JN, Axen SD, Kerfeld CA.** 2011. Comparative analysis of carboxysome shell proteins. Photosynth. Res. **109**:21–32. http://dx.doi.org/10.1007/s11120-011-9624-6.
15. **Bobik TA.** 2006. Polyhedral organelles compartmenting bacterial metabolic processes. Appl. Microbiol. Biotechnol. **70**:517–525. http://dx.doi.org/10.1007/s00253-005-0295-0.
16. **Rae BD, Long BM, Badger MR, Price GD.** 2013. Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO_2 fixation in cyanobacteria and some proteobacteria. Microbiol. Mol. Biol. Rev. **77**:357–379. http://dx.doi.org/10.1128/MMBR.00061-12.
17. **Cheng S, Liu Y, Crowley CS, Yeates TO, Bobik TA.** 2008. Bacterial microcompartments: their properties and paradoxes. Bioessays **30**: 1084–1095. http://dx.doi.org/10.1002/bies.20830.
18. **Yeates TO, Kerfeld CA, Heinhorst S, Cannon GC, Shively JM.** 2008. Protein-based organelles in bacteria: carboxysomes and related microcompartments. Nat. Rev. Microbiol. **6**:681–691. http://dx.doi.org/10.1038/nrmicro1913.
19. **Bobik TA.** 2007. Bacterial microcompartments. Microbe **2**:25–31. http://www.microbemagazine.org/images/stories/arch2007/jan07/znw00107000025.pdf.
20. **Abdul-Rahman F, Petit E, Blanchard JL.** 2013. The distribution of polyhedral bacterial microcompartments suggests frequent horizontal transfer and operon reassembly. J. Phylogenet. Evol. Biol. **1**:4. http://escience central.org/journals/the-distribution-of-polyhedral-bacterial-microcompartments-suggests-frequent-horizontal-transfer-and-operon-reassembly-2329-9002.1000118.pdf.
21. **Penrod JT, Roth JR.** 2006. Conserving a volatile metabolite: a role for carboxysome-like organelles in *Salmonella enterica*. J. Bacteriol. **188**: 2865–2874. http://dx.doi.org/10.1128/JB.188.8.2865-2874.2006.
22. **Havemann GD, Sampson EM, Bobik TA.** 2002. PduA is a shell protein of polyhedral organelles involved in coenzyme B_{12}-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar Typhimurium LT2. J. Bacteriol. **184**:1253–1261. http://dx.doi.org/10.1128/JB.184.5.1253-1261.2002.
23. **Rondon RR, Horswill AR, Escalante-Semerena JC.** 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.
24. **Sampson EM, Bobik TA.** 2008. Microcompartments for B_{12}-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. J. Bacteriol. **190**:2966–2971. http://dx.doi.org/10.1128/JB.01925-07.
25. **Dou Z, Heinhorst S, Williams EB, Murin CD, Shively JM, Cannon GC.** 2008. CO_2 fixation kinetics of *Halothiobacillus neapolitanus* mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO_2. J. Biol. Chem. **283**:10377–10384. http://dx.doi.org/10.1074/jbc.M709285200.
26. **Price GD, Badger MR.** 1989. Expression of human carbonic anhydrase in the cyanobacterium *Synechococcus* PCC7942 creates a high CO_2-requiring phenotype: evidence for a central role for carboxysomes in the CO_2 concentrating mechanism. Plant Physiol. **91**:505–513. http://dx.doi.org/10.1104/pp.91.2.505.
27. **Erbilgin O, McDonald KL, Kerfeld CA.** 2014. Characterization of a planctomycetal organelle: a novel bacterial microcompartment for the aerobic degradation of plant saccharides. Appl. Environ. Microbiol. **80**: 2193–2205. http://dx.doi.org/10.1128/AEM.03887-13.
28. **Bobik TA, Xu Y, Jeter RM, Otto KE, Roth JR.** 1997. Propanediol utilization genes (*pdu*) of *Salmonella typhimurium*: three genes for the propanediol dehydratase. J. Bacteriol. **179**:6633–6639.
29. **Lawrence JG, Roth JR.** 1996. Evolution of coenzyme B_{12} synthesis among enteric bacteria: evidence for loss and reacquisition of a multigene complex. Genetics **142**:11–24.
30. **Thiennimitr P, Winter SE, Winter MG, Xavier MN, Tolstikov V, Huseby DL, Sterzenbach T, Tsolis RM, Roth JR, Baumler AJ.** 2011. Intestinal inflammation allows *Salmonella* to use ethanolamine to compete with the microbiota. Proc. Natl. Acad. Sci. U. S. A. **108**:17480–17485. http://dx.doi.org/10.1073/pnas.1107857108.
31. **Price GD, Badger MR, Woodger FJ, Long BM.** 2008. Advances in understanding the cyanobacterial CO_2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. J. Exp. Bot. **59**:1441–1461. http://dx.doi.org/10.1093/jxb/erm112.
32. **Winter SE, Baumler AJ.** 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. http://dx.doi.org/10.4161/gmic.2.1.14911.
33. **Winter SE, Thiennimitr P, Winter MG, Butler BP, Huseby DL, Crawford RW, Russell JM, Bevins CL, Adams LG, Tsolis RM, Roth JR, Baumler AJ.** 2010. Gut inflammation provides a respiratory electron acceptor for *Salmonella*. Nature **467**:426–429. http://dx.doi.org/10.1038/nature09415.
34. **Conner CP, Heithoff DM, Julio SM, Sinsheimer RL, Mahan MJ.** 1998. Differential patterns of acquired virulence genes distinguish *Salmonella* strains. Proc. Natl. Acad. Sci. U. S. A. **95**:4641–4645. http://dx.doi.org/10.1073/pnas.95.8.4641.
35. **Heithoff DM, Conner CP, Hentschel U, Govantes F, Hanna PC, Mahan MJ.** 1999. Coordinate intracellular expression of *Salmonella* genes induced during infection. J. Bacteriol. **181**:799–807.
36. **Buchrieser C, Rusniok C, Kunst F, Cossart P, Glaser P.** 2003. Comparison of the genome sequences of *Listeria monocytogenes* and *Listeria innocua*: clues for evolution and pathogenicity. FEMS Immunol. Med. Microbiol. **35**:207–213. http://dx.doi.org/10.1016/S0928-8244(02)00448-0.
37. **Joseph B, Przybilla K, Stuhler C, Schauer K, Slaghuis J, Fuchs TM, Goebel W.** 2006. Identification of *Listeria monocytogenes* genes contributing to intracellular replication by expression profiling and mutant

screening. J. Bacteriol. 188:556–568. http://dx.doi.org/10.1128/JB.188.2.556-568.2006.
38. Price-Carter M, Tingey J, Bobik TA, Roth JR. 2001. The alternative electron acceptor tetrathionate supports B_{12}-dependent anaerobic growth of *Salmonella enterica* serovar Typhimurium on ethanolamine or 1,2-propanediol. J. Bacteriol. 183:2463–2475. http://dx.doi.org/10.1128/JB.183.8.2463-2475.2001.
39. Fan C, Bobik TA. 2011. The N-terminal region of the medium subunit (PduD) packages adenosylcobalamin-dependent diol dehydratase (PduCDE) into the Pdu microcompartment. J. Bacteriol. 193:5623–5628. http://dx.doi.org/10.1128/JB.05661-11.
40. Fan C, Cheng S, Liu Y, Escobar CM, Crowley CS, Jefferson RE, Yeates TO, Bobik TA. 2010. Short N-terminal sequences package proteins into bacterial microcompartments. Proc. Natl. Acad. Sci. U. S. A. 107:7509–7514. http://dx.doi.org/10.1073/pnas.0913199107.
41. Fan C, Cheng S, Sinha S, Bobik TA. 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. http://dx.doi.org/10.1073/pnas.1207516109.
42. Choudhary S, Quin MB, Sanders MA, Johnson ET, Schmidt-Dannert C. 2012. Engineered protein nano-compartments for targeted enzyme localization. PLoS One 7:e33342. http://dx.doi.org/10.1371/journal.pone.0033342.
43. Daniel R, Bobik TA, Gottschalk G. 1998. Biochemistry of coenzyme B_{12}-dependent glycerol and diol dehydratases and organization of the encoding genes. FEMS Microbiol. Rev. 22:553–566. http://dx.doi.org/10.1111/j.1574-6976.1998.tb00387.x.
44. Tobimatsu T, Kawata M, Toraya T. 2005. The N-terminal regions of beta and gamma subunits lower the solubility of adenosylcobalamin-dependent diol dehydratase. Biosci. Biotechnol. Biochem. 69:455–462. http://dx.doi.org/10.1271/bbb.69.455.
45. Sutter M, Boehringer D, Gutmann S, Gunther S, Prangishvili D, Loessner MJ, Stetter KO, Weber-Ban E, Ban N. 2008. Structural basis of enzyme encapsulation into a bacterial nanocompartment. Nat. Struct. Mol. Biol. 15:939–947. http://dx.doi.org/10.1038/nsmb.1473.
46. Kinney JN, Salmeen A, Cai F, Kerfeld CA. 2012. Elucidating essential role of conserved carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly. J. Biol. Chem. 287:17729–17736. http://dx.doi.org/10.1074/jbc.M112.355305.
47. Menon BB, Dou Z, Heinhorst S, Shively JM, Cannon GC. 2008. *Halothiobacillus neapolitanus* carboxysomes sequester heterologous and chimeric RubisCO species. PLoS One 3:e3570. http://dx.doi.org/10.1371/journal.pone.0003570.
48. Parsons JB, Frank S, Bhella D, Liang M, Prentice MB, Mulvihill DP, Warren MJ. 2010. Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. Mol. Cell 38:305–315. http://dx.doi.org/10.1016/j.molcel.2010.04.008.
49. Kim EY, Tullman-Ercek D. 2014. A rapid flow cytometry assay for the relative quantification of protein encapsulation into bacterial microcompartments. Biotechnol. J. 9:348–354. http://dx.doi.org/10.1002/biot.201300391.
50. Lawrence AD, Frank S, Newnham S, Lee MJ, Brown IR, Xue W-F, Rowe ML, Mulvihill DP, Prentice MB, Howard MJ, Warren MJ. 30 January 2014. Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. ACS Synth. Biol. http://dx.doi.org/10.1021/sb4001118.
51. Long BM, Rae BD, Badger MR, Price GD. 2011. Over-expression of the beta-carboxysomal CcmM protein in *Synechococcus* PCC7942 reveals a tight co-regulation of carboxysomal carbonic anhydrase (CcaA) and M58 content. Photosynth. Res. 109:33–45. http://dx.doi.org/10.1007/s11120-011-9659-8.
52. Long BM, Tucker L, Badger MR, Price GD. 2010. Functional cyanobacterial beta-carboxysomes have an absolute requirement for both long and short forms of the CcmM protein. Plant Physiol. 153:285–293. http://dx.doi.org/10.1104/pp.110.154948.
53. Long BM, Badger MR, Whitney SM, Price GD. 2007. Analysis of carboxysomes from *Synechococcus* PCC7942 reveals multiple Rubisco complexes with carboxysomal proteins CcmM and CcaA. J. Biol. Chem. 282:29323–29335. http://dx.doi.org/10.1074/jbc.M703896200.
54. Long B, Badger M, Whitney S, Price D. 2007. A structural role for CcmM in beta-carboxysome shell formation. Photosynth. Res. 91:223.
55. Cameron JC, Wilson SC, Bernstein SL, Kerfeld CA. 2013. Biogenesis of a bacterial organelle: the carboxysome assembly pathway. Cell 155:1131–1140. http://dx.doi.org/10.1016/j.cell.2013.10.044.
56. Chen AH, Robinson-Mosher A, Savage DF, Silver PA, Polka JK. 2013. The bacterial carbon-fixing organelle is formed by shell envelopment of preassembled cargo. PLoS One 8:e76127. http://dx.doi.org/10.1371/journal.pone.0076127.
57. Iancu CV, Morris DM, Dou Z, Heinhorst S, Cannon GC, Jensen GJ. 2010. Organization, structure, and assembly of a-carboxysomes determined by electron cryotomography of intact cells. J. Mol. Biol. 396:105–117. http://dx.doi.org/10.1016/j.jmb.2009.11.019.
58. Cheng S, Sinha S, Fan C, Liu Y, Bobik TA. 2011. Genetic analysis of the protein shell of the microcompartments involved in coenzyme B_{12}-dependent 1,2-propanediol degradation by *Salmonella*. J. Bacteriol. 193:1385–1392. http://dx.doi.org/10.1128/JB.01473-10.
59. Bonacci W, Teng PK, Afonso B, Niederholtmeyer H, Grob P, Silver PA, Savage DF. 2012. Modularity of a carbon-fixing protein organelle. Proc. Natl. Acad. Sci. U. S. A. 109:478–483. http://dx.doi.org/10.1073/pnas.1108557109.
60. Parsons JB, Dinesh SD, Deery E, Leech HK, Brindley AA, Heldt D, Frank S, Smales CM, Lunsdorf H, Rambach A, Gass MH, Bleloch A, McClean KJ, Munro AW, Rigby SE, Warren MJ, Prentice MB. 2008. Biochemical and structural insights into bacterial organelle form and biogenesis. J. Biol. Chem. 283:14366–14375. http://dx.doi.org/10.1074/jbc.M709214200.
61. Tanaka S, Sawaya MR, Phillips M, Yeates TO. 2009. Insights from multiple structures of the shell proteins from the beta-carboxysome. Protein Sci. 18:108–120. http://dx.doi.org/10.1002/pro.14.
62. Tanaka S, Kerfeld CA, Sawaya MR, Cai F, Heinhorst S, Cannon GC, Yeates TO. 2008. Atomic-level models of the bacterial carboxysome shell. Science 319:1083–1086. http://dx.doi.org/10.1126/science.1151458.
63. Kerfeld CA, Sawaya MR, Tanaka S, Nguyen CV, Phillips M, Beeby M, Yeates TO. 2005. Protein structures forming the shell of primitive bacterial organelles. Science 309:936–938. http://dx.doi.org/10.1126/science.1113397.
64. Dryden KA, Crowley CS, Tanaka S, Yeates TO, Yeager M. 2009. Two-dimensional crystals of carboxysome shell proteins recapitulate the hexagonal packing of three-dimensional crystals. Protein Sci. 18:2629–2635. http://dx.doi.org/10.1002/pro.272.
65. 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. http://dx.doi.org/10.1073/pnas.0902324106.
66. Klein MG, Zwart P, Bagby SC, Cai F, Chisholm SW, Heinhorst S, Cannon GC, Kerfeld CA. 2009. Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. J. Mol. Biol. 392:319–333. http://dx.doi.org/10.1016/j.jmb.2009.03.056.
67. Heldt D, Frank S, Seyedarabi A, Ladikis D, Parsons JB, Warren MJ, Pickersgill RW. 2009. Structure of a trimeric bacterial microcompartment shell protein, EtuB, associated with ethanol utilization in *Clostridium kluyveri*. Biochem. J. 423:199–207. http://dx.doi.org/10.1042/BJ20090780.
68. Crowley CS, Cascio D, Sawaya MR, Kopstein JS, Bobik TA, Yeates TO. 2010. Structural insight into the mechanisms of transport across the *Salmonella enterica* Pdu microcompartment shell. J. Biol. Chem. 285:37838–37846. http://dx.doi.org/10.1074/jbc.M110.160580.
69. Crowley CS, Sawaya MR, Bobik TA, Yeates TO. 2008. Structure of the PduU shell protein from the Pdu microcompartment of Salmonella. Structure 16:1324–1332. http://dx.doi.org/10.1016/j.str.2008.05.013.
70. Cai F, Sutter M, Cameron JC, Stanley DN, Kinney JN, Kerfeld CA. 2013. The structure of CcmP, a tandem bacterial microcompartment domain protein from the beta-carboxysome, forms a subcompartment within a microcompartment. J. Biol. Chem. 288:16055–16063. http://dx.doi.org/10.1074/jbc.M113.456897.
71. Samborska B, Kimber MS. 2012. A dodecameric CcmK2 structure suggests beta-carboxysomal shell facets have a double-layered organization. Structure 20:1353–1362. http://dx.doi.org/10.1016/j.str.2012.05.013.
72. Price GD, Howitt SM, Harrison K, Badger MR. 1993. Analysis of a genomic DNA region from the cyanobacterium *Synechococcus* sp. strain PCC7942 involved in carboxysome assembly and function. J. Bacteriol. 175:2871–2879.
73. Cai F, Menon BB, Cannon GC, Curry KJ, Shively JM, Heinhorst S. 2009. The pentameric vertex proteins are necessary for the icosahedral

carboxysome shell to function as a CO_2 leakage barrier. PLoS One 4:e7521. http://dx.doi.org/10.1371/journal.pone.0007521.
74. **Forouhar F, Kuzin A, Seetharaman J, Lee I, Zhou W, Abashidze M, Chen Y, Yong W, Janjua H, Fang Y, Wang D, Cunningham K, Xiao R, Acton TB, Pichersky E, Klessig DF, Porter CW, Montelione GT, Tong L.** 2007. Functional insights from structural genomics. J. Struct. Funct. Genomics 8:37–44. http://dx.doi.org/10.1007/s10969-007-9018-3.
75. **Wheatley NM, Gidaniyan SD, Liu Y, Cascio D, Yeates TO.** 2013. Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. Protein Sci. 22:660–665. http://dx.doi.org/10.1002/pro.2246.
76. **Tanaka S, Sawaya MR, Yeates TO.** 2010. Structure and mechanisms of a protein-based organelle in *Escherichia coli*. Science 327:81–84. http://dx.doi.org/10.1126/science.1179513.
77. **Pang A, Liang M, Prentice MB, Pickersgill RW.** 2012. Substrate channels revealed in the trimeric *Lactobacillus reuteri* bacterial microcompartment shell protein PduB. Acta Crystallogr. D Biol. Crystallogr. 68:1642–1652. http://dx.doi.org/10.1107/S0907444912039315.
78. **Pang A, Warren MJ, Pickersgill RW.** 2011. Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe-4S cluster-binding site. Acta Crystallogr. D Biol. Crystallogr. 67:91–96. http://dx.doi.org/10.1107/S0907444910050201.
79. **Cheng S, Fan C, Sinha S, Bobik TA.** 2012. The PduQ enzyme is an alcohol dehydrogenase used to recycle NAD^+ internally within the Pdu microcompartment of *Salmonella enterica*. PLoS One 7:e47144. http://dx.doi.org/10.1371/journal.pone.0047144.
80. **Cheng S, Bobik TA.** 2010. Characterization of the PduS cobalamin reductase of *Salmonella enterica* and its role in the Pdu microcompartment. J. Bacteriol. 192:5071–5080. http://dx.doi.org/10.1128/JB.00575-10.
81. **Huseby DL, Roth JR.** 2013. Evidence that a metabolic microcompartment contains and recycles private cofactor pools. J. Bacteriol. 195:2864–2879. http://dx.doi.org/10.1128/JB.02179-12.
82. **Havemann GD, Bobik TA.** 2003. Protein content of polyhedral organelles involved in coenzyme B_{12}-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar Typhimurium LT2. J. Bacteriol. 185:5086–5095. http://dx.doi.org/10.1128/JB.185.17.5086-5095.2003.
83. **Leal NA, Havemann GD, Bobik TA.** 2003. PduP is a coenzyme-A-acylating propionaldehyde dehydrogenase associated with the polyhedral bodies involved in B_{12}-dependent 1,2-propanediol degradation by *Salmonella enterica* serovar Typhimurium LT2. Arch. Microbiol. 180:353–361. http://dx.doi.org/10.1007/s00203-003-0601-0.
84. **Sinha S, Cheng S, Fan C, Bobik TA.** 2012. The PduM protein is a structural component of the microcompartments involved in coenzyme B_{12}-dependent 1,2-propanediol degradation by *Salmonella enterica*. J. Bacteriol. 194:1912–1918. http://dx.doi.org/10.1128/JB.06529-11.
85. **Toraya T, Honda S, Fukui S.** 1979. Fermentation of 1,2-propanediol with 1,2-ethanediol by some genera of *Enterobacteriaceae*, involving coenzyme B_{12}-dependent diol dehydratase. J. Bacteriol. 139:39–47.
86. **Obradors N, Badia J, Baldoma L, Aguilar J.** 1988. Anaerobic metabolism of the L-rhamnose fermentation product 1,2-propanediol in *Salmonella typhimurium*. J. Bacteriol. 170:2159–2162.
87. **Jeter RM.** 1990. Cobalamin-dependent 1,2-propanediol utilization by *Salmonella typhimurium*. J. Gen. Microbiol. 136:887–896. http://dx.doi.org/10.1099/00221287-136-5-887.
88. **Roth JR, Lawrence JG, Bobik TA.** 1996. Cobalamin (coenzyme B_{12}): synthesis and biological significance. Annu. Rev. Microbiol. 50:137–181. http://dx.doi.org/10.1146/annurev.micro.50.1.137.
89. **Liu Y, Leal NA, Sampson EM, Johnson CL, Havemann GD, Bobik TA.** 2007. PduL is an evolutionarily distinct phosphotransacylase involved in B_{12}-dependent 1,2-propanediol degradation by *Salmonella enterica* serovar Typhimurium LT2. J. Bacteriol. 189:1589–1596. http://dx.doi.org/10.1128/JB.01151-06.
90. **Horswill AR, Escalante-Semerena JC.** 1997. Propionate catabolism in *Salmonella typhimurium* LT2: two divergently transcribed units comprise the *prp* locus at 8.5 centisomes, *prpR* encodes a member of the sigma-54 family of activators, and the *prpBCDE* genes constitute an operon. J. Bacteriol. 179:928–940.
91. **Horswill AR, Escalante-Semerena JC.** 1999. *Salmonella typhimurium* LT2 catabolizes propionate via the 2-methylcitric acid cycle. J. Bacteriol. 181:5615–5623.
92. **Palacios S, Starai VJ, Escalante-Semerena JC.** 2003. Propionyl coenzyme A is a common intermediate in the 1,2-propanediol and propionate catabolic pathways needed for expression of the *prpBCDE* operon during growth of *Salmonella enterica* on 1,2-propanediol. J. Bacteriol. 185:2802–2810. http://dx.doi.org/10.1128/JB.185.9.2802-2810.2003.
93. **Jeter RM, Olivera BM, Roth JR.** 1984. *Salmonella typhimurium* synthesizes cobalamin (vitamin B_{12}) de novo under anaerobic growth conditions. J. Bacteriol. 159:206–213.
94. **Chen P, Ailion M, Bobik T, Stormo G, Roth J.** 1995. Five promoters integrate control of the *cob/pdu* regulon in *Salmonella typhimurium*. J. Bacteriol. 177:5401–5410.
95. **Bobik TA, Ailion M, Roth JR.** 1992. A single regulatory gene integrates control of vitamin B_{12} synthesis and propanediol degradation. J. Bacteriol. 174:2253–2266.
96. **Rondon MR, Escalante-Semerena JC.** 1992. The *poc* locus is required for 1,2-propanediol-dependent transcription of the cobalamin biosynthetic (*cob*) and propanediol utilization (*pdu*) genes of *Salmonella typhimurium*. J. Bacteriol. 174:2267–2272.
97. **Roth JR, Lawrence JG, Rubenfield M, Kieffer-Higgins S, Church GM.** 1993. Characterization of the cobalamin (vitamin B_{12}) biosynthetic genes of *Salmonella typhimurium*. J. Bacteriol. 175:3303–3016.
98. **Andersson DI.** 1992. Involvement of the Arc system in redox regulation of the Cob operon in *Salmonella typhimurium*. Mol. Microbiol. 6:1491–1494. http://dx.doi.org/10.1111/j.1365-2958.1992.tb00869.x.
99. **Rondon MR, Escalante-Semerena JC.** 1996. In vitro analysis of the interactions between the PocR regulatory protein and the promoter region of the cobalamin biosynthetic (*cob*) operon of *Salmonella typhimurium* LT2. J. Bacteriol. 178:2196–2203.
100. **Ailion M, Bobik TA, Roth JR.** 1993. Two global regulatory systems (Crp and Arc) control the cobalamin/propanediol regulon of *Salmonella typhimurium*. J. Bacteriol. 175:7200–7208.
101. **Escalante-Semerena JC, Roth JR.** 1987. Regulation of cobalamin biosynthetic operons in *Salmonella typhimurium*. J. Bacteriol. 169:2251–2258.
102. **Rondon MR, Kazmierczak R, Escalante-Semerena JC.** 1995. Glutathione is required for maximal transcription of the cobalamin biosynthetic and 1,2-propanediol utilization (*cob/pdu*) regulon and for the catabolism of ethanolamine, 1,2-propanediol, and propionate in *Salmonella typhimurium* LT2. J. Bacteriol. 177:5434–5439.
103. **Chang MC, Keasling JD.** 2006. Production of isoprenoid pharmaceuticals by engineered microbes. Nat. Chem. Biol. 2:674–681. http://dx.doi.org/10.1038/nchembio836.
104. **Johnson CL, Buszko ML, Bobik TA.** 2004. Purification and initial characterization of the *Salmonella enterica* PduO ATP:cob(I)alamin adenosyltransferase. J. Bacteriol. 186:7881–7887. http://dx.doi.org/10.1128/JB.186.23.7881-7887.2004.
105. **Johnson CL, Pechonick E, Park SD, Havemann GD, Leal NA, Bobik TA.** 2001. Functional genomic, biochemical, and genetic characterization of the *Salmonella pduO* gene, an ATP:cob(I)alamin adenosyltransferase gene. J. Bacteriol. 183:1577–1584. http://dx.doi.org/10.1128/JB.183.5.1577-1584.2001.
106. **Parsons JB, Lawrence AD, McLean KJ, Munro AW, Rigby SE, Warren MJ.** 2010. Characterisation of PduS, the pdu metabolosome corrin reductase, and evidence of substructural organisation within the bacterial microcompartment. PLoS One 5:e14009. http://dx.doi.org/10.1371/journal.pone.0014009.
107. **Sampson EM, Johnson CL, Bobik TA.** 2005. Biochemical evidence that the pduS gene encodes a bifunctional cobalamin reductase. Microbiology 151:1169–1177. http://dx.doi.org/10.1099/mic.0.27755-0.
108. **Abeles RH, Lee HA, Jr.** 1961. An intramolecular oxidation-reduction requiring a cobamide coenzyme. J. Biol. Chem. 236:2347–2350.
109. **Walter D, Ailion M, Roth J.** 1997. Genetic characterization of the *pdu* operon: use of 1,2-propanediol in *Salmonella typhimurium*. J. Bacteriol. 179:1013–1022.
110. **Poznanskaja AA, Tanizawa K, Soda K, Toraya T, Fukui S.** 1979. Coenzyme B_{12}-dependent diol dehydrase: purification, subunit heterogeneity, and reversible association. Arch. Biochem. Biophys. 194:379–386. http://dx.doi.org/10.1016/0003-9861(79)90630-1.
111. **Sauvageot N, Muller C, Hartke A, Auffray Y, Laplace J-M.** 2002. Characterisation of the diol dehydratase *pdu* operon of *Lactobacillus collinoides*. FEMS Microbiol. Lett. 209:69–74. http://dx.doi.org/10.1016/S0378-1097(02)00471-8.
112. **Sauvageot N, Pichereau V, Louarme L, Hartke A, Auffray Y, Laplace JM.** 2002. Purification, characterization and subunits identification of the diol dehydratase of *Lactobacillus collinoides*. Eur. J. Biochem. 269:5731–5737. http://dx.doi.org/10.1046/j.1432-1033.2002.03288.x.

carboxysome shell to function as a CO_2 leakage barrier. PLoS One 4:e7521. http://dx.doi.org/10.1371/journal.pone.0007521.
74. **Forouhar F, Kuzin A, Seetharaman J, Lee I, Zhou W, Abashidze M, Chen Y, Yong W, Janjua H, Fang Y, Wang D, Cunningham K, Xiao R, Acton TB, Pichersky E, Klessig DF, Porter CW, Montelione GT, Tong L.** 2007. Functional insights from structural genomics. J. Struct. Funct. Genomics 8:37–44. http://dx.doi.org/10.1007/s10969-007-9018-3.
75. **Wheatley NM, Gidaniyan SD, Liu Y, Cascio D, Yeates TO.** 2013. Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. Protein Sci. 22:660–665. http://dx.doi.org/10.1002/pro.2246.
76. **Tanaka S, Sawaya MR, Yeates TO.** 2010. Structure and mechanisms of a protein-based organelle in *Escherichia coli*. Science 327:81–84. http://dx.doi.org/10.1126/science.1179513.
77. **Pang A, Liang M, Prentice MB, Pickersgill RW.** 2012. Substrate channels revealed in the trimeric *Lactobacillus reuteri* bacterial microcompartment shell protein PduB. Acta Crystallogr. D Biol. Crystallogr. 68:1642–1652. http://dx.doi.org/10.1107/S0907444912039315.
78. **Pang A, Warren MJ, Pickersgill RW.** 2011. Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe-4S cluster-binding site. Acta Crystallogr. D Biol. Crystallogr. 67:91–96. http://dx.doi.org/10.1107/S0907444910050201.
79. **Cheng S, Fan C, Sinha S, Bobik TA.** 2012. The PduQ enzyme is an alcohol dehydrogenase used to recycle NAD^+ internally within the Pdu microcompartment of *Salmonella enterica*. PLoS One 7:e47144. http://dx.doi.org/10.1371/journal.pone.0047144.
80. **Cheng S, Bobik TA.** 2010. Characterization of the PduS cobalamin reductase of *Salmonella enterica* and its role in the Pdu microcompartment. J. Bacteriol. 192:5071–5080. http://dx.doi.org/10.1128/JB.00575-10.
81. **Huseby DL, Roth JR.** 2013. Evidence that a metabolic microcompartment contains and recycles private cofactor pools. J. Bacteriol. 195:2864–2879. http://dx.doi.org/10.1128/JB.02179-12.
82. **Havemann GD, Bobik TA.** 2003. Protein content of polyhedral organelles involved in coenzyme B_{12}-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar Typhimurium LT2. J. Bacteriol. 185:5086–5095. http://dx.doi.org/10.1128/JB.185.17.5086-5095.2003.
83. **Leal NA, Havemann GD, Bobik TA.** 2003. PduP is a coenzyme-A-acylating propionaldehyde dehydrogenase associated with the polyhedral bodies involved in B_{12}-dependent 1,2-propanediol degradation by *Salmonella enterica* serovar Typhimurium LT2. Arch. Microbiol. 180:353–361. http://dx.doi.org/10.1007/s00203-003-0601-0.
84. **Sinha S, Cheng S, Fan C, Bobik TA.** 2012. The PduM protein is a structural component of the microcompartments involved in coenzyme B_{12}-dependent 1,2-propanediol degradation by *Salmonella enterica*. J. Bacteriol. 194:1912–1918. http://dx.doi.org/10.1128/JB.06529-11.
85. **Toraya T, Honda S, Fukui S.** 1979. Fermentation of 1,2-propanediol with 1,2-ethanediol by some genera of *Enterobacteriaceae*, involving coenzyme B_{12}-dependent diol dehydratase. J. Bacteriol. 139:39–47.
86. **Obradors N, Badia J, Baldoma L, Aguilar J.** 1988. Anaerobic metabolism of the L-rhamnose fermentation product 1,2-propanediol in *Salmonella typhimurium*. J. Bacteriol. 170:2159–2162.
87. **Jeter RM.** 1990. Cobalamin-dependent 1,2-propanediol utilization by *Salmonella typhimurium*. J. Gen. Microbiol. 136:887–896. http://dx.doi.org/10.1099/00221287-136-5-887.
88. **Roth JR, Lawrence JG, Bobik TA.** 1996. Cobalamin (coenzyme B_{12}): synthesis and biological significance. Annu. Rev. Microbiol. 50:137–181. http://dx.doi.org/10.1146/annurev.micro.50.1.137.
89. **Liu Y, Leal NA, Sampson EM, Johnson CL, Havemann GD, Bobik TA.** 2007. PduL is an evolutionarily distinct phosphotransacylase involved in B_{12}-dependent 1,2-propanediol degradation by *Salmonella enterica* serovar Typhimurium LT2. J. Bacteriol. 189:1589–1596. http://dx.doi.org/10.1128/JB.01151-06.
90. **Horswill AR, Escalante-Semerena JC.** 1997. Propionate catabolism in *Salmonella typhimurium* LT2: two divergently transcribed units comprise the *prp* locus at 8.5 centisomes, *prpR* encodes a member of the sigma-54 family of activators, and the *prpBCDE* genes constitute an operon. J. Bacteriol. 179:928–940.
91. **Horswill AR, Escalante-Semerena JC.** 1999. *Salmonella typhimurium* LT2 catabolizes propionate via the 2-methylcitric acid cycle. J. Bacteriol. 181:5615–5623.
92. **Palacios S, Starai VJ, Escalante-Semerena JC.** 2003. Propionyl coenzyme A is a common intermediate in the 1,2-propanediol and propionate catabolic pathways needed for expression of the *prpBCDE* operon during growth of *Salmonella enterica* on 1,2-propanediol. J. Bacteriol. 185:2802–2810. http://dx.doi.org/10.1128/JB.185.9.2802-2810.2003.
93. **Jeter RM, Olivera BM, Roth JR.** 1984. *Salmonella typhimurium* synthesizes cobalamin (vitamin B_{12}) de novo under anaerobic growth conditions. J. Bacteriol. 159:206–213.
94. **Chen P, Ailion M, Bobik T, Stormo G, Roth J.** 1995. Five promoters integrate control of the *cob/pdu* regulon in *Salmonella typhimurium*. J. Bacteriol. 177:5401–5410.
95. **Bobik TA, Ailion M, Roth JR.** 1992. A single regulatory gene integrates control of vitamin B_{12} synthesis and propanediol degradation. J. Bacteriol. 174:2253–2266.
96. **Rondon MR, Escalante-Semerena JC.** 1992. The *poc* locus is required for 1,2-propanediol-dependent transcription of the cobalamin biosynthetic (*cob*) and propanediol utilization (*pdu*) genes of *Salmonella typhimurium*. J. Bacteriol. 174:2267–2272.
97. **Roth JR, Lawrence JG, Rubenfield M, Kieffer-Higgins S, Church GM.** 1993. Characterization of the cobalamin (vitamin B_{12}) biosynthetic genes of *Salmonella typhimurium*. J. Bacteriol. 175:3303–3016.
98. **Andersson DI.** 1992. Involvement of the Arc system in redox regulation of the Cob operon in *Salmonella typhimurium*. Mol. Microbiol. 6:1491–1494. http://dx.doi.org/10.1111/j.1365-2958.1992.tb00869.x.
99. **Rondon MR, Escalante-Semerena JC.** 1996. In vitro analysis of the interactions between the PocR regulatory protein and the promoter region of the cobalamin biosynthetic (*cob*) operon of *Salmonella typhimurium* LT2. J. Bacteriol. 178:2196–2203.
100. **Ailion M, Bobik TA, Roth JR.** 1993. Two global regulatory systems (Crp and Arc) control the cobalamin/propanediol regulon of *Salmonella typhimurium*. J. Bacteriol. 175:7200–7208.
101. **Escalante-Semerena JC, Roth JR.** 1987. Regulation of cobalamin biosynthetic operons in *Salmonella typhimurium*. J. Bacteriol. 169:2251–2258.
102. **Rondon MR, Kazmierczak R, Escalante-Semerena JC.** 1995. Glutathione is required for maximal transcription of the cobalamin biosynthetic and 1,2-propanediol utilization (*cob/pdu*) regulon and for the catabolism of ethanolamine, 1,2-propanediol, and propionate in *Salmonella typhimurium* LT2. J. Bacteriol. 177:5434–5439.
103. **Chang MC, Keasling JD.** 2006. Production of isoprenoid pharmaceuticals by engineered microbes. Nat. Chem. Biol. 2:674–681. http://dx.doi.org/10.1038/nchembio836.
104. **Johnson CL, Buszko ML, Bobik TA.** 2004. Purification and initial characterization of the *Salmonella enterica* PduO ATP:cob(I)alamin adenosyltransferase. J. Bacteriol. 186:7881–7887. http://dx.doi.org/10.1128/JB.186.23.7881-7887.2004.
105. **Johnson CL, Pechonick E, Park SD, Havemann GD, Leal NA, Bobik TA.** 2001. Functional genomic, biochemical, and genetic characterization of the *Salmonella pduO* gene, an ATP:cob(I)alamin adenosyltransferase gene. J. Bacteriol. 183:1577–1584. http://dx.doi.org/10.1128/JB.183.5.1577-1584.2001.
106. **Parsons JB, Lawrence AD, McLean KJ, Munro AW, Rigby SE, Warren MJ.** 2010. Characterisation of PduS, the pdu metabolosome corrin reductase, and evidence of substructural organisation within the bacterial microcompartment. PLoS One 5:e14009. http://dx.doi.org/10.1371/journal.pone.0014009.
107. **Sampson EM, Johnson CL, Bobik TA.** 2005. Biochemical evidence that the pduS gene encodes a bifunctional cobalamin reductase. Microbiology 151:1169–1177. http://dx.doi.org/10.1099/mic.0.27755-0.
108. **Abeles RH, Lee HA, Jr.** 1961. An intramolecular oxidation-reduction requiring a cobamide coenzyme. J. Biol. Chem. 236:2347–2350.
109. **Walter D, Ailion M, Roth J.** 1997. Genetic characterization of the *pdu* operon: use of 1,2-propanediol in *Salmonella typhimurium*. J. Bacteriol. 179:1013–1022.
110. **Poznanskaja AA, Tanizawa K, Soda K, Toraya T, Fukui S.** 1979. Coenzyme B_{12}-dependent diol dehydrase: purification, subunit heterogeneity, and reversible association. Arch. Biochem. Biophys. 194:379–386. http://dx.doi.org/10.1016/0003-9861(79)90630-1.
111. **Sauvageot N, Muller C, Hartke A, Auffray Y, Laplace J-M.** 2002. Characterisation of the diol dehydratase *pdu* operon of *Lactobacillus collinoides*. FEMS Microbiol. Lett. 209:69–74. http://dx.doi.org/10.1016/S0378-1097(02)00471-8.
112. **Sauvageot N, Pichereau V, Louarme L, Hartke A, Auffray Y, Laplace JM.** 2002. Purification, characterization and subunits identification of the diol dehydratase of *Lactobacillus collinoides*. Eur. J. Biochem. 269:5731–5737. http://dx.doi.org/10.1046/j.1432-1033.2002.03288.x.

113. **Tobimatsu T, Sakai T, Hashida Y, Mizoguchi N, Miyoshi S, Toraya T.** 1997. Heterologous expression, purification, and properties of diol dehydratase, an adenosylcobalamin-dependent enzyme of *Klebsiella oxytoca*. Arch. Biochem. Biophys. **347**:132–140. http://dx.doi.org/10.1006/abbi.1997.0325.
114. **Toraya T, Honda S, Mori K.** 2010. Coenzyme B12-dependent diol dehydratase is a potassium ion-requiring calcium metalloenzyme: evidence that the substrate-coordinated metal ion is calcium. Biochemistry **49**:7210–7217. http://dx.doi.org/10.1021/bi100561m.
115. **Shibata N, Masuda J, Tobimatsu T, Toraya T, Suto K, Morimoto Y, Yasuoka N.** 1999. A new mode of B_{12} binding and the direct participation of a potassium ion in enzyme catalysis: X-ray structure of diol dehydratase. Structure **7**:997–1008. http://dx.doi.org/10.1016/S0969-2126(99)80126-9.
116. **Luo LH, Kim CH, Heo SY, Oh BR, Hong WK, Kim S, Kim DH, Seo JW.** 2012. Production of 3-hydroxypropionic acid through propionaldehyde dehydrogenase PduP mediated biosynthetic pathway in *Klebsiella pneumoniae*. Bioresour. Technol. **103**:1–6. http://dx.doi.org/10.1016/j.biortech.2011.09.099.
117. **Luo LH, Seo JW, Baek JO, Oh BR, Heo SY, Hong WK, Kim DH, Kim CH.** 2011. Identification and characterization of the propanediol utilization protein PduP of *Lactobacillus reuteri* for 3-hydroxypropionic acid production from glycerol. Appl. Microbiol. Biotechnol. **89**:697–703. http://dx.doi.org/10.1007/s00253-010-2887-6.
118. **Tsang AW, Horswill AR, Escalante-Semerena JC.** 1998. Studies of regulation of expression of the propionate (*prpBCDE*) operon provide insights into how *Salmonella typhimurium* LT2 integrates its 1,2-propanediol and propionate catabolic pathways. J. Bacteriol. **180**:6511–6518.
119. **Honda S, Toraya T, Fukui S.** 1980. In situ reactivation of glycerol-inactivated coenzyme B_{12}-dependent enzymes, glycerol dehydratase and diol dehydratase. J. Bacteriol. **143**:1458–1465.
120. **Ushio K, Honda S, Toraya T, Fukui S.** 1982. The mechanism of in situ reactivation of glycerol-inactivated coenzyme B_{12}-dependent enzymes, glycerol dehydratase and diol dehydratase. J. Nutr. Sci. Vitaminol. (Tokyo) **28**:225–236. http://dx.doi.org/10.3177/jnsv.28.225.
121. **Toraya T.** 2000. Radical catalysis of B_{12} enzymes: structure, mechanism, inactivation, and reactivation of diol and glycerol dehydratases. Cell. Mol. Life Sci. **57**:106–127. http://dx.doi.org/10.1007/s000180050502.
122. **Mori K, Toraya T.** 1999. Mechanism of reactivation of coenzyme B_{12}-dependent diol dehydratase by a molecular chaperone-like reactivating factor. Biochemistry **38**:13170–13178. http://dx.doi.org/10.1021/bi9911738.
123. **Mori K, Tobimatsu T, Hara T, Toraya T.** 1997. Characterization, sequencing, and expression of the genes encoding a reactivating factor for glycerol-inactivated adenosylcobalamin-dependent diol dehydratase. J. Biol. Chem. **272**:32034–32041. http://dx.doi.org/10.1074/jbc.272.51.32034.
124. **Mori K, Tobimatsu T, Toraya T.** 1997. A protein factor is essential for in situ reactivation of glycerol-inactivated adenosylcobalamin-dependent diol dehydratase. Biosci. Biotechnol. Biochem. **61**:1729–1733. http://dx.doi.org/10.1271/bbb.61.1729.
125. **Tobimatsu T, Kajiura H, Yunoki M, Azuma M, Toraya T.** 1999. Identification and expression of the genes encoding a reactivating factor for adenosylcobalamin-dependent glycerol dehydratase. J. Bacteriol. **181**:4110–4113.
126. **Toraya T, Mori K.** 1999. A reactivating factor for coenzyme B_{12}-dependent diol dehydratase. J. Biol. Chem. **274**:3372–3377. http://dx.doi.org/10.1074/jbc.274.6.3372.
127. **Toraya T, Mori K, Hara T, Tobimatsu T.** 2000. A reactivating factor for coenzyme B_{12}-dependent diol dehydratase. Biofactors **11**:105–107. http://dx.doi.org/10.1002/biof.5520110131.
128. **Escalante-Semerena JC, Suh SJ, Roth JR.** 1990. *cobA* function is required for both de novo cobalamin biosynthesis and assimilation of exogenous corrinoids in *Salmonella typhimurium*. J. Bacteriol. **172**:273–280.
129. **Fonseca MV, Escalante-Semerena JC.** 2000. Reduction of cob(III)alamin to cob(II)alamin in *Salmonella enterica* serovar Typhimurium LT2. J. Bacteriol. **182**:4304–4309. http://dx.doi.org/10.1128/JB.182.15.4304-4309.2000.
130. **Fonseca MV, Escalante-Semerena JC.** 2001. An in vitro reducing system for the enzymic conversion of cobalamin to adenosylcobalamin. J. Biol. Chem. **276**:32101–32108. http://dx.doi.org/10.1074/jbc.M102510200.
131. **Pena KL, Castel SE, de Araujo C, Espie GS, Kimber MS.** 2010. Structural basis of the oxidative activation of the carboxysomal g-carbonic anhydrase, CcmM. Proc. Natl. Acad. Sci. U. S. A. **107**:2455–2460. http://dx.doi.org/10.1073/pnas.0910866107.
132. **Park K, Mera PE, Escalante-Semerena JC, Brunold TC.** 2008. Kinetic and spectroscopic studies of the ATP:corrinoid adenosyltransferase PduO from *Lactobacillus reuteri*: substrate specificity and insights into the mechanism of Co(II)corrinoid reduction. Biochemistry **47**:9007–9015. http://dx.doi.org/10.1021/bi800419e.
133. **St Maurice M, Mera PE, Taranto MP, Sesma F, Escalante-Semerena JC, Rayment I.** 2007. Structural characterization of the active site of the PduO-type ATP:co(I)rrinoid adenosyltransferase from *Lactobacillus reuteri*. J. Biol. Chem. **282**:2596–2605. http://dx.doi.org/10.1074/jbc.M609557200.
134. **Stich TA, Buan NR, Escalante-Semerena JC, Brunold TC.** 2005. Spectroscopic and computational studies of the ATP:corrinoid adenosyltransferase (CobA) from *Salmonella enterica*: insights into the mechanism of adenosylcobalamin biosynthesis. J. Am. Chem. Soc. **127**:8710–8719. http://dx.doi.org/10.1021/ja042142p.
135. **Stich TA, Yamanishi M, Banerjee R, Brunold TC.** 2005. Spectroscopic evidence for the formation of a four-coordinate Co^{2+} cobalamin species upon binding to the human ATP:cobalamin adenosyltransferase. J. Am. Chem. Soc. **127**:7660–7661. http://dx.doi.org/10.1021/ja050546r.
136. **Leal NA, Park SD, Kima PE, Bobik TA.** 2003. Identification of the human and bovine ATP:cob(I)alamin adenosyltransferase cDNAs based on complementation of a bacterial mutant. J. Biol. Chem. **278**:9227–9234. http://dx.doi.org/10.1074/jbc.M212739200.
137. **Dobson CM, Wai T, Leclerc D, Kadir H, Narang M, Lerner-Ellis JP, Hudson TJ, Rosenblatt DS, Gravel RA.** 2002. Identification of the gene responsible for the *cblB* complementation group of vitamin B_{12}-dependent methylmalonic aciduria. Hum. Mol. Genet. **11**:3361–3369. http://dx.doi.org/10.1093/hmg/11.26.3361.
138. **Zhang J, Dobson CM, Wu XC, Lerner-Ellis J, Rosenblatt DS, Gravel RA.** 2006. Impact of *cblB* mutations on the function of ATP:cob(I)alamin adenosyltransferase in disorders of vitamin B_{12} metabolism. Mol. Genet. Metab. **87**:315–322. http://dx.doi.org/10.1016/j.ymgme.2005.12.003.
139. **Mera PE, St Maurice M, Rayment I, Escalante-Semerena JC.** 2007. Structural and functional analyses of the human-type corrinoid adenosyltransferase (PduO) from *Lactobacillus reuteri*. Biochemistry **46**:13829–13836. http://dx.doi.org/10.1021/bi701622j.
140. **Saridakis V, Yakunin A, Xu X, Anandakumar P, Pennycooke M, Gu J, Cheung F, Lew JM, Sanishvili R, Joachimiak A, Arrowsmith CH, Christendat D, Edwards AM.** 2004. The structural basis for methylmalonic aciduria. The crystal structure of archaeal ATP:cobalamin adenosyltransferase. J. Biol. Chem. **279**:23646–23653. http://dx.doi.org/10.1074/jbc.M401395200.
141. **Mera PE, St Maurice M, Rayment I, Escalante-Semerena JC.** 2009. Residue Phe112 of the human-type corrinoid adenosyltransferase (PduO) enzyme of *Lactobacillus reuteri* is critical to the formation of the four-coordinate Co(II) corrinoid substrate and to the activity of the enzyme. Biochemistry **48**:3138–3145. http://dx.doi.org/10.1021/bi9000134.
142. **Toraya T, Shirakashi T, Kosuga T, Fukui S.** 1976. Substrate specificity of coenzyme B_{12}-dependent diol dehydrase: glycerol as both a good substrate and a potent inactivator. Biochem. Biophys. Res. Commun. **69**:475–480. http://dx.doi.org/10.1016/0006-291X(76)90546-5.
143. **Schneider Z, Pawelkiewicz J.** 1966. The properties of glycerol dehydratase isolated from *Aerobacter aerogenes*, and the properties of the apoenzyme subunits. Acta Biochim. Pol. **13**:311–328.
144. **Shibata N, Mori K, Hieda N, Higuchi Y, Yamanishi M, Toraya T.** 2005. Release of a damaged cofactor from a coenzyme B_{12}-dependent enzyme: X-ray structures of diol dehydratase-reactivating factor. Structure **13**:1745–1754. http://dx.doi.org/10.1016/j.str.2005.08.011.
145. **Laganowsky A, Liu C, Sawaya MR, Whitelegge JP, Park J, Zhao M, Pensalfini A, Soriaga AB, Landau M, Teng PK, Cascio D, Glabe C, Eisenberg D.** 2012. Atomic view of a toxic amyloid small oligomer. Science **335**:1228–1231. http://dx.doi.org/10.1126/science.1213151.
146. **Schulz GE.** 2002. The structure of bacterial outer membrane proteins. Biochim. Biophys. Acta **1565**:308–317. http://dx.doi.org/10.1016/S0005-2736(02)00577-1.
147. **Fan C, Bobik TA.** 2008. The PduX enzyme of *Salmonella enterica* is an

L-threonine kinase used for coenzyme B_{12} synthesis. J. Biol. Chem. 283: 11322–11329. http://dx.doi.org/10.1074/jbc.M800287200.
148. Fan C, Fromm HJ, Bobik TA. 2009. Kinetic and functional analysis of L-threonine kinase, the PduX enzyme of *Salmonella enterica*. J. Biol. Chem. 284:20240–20248. http://dx.doi.org/10.1074/jbc.M109.027425.
149. Sriramulu DD, Liang M, Hernandez-Romero D, Raux-Deery E, Lunsdorf H, Parsons JB, Warren MJ, Prentice MB. 2008. *Lactobacillus reuteri* DSM 20016 produces cobalamin-dependent diol dehydratase in metabolosomes and metabolizes 1,2-propanediol by disproportionation. J. Bacteriol. 190:4559–4567. http://dx.doi.org/10.1128/JB.01535-07.
150. Talarico TL, Axelsson LT, Novotny J, Fiuzat M, Dobrogosz WJ. 1990. Utilization of glycerol as a hydrogen acceptor by *Lactobacillus reuteri*: purification of 1,3-propanediol:NAD oxidoreductase. Appl. Environ. Microbiol. 56:943–948.
151. Luthi-Peng Q, Dileme FB, Puhan Z. 2002. Effect of glucose on glycerol bioconversion by *Lactobacillus reuteri*. Appl. Microbiol. Biotechnol. 59: 289–296. http://dx.doi.org/10.1007/s00253-002-1002-z.
152. Axelsson LT, Chung TC, Dobrogosz WJ, Lindgren SE. 1989. Production of a broad spectrum antimicrobial substance by *Lactobacillus reuteri*. Microb. Ecol. Health Dis. 2:131–136. http://dx.doi.org/10.3109/08910608909140210.
153. Talarico TL, Casas IA, Chung TC, Dobrogosz WJ. 1988. Production and isolation of reuterin, a growth inhibitor produced by *Lactobacillus reuteri*. Antimicrob. Agents Chemother. 32:1854–1858. http://dx.doi.org/10.1128/AAC.32.12.1854.
154. Morita H, Toh H, Fukuda S, Horikawa H, Oshima K, Suzuki T, Murakami M, Hisamatsu S, Kato Y, Takizawa T, Fukuoka H, Yoshimura T, Itoh K, O'Sullivan DJ, McKay LL, Ohno H, Kikuchi J, Masaoka T, Hattori M. 2008. Comparative genome analysis of *Lactobacillus reuteri* and *Lactobacillus fermentum* reveal a genomic island for reuterin and cobalamin production. DNA Res. 15:151–161. http://dx.doi.org/10.1093/dnares/dsn009.
155. Bradbeer C. 1965. The clostridial fermentations of choline and ethanolamine. II. Requirement for a cobamide coenzyme by an ethanolamine deaminase. J. Biol. Chem. 240:4675–4681.
156. Chang GW, Chang JT. 1975. Evidence for the B_{12}-dependent enzyme ethanolamine deaminase in *Salmonella*. Nature 254:150–151. http://dx.doi.org/10.1038/254150a0.
157. Tsoy O, Ravcheev D, Mushegian A. 2009. Comparative genomics of ethanolamine utilization. J. Bacteriol. 191:7157–7164. http://dx.doi.org/10.1128/JB.00838-09.
158. Shively JM, Bradburne CE, Aldrich HC, Bobik TA, Mehlman JL, Jin S, Baker SH. 1998. Sequence homologs of the carboxysomal polypeptide CsoS1 of the thiobacilli are present in cyanobacteria and enteric bacteria that form carboxysomes-polyhedral bodies. Can. J. Bot. 76:906–916. http://dx.doi.org/10.1139/b98-088.
159. Garsin DA. 2010. Ethanolamine utilization in bacterial pathogens: roles and regulation. Nat. Rev. Microbiol. 8:290–295. http://dx.doi.org/10.1038/nrmicro2334.
160. Roof DM, Roth JR. 1988. Ethanolamine utilization in *Salmonella typhimurium*. J. Bacteriol. 170:3855–3863.
161. Roof DM, Roth JR. 1989. Functions required for vitamin B_{12}-dependent ethanolamine utilization in *Salmonella typhimurium*. J. Bacteriol. 171: 3316–3323.
162. 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. http://dx.doi.org/10.1128/JB.186.20.6845-6854.2004.
163. Buan NR, Escalante-Semerena JC. 2006. Purification and initial biochemical characterization of ATP:cob(I)alamin adenosyltransferase (EutT) enzyme of *Salmonella enterica*. J. Biol. Chem. 281:16971–16977. http://dx.doi.org/10.1074/jbc.M603069200.
164. Buan NR, Suh SJ, Escalante-Semerena JC. 2004. The *eutT* gene of *Salmonella enterica* encodes an oxygen-labile, metal-containing ATP: corrinoid adenosyltransferase enzyme. J. Bacteriol. 186:5708–5714. http://dx.doi.org/10.1128/JB.186.17.5708-5714.2004.
165. Sheppard DE, Penrod JT, Bobik T, Kofoid E, Roth JR. 2004. Evidence that a B_{12}-adenosyl transferase is encoded within the ethanolamine operon of *Salmonella enterica*. J. Bacteriol. 186:7635–7644. http://dx.doi.org/10.1128/JB.186.22.7635-7644.2004.
166. Penrod JT, Mace CC, Roth JR. 2004. A pH-sensitive function and phenotype: evidence that EutH facilitates diffusion of uncharged ethanolamine in *Salmonella enterica*. J. Bacteriol. 186:6885–6890. http://dx.doi.org/10.1128/JB.186.20.6885-6890.2004.
167. Roof DM, Roth JR. 1992. Autogenous regulation of ethanolamine utilization by a transcriptional activator of the *eut* operon in *Salmonella typhimurium*. J. Bacteriol. 174:6634–6643.
168. Sheppard DE, Roth JR. 1994. A rationale for autoinduction of a transcriptional activator: ethanolamine ammonia-lyase (EutBC) and the operon activator (EutR) compete for adenosyl-cobalamin in *Salmonella typhimurium*. J. Bacteriol. 176:1287–1296.
169. Fox KA, Ramesh A, Stearns JE, Bourgogne A, Reyes-Jara A, Winkler WC, Garsin DA. 2009. Multiple posttranscriptional regulatory mechanisms partner to control ethanolamine utilization in *Enterococcus faecalis*. Proc. Natl. Acad. Sci. U. S. A. 106:4435–4440. http://dx.doi.org/10.1073/pnas.0812194106.
170. Beeby M, Bobik TA, Yeates TO. 2009. Exploiting genomic patterns to discover new supramolecular protein assemblies. Protein Sci. 18:69–79. http://dx.doi.org/10.1002/pro.1.
171. Scott KP, Martin JC, Campbell G, Mayer CD, Flint HJ. 2006. Whole-genome transcription profiling reveals genes up-regulated by growth on fucose in the human gut bacterium "*Roseburia inulinivorans*." J. Bacteriol. 188:4340–4349. http://dx.doi.org/10.1128/JB.00137-06.
172. O'Brien JR, Raynaud C, Croux C, Girbal L, Soucaille P, Lanzilotta WN. 2004. Insight into the mechanism of the B_{12}-independent glycerol dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization. Biochemistry 43:4635–4645. http://dx.doi.org/10.1021/bi035930k.
173. Raynaud C, Sarcabal P, Meynial-Salles I, Croux C, Soucaille P. 2003. Molecular characterization of the 1,3-propanediol (1,3-PD) operon of *Clostridium butyricum*. Proc. Natl. Acad. Sci. U. S. A. 100:5010–5015. http://dx.doi.org/10.1073/pnas.0734105100.
174. Petit E, LaTouf WG, Coppi MV, Warnick TA, Currie D, Romashko I, Deshpande S, Haas K, Alvelo-Maurosa JG, Wardman C, Schnell DJ, Leschine SB, Blanchard JL. 2013. Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by *Clostridium phytofermentans*. PLoS One 8:e54337. http://dx.doi.org/10.1371/journal.pone.0054337.
175. Seedorf H, Fricke WF, Veith B, Bruggemann H, Liesegang H, Strittmatter A, Miethke M, Buckel W, Hinderberger J, Li F, Hagemeier C, Thauer RK, Gottschalk G. 2008. The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic features. Proc. Natl. Acad. Sci. U. S. A. 105:2128–2133. http://dx.doi.org/10.1073/pnas.0711093105.
176. Hurtley S. 2009. Spatial cell biology. Location, location, location. Introduction. Science 326:1205. http://dx.doi.org/10.1126/science.326.5957.1205.
177. Lopez-Gallego F, Schmidt-Dannert C. 2010. Multi-enzymatic synthesis. Curr. Opin. Chem. Biol. 14:174–183. http://dx.doi.org/10.1016/j.cbpa.2009.11.023.
178. Held M, Quin MB, Schmidt-Dannert C. 2013. Eut bacterial microcompartments: insights into their function, structure, and bioengineering applications. J. Mol. Microbiol. Biotechnol. 23:308–320. http://dx.doi.org/10.1159/000351343.
179. Chen AH, Silver PA. 2012. Designing biological compartmentalization. Trends Cell Biol. 22:662–670. http://dx.doi.org/10.1016/j.tcb.2012.07.002.
180. Gehrmann W, Elsner M. 2011. A specific fluorescence probe for hydrogen peroxide detection in peroxisomes. Free Radic. Res. 45:501–516. http://dx.doi.org/10.3109/10715762.2011.560148.
181. Moon TS, Dueber JE, Shiue E, Prather KL. 2010. Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. Metab. Eng. 12:298–305. http://dx.doi.org/10.1016/j.ymben.2010.01.003.
182. Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, Ullal AV, Prather KL, Keasling JD. 2009. Synthetic protein scaffolds provide modular control over metabolic flux. Nat. Biotechnol. 27:753–759. http://dx.doi.org/10.1038/nbt.1557.
183. Conrado RJ, Mansell TJ, Varner JD, DeLisa MP. 2007. Stochastic reaction-diffusion simulation of enzyme compartmentalization reveals improved catalytic efficiency for a synthetic metabolic pathway. Metab. Eng. 9:355–363. http://dx.doi.org/10.1016/j.ymben.2007.05.002.
184. Kim EY, Tullman-Ercek D. 2013. Engineering nanoscale protein compartments for synthetic organelles. Curr. Opin. Biotechnol. 24:627–632. http://dx.doi.org/10.1016/j.copbio.2012.11.012.
185. Worsdorfer B, Woycechowsky KJ, Hilvert D. 2011. Directed evolution

of a protein container. Science 331:589–592. http://dx.doi.org/10.1126/science.1199081.
186. Chen HN, Woycechowsky KJ. 2012. Conversion of a dodecahedral protein capsid into pentamers via minimal point mutations. Biochemistry 51:4704–4712. http://dx.doi.org/10.1021/bi3003555.
187. Lee H, DeLoache WC, Dueber JE. 2012. Spatial organization of enzymes for metabolic engineering. Metab. Eng. 14:242–251. http://dx.doi.org/10.1016/j.ymben.2011.09.003.
188. Sargent F, Davidson FA, Kelly CL, Binny R, Christodoulides N, Gibson D, Johansson E, Kozyrska K, Lado LL, Maccallum J, Montague R, Ortmann B, Owen R, Coulthurst SJ, Dupuy L, Prescott AR, Palmer T. 2013. A synthetic system for expression of components of a bacterial microcompartment. Microbiology 159:2427–2436. http://dx.doi.org/10.1099/mic.0.069922-0.
189. Lassila JK, Bernstein SL, Kinney JN, Axen SD, Kerfeld CA. 11 March 2014. Assembly of robust bacterial microcompartment shells using building blocks from an organelle of unknown function. J. Mol. Biol. http://dx.doi.org/10.1016/j.jmb.2014.02.025.
190. Tsai SJ, Yeates TO. 2011. Bacterial microcompartments insights into the structure, mechanism, and engineering applications. Prog. Mol. Biol. Transl. Sci. 103:1–20. http://dx.doi.org/10.1016/B978-0-12-415906-8.00008-X.

Chapter 3

The central pore of PduA hexamers serves as conduit for metabolite transport

Introduction

The structure of PduA solved by Crowley et al. (2010) revealed a hexamer assembly with a central pore of diameter 5.6Å, lined with positively charged residues and peptide backbone atoms[14]. The pore could easily accommodate passage of small three carbon metabolites of the Pdu pathway, so it was hypothesized that these pores are the routes that metabolites use to traverse the BMC shell. Kerfeld et al. (2005) had also previously suggested in their work on carboxysome shell proteins CcmK2 and CcmK4 that the central pores of these hexagonal assemblies (7Å and 4Å in diameter, respectively) may serve as conduits of small molecule transport[22]. However, no study to date has investigated these hypotheses.

In the following article, the PduA pore was studied using site-directed mutagenesis to mutate the pore residue Ser40 to change the pore size or the pore polarity. S40 mutants were analyzed by *in* vivo growth assays and X-ray crystallography for structure determination. Our results confirmed that the PduA hexamer pore is indeed a route of small molecule transport. We also showed that the PduA pore serves to discriminate substrate from toxic intermediate molecules via a mechanism involving the key pore residue S40. Our study highlights the importance of biochemical investigation and structural studies with atomic resolution in increasing our understanding of how BMCs function.

Selective molecular transport through the protein shell of a bacterial microcompartment organelle

Chiranjit Chowdhury[a,1], Sunny Chun[b,1], Allan Pang[c], Michael R. Sawaya[c], Sharmistha Sinha[a], Todd O. Yeates[b,c,d,2], and Thomas A. Bobik[a,2]

[a]Roy J. Carver Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University, Ames, IA 50011; [b]Molecular Biology Institute, [c]UCLA–DOE Institute for Genomics and Proteomics, and [d]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095

Edited by Johann Deisenhofer, University of Texas Southwestern Medical Center, Dallas, TX, and approved January 29, 2015 (received for review December 11, 2014)

Bacterial microcompartments are widespread prokaryotic organelles that have important and diverse roles ranging from carbon fixation to enteric pathogenesis. Current models for microcompartment function propose that their outer protein shell is selectively permeable to small molecules, but whether a protein shell can mediate selective permeability and how this occurs are unresolved questions. Here, biochemical and physiological studies of structure-guided mutants are used to show that the hexameric PduA shell protein of the 1,2-propanediol utilization (Pdu) microcompartment forms a selectively permeable pore tailored for the influx of 1,2-propanediol (the substrate of the Pdu microcompartment) while restricting the efflux of propionaldehyde, a toxic intermediate of 1,2-propanediol catabolism. Crystal structures of various PduA mutants provide a foundation for interpreting the observed biochemical and phenotypic data in terms of molecular diffusion across the shell. Overall, these studies provide a basis for understanding a class of selectively permeable channels formed by nonmembrane proteins.

microcompartment | protein channel | carboxysome | Salmonella | B_{12}

The complex behavior of biological systems depends fundamentally on the controlled movement of molecules between cellular compartments. Such processes occur in a wide range of biological contexts through the movement of ions and small molecules across lipid bilayers via proteins—channels and pumps—embedded in the bilayer. Achievements in understanding molecular transport in transmembrane systems have contributed to scientific disciplines from cell biology and physiology to membrane biophysics (1, 2). Interestingly, there exists a second type of system for molecular transport through proteins that is fundamentally different and much less understood. Hundreds of species of bacteria produce large subcellular organelles known as microcompartments (MCPs), which consist of metabolic enzymes encapsulated within proteinaceous shells reminiscent of viral capsids (reviewed in ref. 3). For MCPs to function, substrates and products must move across their outer protein shell, which lacks any lipid-based membrane. In the last several years, 3D structures of the proteins that comprise MCP shells have revealed narrow pores through their centers that have been hypothesized to be the routes by which substrates enter (and products escape from) MCPs (4; reviewed in ref. 5). However, experimental evidence to support this key hypothesis and the molecular principles involved is lacking.

The overarching function of MCPs is to optimize metabolic pathways that have toxic or volatile intermediates. MCPs are present across at least 11 different bacterial phyla, where they carry out diverse metabolic processes (6–12). The carboxysome MCP is used to enhance CO_2 fixation in nearly all bacteria that use the Calvin cycle, and it has been estimated that 25% of the carbon fixation on Earth occurs within this proteinaceous bacterial organelle (9). The 1,2-propanediol utilization (Pdu) and ethanolamine utilization (Eut) MCPs are used to optimize 1,2-propanediol (1,2-PD) and ethanolamine catabolism, respectively (13–15), and the degradation of these compounds is thought to promote enteric pathogenesis (12, 16, 17). Although the Pdu and Eut MCPs, the carboxysome, and other metabolically diverse MCPs of unknown function encapsulate distinct sets of enzymes, all have shells built from homologous proteins suggesting they operate by conserved functional principles. Most models of MCP function propose that the protein shell acts as a diffusion barrier that allows passage of substrates (and products) while limiting the escape of a toxic or volatile metabolic intermediate such as CO_2 or toxic aldehyde (9, 18), but selective permeability by MCP shells has not been established experimentally.

The shells of MCPs are assembled primarily from a family of small proteins that have so-called bacterial microcompartment (BMC) domains (5). Many BMC domain proteins form flat, hexagonally shaped oligomers that tile into extended sheets that form the basis of the MCP shell (4, 19, 20) (Fig. 1). In most cases, MCP shells are composed of four to eight different types of functionally diversified BMC domain proteins, some of which have pores proposed to mediate the selective movement of metabolites across the shell (4, 7, 8, 18). For example, the PduA shell protein from the Pdu MCP has a small central pore (~6 Å) that is lined with numerous hydrogen-bond donors and acceptors, leading to a suggested role in the preferential movement of 1,2-PD over the less polar propionaldehyde (a toxic intermediate) (21). In

Significance

Here, we describe a type of selective channel formed by a nonmembrane protein that is used to control metabolite movement between cellular compartments. In this case, the channel controls the movement of metabolites between the bacterial cytoplasm and the lumen of a bacterial microcompartment, a primitive organelle that is bounded by a protein shell and lacks any lipid bilayer. These studies are the first to our knowledge to prove selective molecular transport through a protein-based barrier. These findings broaden our understanding of how cells control the movement of molecules between cellular compartments, which is fundamental to biological systems. It also solves a key question about the function of bacterial microcompartments, which themselves impact many diverse and important biological processes.

Author contributions: C.C., S.C., A.P., T.O.Y., and T.A.B. designed research; C.C., S.C., A.P., M.R.S., and S.S. performed research; C.C., S.C., A.P., M.R.S., S.S., T.O.Y., and T.A.B. analyzed data; and C.C., S.C., T.O.Y., and T.A.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, www.pdb.org [PDB ID codes 4RBT (PduA-S40L), 4RBV (PduA-S40GSG), 4QIF (PduA-S40H), 4QIG (PduA-S40C), and 4RBU (PduA-S40Q)].

[1]C.C. and S.C. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: bobik@iastate.edu or yeates@mbi.ucla.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1423672112/-/DCSupplemental.

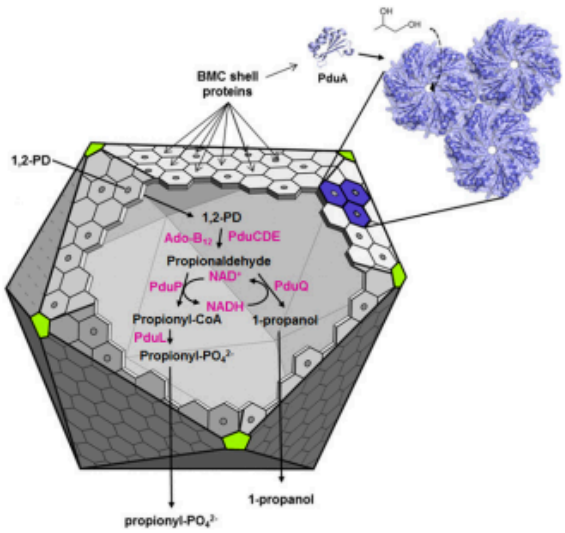


Fig. 1. Structure and function of the propanediol utilization (Pdu) bacterial microcompartment. A few thousand shell proteins (mostly of the BMC family) encapsulate a series of enzymes for metabolizing 1,2-PD. The protein shell of the Pdu MCP has been hypothesized to be selectively permeable allowing substrates such as 1,2-PD to enter through small pores in the center of hexameric shell proteins while restricting the efflux of propionaldehyde, which is toxic to the cell. For clarity, the reaction scheme has been simplified by omitting some of the steps involved in coenzyme B_{12} recycling.

addition, a subgroup of BMC proteins have been crystallized in two distinct conformations where the central pore is either fully closed or opened widely (12–15 Å), suggesting that a gating mechanism might control the movement of larger molecules (such as enzymatic cofactors) across the MCP shell (22, 23). However, no physiological or biochemical studies demonstrating transport or selective movement specifically through any MCP pore have been reported. As a result, the idea that the MCP protein shell is capable of mediating selective diffusion has lacked a clear experimental basis. Furthermore, a recent alternative model for MCP function proposes that enzymes embedded in or tightly associated with the shell could move metabolites into MCPs by vectorial catalysis, in which case functional pores might not be required for metabolite movement (24).

Here, we use the known structure of PduA—a canonical, hexameric BMC-type shell protein in the Pdu MCP—to design a series of mutant shell proteins having central pores with altered sizes and physicochemical properties. A combination of physiological studies on mutant bacteria, biochemical studies on isolated mutant MCPs, and crystal structure studies on the mutant shell proteins, show that the PduA pore serves as a key route for entry of the metabolic substrate (1,2-PD), and that the chemical properties of the PduA pore are tuned to limit the escape of the toxic propionaldehyde intermediate.

Results

PduA Pore Mutants and Their Structures. The PduA protein is a major component of the shell of the Pdu MCP, whose function is diagrammed in Fig. 1 (25). To investigate the role of the pore in molecular transport, we made eight PduA pore variants in *Salmonella enterica* serovar Typhimurium LT2 via chromosomal mutations: PduA-S40T, PduA-S40Q, PduA-S40C, PduA-S40L, PduA-S40A, PduA-S40H, and PduA-S40M, and an insertion mutation (PduA-S40GSG) that was intended to occlude the pore by the addition of two glycine residues (one on each side of S40).

For all eight PduA mutants, electron microscopy and SDS/PAGE showed that the mutated shell protein assembled into normal-appearing MCPs with protein compositions similar to wild-type, indicating that these mutations did not cause major structural changes (*SI Appendix*, Figs. S1 and S2).

To characterize the detailed structural effects of the pore mutations, X-ray crystallographic studies were undertaken on the individual shell protein mutants. Crystal structures of five of the PduA mutants were successfully determined (PduA-S40L, PduA-S40H, PduA-S40C, PduA-S40Q, and PduA-S40GSG) in the background context of a K26A edge mutation that mitigates crystallographic problems that often arise from the side-by-side aggregation of hexameric units (*SI Appendix*, Table S1) (26). In all five crystal structures elucidated, the protein assembled into homohexamers (despite the potentially disrupting effects of having six closely spaced changes in the native pore region) and revealed differences relative to the wild-type PduA structure, mainly in the pore region as intended (Fig. 2 and *SI Appendix*,

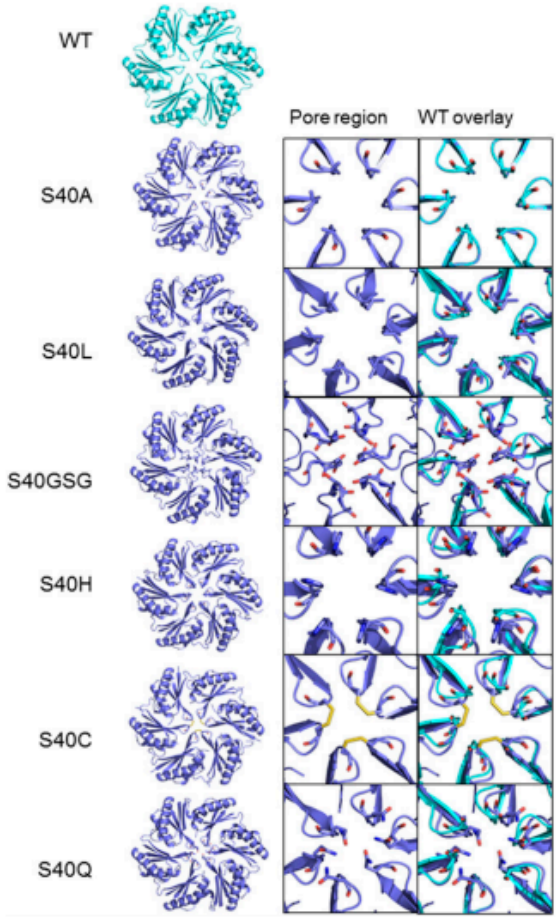


Fig. 2. Three-dimensional crystal structures of PduA and pore mutants. PduA hexamer assemblies are shown in cartoon format (*Left*) with close-up view of the pore region (*Middle*) and overlaid with wild-type (*Right*). From top to bottom: Wild-type (WT) PduA (PDB ID code: 3NGK, cyan), PduA-S40A (model, marine blue), PduA-S40L (PDB ID code: 4RBT), PduA-S40GSG (PDB ID code: 4RBV), PduA-S40H (PDB ID code: 4QIF), PduA-S40C (PDB ID code: 4QIG), and PduA-S40Q (PDB ID code: 4RBU).

Table 1. Pore diameter and hydrophobicity

PduA mutant	Pore diameter, Å	Pore segment	Solvation free energy difference in the pore, kcal·mol^{-1}
Wild-type	5.6	I38-L42	—
PduA-S40A	5.9	I38-L42	−2.1 ± 0.5
PduA-S40L	5.5	I38-L42	4.7 ± 0.7
PduA-S40GSG	1.9	I38-L44	−1.3 ± 0.9
PduA-S40Q	0.5	I38-L42	−0.2 ± 0.1
PduA-S40H	4.3	I38-L42	1.5 ± 0.2

In cases where multiple instances of a hexamer were present in a crystal structure, the reported pore diameters and solvation-free energies are average values.

Table S2). The pore of PduA-S40L is considerably more hydrophobic, but not more occluded than the wild type. The S40H mutant has a pore whose natural sixfold symmetry is broken, but which is likewise not occluded. The S40C mutant shows polymorphic behavior; three distinct instances of the hexamer occur in the crystal form, and the pore can be open or occluded depending on the cysteine side chain orientations and whether disulfide bonds are formed between them. The S40GSG insertion mutation has a fully occluded pore (which was the intended outcome) in two of three distinct hexamers of the crystal structure. The S40Q mutant also has an occluded pore in both distinct hexamers of the crystal structure (*SI Appendix*, Table S2). Three other mutants (S40A, S40M, and S40T) formed essentially normal-appearing MCPs but were not characterized by crystallography. Models of those mutants were generated by conservative computer modeling. In particular, the S40A mutant, whose physiological behavior was notable, was modeled simply by removing the serine hydroxyl group. The PduA-S40A model has a central pore that is not significantly larger than wild type (despite loss of the hydroxyl groups), but which is less polar (Table 1 and *SI Appendix*, Table S2 and Fig. S6).

Effects of PduA Pore Mutations on Shell Function. Prior studies showed that mutational impairment of the shell of the Pdu MCP results in faster growth of *Salmonella* on 1,2-PD at limiting B_{12} concentrations, whereas shell defects lead to propionaldehyde toxicity at saturating B_{12} (27, 28). When B_{12} is saturating, diol dehydratase rapidly converts 1,2-PD to propionaldehyde, which leaks from defective MCPs causing toxicity and growth inhibition (27–29). At limiting B_{12}, a broken or damaged shell results in faster growth due to the increased availability of enzyme substrates and cofactors to the 1,2-PD degradative enzymes encased within the MCP (27, 29). No aldehyde toxicity occurs at limiting B_{12} due to the lower rate of propionaldehyde formation by diol dehydratase. Hence, growth rate at saturating B_{12} provides a means to measure the extent of propionaldehyde leakage from the MCP, whereas growth rate at limiting B_{12} allows an estimate of the permeability of the shell to substrates (27, 29).

In our initial tests, we looked at the effects of the eight PduA pore mutants (described above) on shell permeability by measuring growth of *Salmonella* on 1,2-PD at limiting B_{12}. PduA-S40T, PduA-S40A, PduA-S40H, and PduA-S40M mutants grew similarly to wild type (*SI Appendix*, Fig. S3), indicating no significant change in shell permeability. In contrast, the PduA-S40L, PduA-S40C, PduA-S40Q, and PduA-S40GSG mutants grew more slowly than wild type on 1,2-PD (*SI Appendix*, Fig. S3). This finding suggested that the PduA-S40L, PduA-S40C, PduA-S40Q, and PduA-S40GSG pore mutations resulted in a shell that is less permeable to substrates. Notably, three of these slow-growing mutants (PduA-S40GSG, PduA-S40Q, and PduA-S40C) had occluded pores in the crystal structures. Moreover, additional tests with the PduA-S40L, PduA-S40GSG, PduA-S40C, and PduA-S40Q mutants showed that increasing the 1,2-PD concentration from 0.4 to 2.4% corrected their slow-growth phenotype, consistent with an impaired diffusion of 1,2-PD into the MCP that could be offset by a much higher concentration gradient (*SI Appendix*, Fig. S4).

To more directly test whether PduA pore mutations impaired the movement of 1,2-PD across the shell of the Pdu MCP, we measured the coenzyme B_{12}-dependent diol dehydratase (DDH) activity in purified MCPs. DDH is an MCP lumen enzyme that catalyzes the first step of 1,2-PD degradation (the conversion of 1,2-PD to propionaldehyde); hence, its activity depends on the diffusion of 1,2-PD across the MCP shell (13). MCPs purified from PduA-S40L, PduA-S40GSG, and PduA-S40Q mutants had ∼50%, 62%, and 75% DDH activity compared with the wild-type level, whereas MCPs from PduA-S40T, PduA-S40C, PduA-S40A, PduA-S40H, and PduA-S40M mutants had activities similar to wild type (Table 2). The markedly reduced DDH activity of the PduA-S40L, PduA-S40Q, and PduA-S40GSG mutants are most simply explained by altered structural features that block the influx of 1,2-PD and/or that impair the egress of propionaldehyde (which might lead to product inhibition). The former explanation is most consistent with the experiments noted above in which increased 1,2-PD concentration mitigated the growth defect.

It was notable that overall the DDH activities of MCPs purified from the PduA pore mutants correlated well with the growth rates of those mutants on 1,2-PD, providing both in vivo and in vitro evidence that certain PduA pore variants limit substrate influx. The PduA-S40C mutant, however, provided one interesting exception. It showed normal DDH activity in vitro for purified MCPs, although it showed a reduced growth rate on 1,2-PD (Table 2 and *SI Appendix*, Fig. S3). To test whether the in vitro activity of this mutant might be lowered by environmental conditions likely to occlude the pore, we added iron ($FeCl_2$) under reducing conditions (adding DTT) with the expectation that the multiple proximal cysteine residues might bind iron at the center of the pore. Indeed, under these conditions DDH activity for the S40C mutant dropped to 66% (similar treatment lowered the activity in the wild type case to 87%) (*SI Appendix*, Table S3). This finding provided further evidence that occluding the pore of the PduA shell protein limits flux into the

Table 2. DDH activity of purified mutant MCPs

Type of purified MCPs	Specific activity, μmol·min^{-1}·mg^{-1}
Wild-type (LT2)	28.4 ± 0.6
PduA-S40L	14.8 ± 0.8
PduA-S40GSG	17.5 ± 0.5
PduA-S40Q	21.2 ± 0.7
PduA-S40H	28.5 ± 0.4
PduA-S40T	29.4 ± 0.4
PduA-S40M	29.2 ± 0.2
PduA-S40N	27.8 ± 0.8
PduA-S40A	26.1 ± 0.8
PduA-S40C	27.1 ± 0.2

MCP and also suggested that the inconsistency between growth test and DDH assays (noted above) might be due to differences between in vivo and in vitro conditions.

Control experiments were performed to rule out ancillary effects of pore mutations. If the PduA-S40L and PduA-S40GSG mutants reduced DDH activity by impeding the movement of 1,2-PD across the MCP shell (rather than by some other unanticipated effect), then breaking the MCP should return the DDH activity to wild-type levels. The measured activities of wild type, PduA-S40L, and PduA-S40GSG MCPs following disruption using standard protocols were all very similar (*SI Appendix*, Table S4).

Movement of Propionaldehyde Through the PduA Pore. The hypothesized role of the Pdu MCP shell is to retain propionaldehyde while allowing 1,2-PD entry. We therefore tested the effects of selected PduA pore mutants on the efflux of propionaldehyde from the Pdu MCP by measuring the amount of propionaldehyde that diffused out of the MCP, through the bacterial cell membrane, and into the culture medium during growth of *Salmonella* on 1,2-PD as described (28) (Fig. 3). Simultaneously, we measured growth rate and 1,2-PD consumption. The PduA-S40L and PduA-S40GSG mutants (whose pores were shown to diminish 1,2-PD uptake) consumed 1,2-PD more slowly than wild

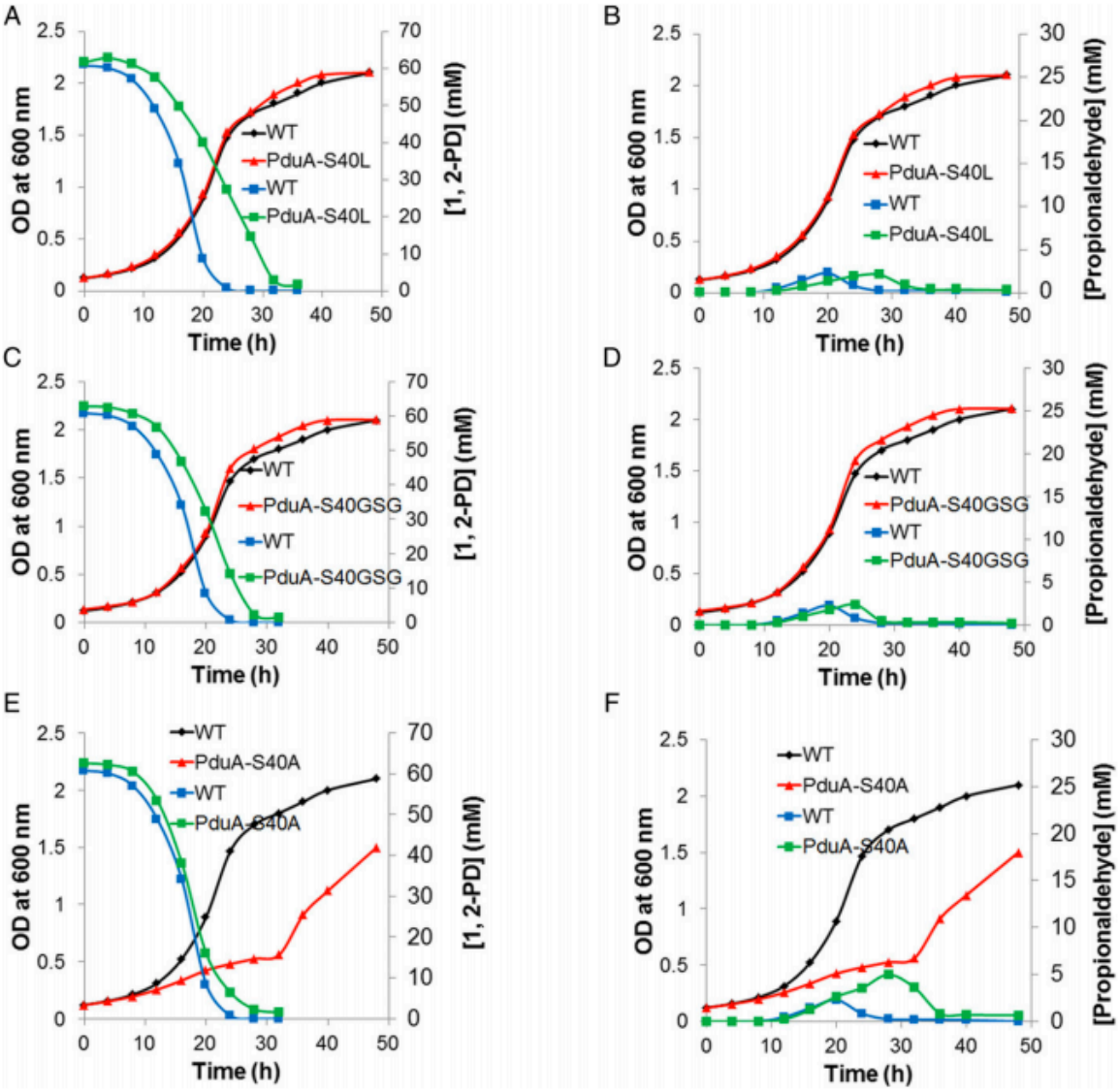


Fig. 3. Consumption of 1,2-PD and propionaldehyde release by select PduA pore mutants. 1,2-PD consumption and propionaldehyde release was monitored during growth of wild-type and various PduA pore mutants on 1,2-PD minimal medium. Wild-type *Salmonella* (WT) compared with a PduA-S40L mutant (*A* and *B*), a PduA-S40GSG mutant (*C* and *D*), and a PduA-S40A mutant (*E* and *F*). Left column, red and black lines are OD_{600} and green and blue lines are 1,2-PD. Right column, red and black lines are OD_{600} and green and blue lines are propionaldehyde.

type, but excreted propionaldehyde in similar amounts. Most notably, the PduA-S40A mutant consumed 1,2-PD at a rate similar to wild type, but excreted approximately 2.5-fold more propionaldehyde. Increased excretion led to a marked inhibition of growth (from 12 to 30 h) as a result of propionaldehyde toxicity as was previously observed for deletion mutations that disrupt the MCP shell (27) (Fig. 3). These results indicated that the wild-type PduA pore provides a key route for inward diffusion of 1,2-PD, and that its detailed structural features are important for restricting the egress of propionaldehyde to mitigate aldehyde toxicity.

A second test used to assess the permeability of PduA pore variants to propionaldehyde was to grow *Salmonella* on 1-propanol under conditions where growth was limited by diffusion of propionaldehyde, produced by oxidation of 1-propanol in the cytoplasm, into the Pdu MCP (*SI Appendix, SI Materials and Methods*). These tests indicated that the relative rate of propionaldehyde movement into the Pdu MCP was PduAS40A > PduS40L > wild type ~ PduAS40T (*SI Appendix*, Fig. S5). The modest increase in the permeability of S40L compared with wild type for propionaldehyde influx (*SI Appendix*, Fig. S5) can be reconciled with the unaffected rate of aldehyde efflux during growth on 1,2-PD (Fig. 3) by a reduced rate of internal aldehyde production (by diol dehydratase) in this mutant (Table 2). Most importantly, however, this finding was the second experiment that indicated the S40A mutant has an increased permeability to propionaldehyde compared with wild type.

Effects of Polarity of the PduA Pore on Small Molecule Diffusion. Some of the observed data on transport through mutant PduA pores (summarized in *SI Appendix*, Table S6) are easily interpreted, such as the blockage of 1,2-PD influx in the fully occluded S40GSG mutant and in the pore-constricted S40Q mutant, whereas the behavior in other cases calls for a more detailed analysis. Two observations in particular are worth examining. First, the S40L mutant did not produce a pore with a narrower diameter, yet influx of 1,2-PD was diminished. Second, the S40A mutation did not significantly increase the minimal radius of the pore, yet propionaldehyde efflux was enhanced. We examined the physicochemical properties of the mutant pores to shed light on these phenomena. We evaluated the polarity of the pore in the different structures according to a calculated estimate of the difference in solvation free energy of the pore (30). The trend in solvation free energies sheds light on the transport of small molecules through the pore. Not unexpectedly, the S40L pore is much more hydrophobic (i.e., higher solvation energy) than the wild type (Table 1 and *SI Appendix*, Fig. S6). In addition, the longer leucine side chain produces a pore with an extended length compared with the wild-type structure, wherein the pore narrows at a more focused point. Although the pore is not occluded, the hydrophobic nature of the S40L pore likely creates an energetic barrier for the 1,2-PD substrate. The 1,2-PD is presumably strongly hydrated and hydrogen bonded to water in the bulk phase, and would need to be desolvated during its passage along the length of the hydrophobic pore in the S40L mutant. In the case of the S40A mutant, removal of the hydroxyl group (and its hydrogen bonding capacity) did not hinder 1,2-PD influx under the conditions tested. It did, however, substantially enhance escape of the propionaldehyde intermediate. This finding suggests that the polar character, and perhaps the hydrogen bonding capacity, of the native serine hydroxyl at residue 40 creates a relatively lower energy barrier for passage of 1,2-PD compared with propionaldehyde, which is less polar than 1,2-PD. The substantial effects on transport properties that are caused by subtle changes in the structure of the PduA pore attest to its finely tuned role in the Pdu MCP.

The Movement of Large Cofactors into Purified Mutant and Wild-Type MCPs. Prior studies indicated that PduP, which catalyzes the conversion of propionaldehyde + HS-CoA + NAD^+ → propionyl-CoA and NADH, localizes to the lumen of the Pdu MCP (31). Therefore, measurement of PduP activity provides a way to assess the movement of the large cofactors, HS-CoA and NAD^+, across the MCP shell. Enzyme assays on purified MCPs (both broken and intact) showed that wild-type and mutant MCPs all had similar levels of PduP activity (approximately 2.8 $\mu mol \cdot min^{-1} \cdot mg^{-1}$) when assayed under conditions identical to those used to show that the PduA-S40L and PduA-S40GSG mutations restricted the influx of 1,2-PD (*SI Appendix*, Table S4). This finding suggests that large cofactors can enter mutant Pdu MCPs capable of restricting the movement of 1,2-PD and propionaldehyde. The absence of effect with mutants in PduA supports earlier hypotheses that large cofactors cross the MCP through other specialized shell proteins that have gated pores that open and close when triggered for specific transport (22, 23, 32).

Selectivity of the PduA Pore for 1,2-PD and Glycerol. A multiple sequence alignment showed that the serine at position 40 of PduA is widely conserved among diverse genera of bacteria except for the *Lactobacilli*, where all species examined had a histidine at that position (*SI Appendix*, Fig. S7). Interestingly, previous studies showed that *Lactobacillus* uses the Pdu MCP for the catabolism of both 1,2-PD and glycerol (33, 34). In contrast, *Salmonella* uses the Pdu MCP for 1,2-PD degradation but catabolizes glycerol by alternative pathways. To test whether the H40 amino acid of the *Lactobacillus* PduA protein might be an adaptation that allows movement of both glycerol and 1,2-PD into the *Lactobacillus* Pdu MCP, we constructed a chromosomal PduA-S40H mutation in *Salmonella*, and measured its effects on the influx of 1,2-PD and glycerol into the Pdu MCP. The PduA-S40H mutation had no significant effect on growth of *Salmonella* on 1,2-PD or on the DDH activity of purified MCPs (*SI Appendix*, Fig. S3 and Table 2), indicating that H40 allowed normal entry of 1,2-PD into the Pdu MCP under the conditions used. To examine molecular transport in the PduA S40H mutant in vitro, we looked at glycerol inhibition of DDH in purified Pdu MCPs; glycerol is both a substrate and an inhibitor of DDH (35). The DDH activity of purified PduA-S40H MCPs was inhibited by glycerol at a five times lower concentration compared with wild-type MCPs, indicating substantially greater permeability toward glycerol (Table 3). Conversely, DDH inhibition in the PduA-S40L and PduA-S40GSG mutants required five times higher concentrations of glycerol. This finding indicated a change in the relative permeability of the pore to 1,2-PD and glycerol, which compete for the DDH active site. We also found that purified recombinant DDH (*SI Appendix*, Fig. S2) was particularly sensitive to glycerol inhibition, further indicating the selectivity of the MCP shell in general (Table 3). Collectively, these results emphasize the role of the MCP shell and the PduA in dictating transport selectivity in the Pdu MCP.

Discussion

A fundamental question about bacterial MCPs is how their protein shells restrict the efflux of volatile/toxic pathway intermediates while allowing enzyme substrates, cofactors, and products to pass. In this report, we show that the PduA protein forms selectively permeable pores across the shell of the Pdu MCP that allow the retention of the propionaldehyde intermediate while

Table 3. DDH inhibition by glycerol in MCPs purified from different pore mutants

Type of purified MCP	IC_{50} of glycerol, mM
Wild-type (LT2)	291.7 ± 2.2
PduA-S40L	1,560 ± 50.1
PduA-S40GSG	1,410 ± 35.9
PduA-S40H	55.3 ± 1.7
Purified DDH	1.2 ± 0.1

allowing influx of the 1,2-PD substrate. These studies demonstrate that the pores of MCP shell proteins are selectively permeable to small molecules. In the work presented here, crystal structures of several of the PduA pore mutants provided a framework for interpreting the phenotypic data in terms of physical properties of the pore. Mutants unable to take up the 1,2-PD substrate could be correlated with pores that were fully occluded or substantially more hydrophobic. The behavior and structure of the S40A mutant shed light on the importance of the detailed properties of the pore for aldehyde retention. In this case, the removal of (six copies of) a hydroxyl group in the pore does not have a significant effect on 1,2-PD entrance, but propionaldehyde escape is allowed. The hydroxyl groups at this constricted point in the pore provide a highly polar environment that evidently disfavors occupancy by propionaldehyde (which is less polar than 1,2-PD). The S40H mutant showed altered transport specificity, allowing influx of glycerol, consistent with the metabolic capabilities of bacterial species that have this particular amino acid at the key residue 40 in their PduA protein sequence. Together, these results provide substantial evidence that MCP shell proteins have evolved to meet specific transport roles and that a selectively permeable protein shell is important to the function of diverse MCPs.

In summary, these studies highlight the existence of a family of proteins—entirely distinct from transmembrane protein channels and pumps—that have evolved diverse selective transport functions in bacteria. The distinct features of these protein-based transport systems, particularly the potential advantage of their nonlipidic environments, should motivate vigorous investigations of their mechanisms. Important future goals include understanding the dynamic mechanisms that underlie molecular transport in MCP shell proteins, and intelligently modifying MCP proteins to create selectively permeable molecular chambers for synthetic biology applications (36).

Methods and Materials

MCP purification, enzyme assays, and growth studies were performed as described (27, 31, 37). Site-directed mutants were made by using standard protocols (26). Crystals were prepared by the hanging drop method. Some minor modifications to prior protocols are described in *SI Appendix, SI Methods and Materials*.

ACKNOWLEDGMENTS. We thank the Iowa State University nanoimaging and DNA sequencing and synthesis facilities for help with the electron microscopy and DNA sequencing, Michael Collazo and Duilio Cascio for assistance in protein crystallization and data collection, and the synchrotron staff at Advanced Photon Source (APS) Beamline 24-ID-C. We also thank M. Capel, K. Rajashankar, N. Sukumar, J. Schuermann, I. Kourinov, and F. Murphy (Northeastern Collaborative Access Team Beamline 24-ID-C at APS, which is supported by National Center for Research Resources Grant 5P41RR015301-10 and National Institute of General Medical Sciences Grant 8 P41 GM103403-10 from NIH). This work was supported by NIH Grant R01AI081146 (to T.O.Y. and T.A.B.). Work in the T.O.Y. laboratory was supported by the Biological and Environmental Research program of the Department of Energy (DOE) Office of Science. S.C. was supported by a Chemistry-Biology Interface NIH Training Grant T32-GM008496 and by the University of California, Los Angeles Graduate Division. Use of the APS is supported by the DOE under Contract DE-AC02-06CH11357.

1. Doyle DA, et al. (1998) The structure of the potassium channel: Molecular basis of K^+ conduction and selectivity. *Science* 280(5360):69–77.
2. Luckey M (2014) *Membrane Structural Biology: With Biochemical and Biophysical Foundations* (Cambridge Univ Press, Cambridge), 2nd Ed, pp 284–357.
3. Chowdhury C, Sinha S, Chun S, Yeates TO, Bobik TA (2014) Diverse bacterial microcompartment organelles. *Microbiol Mol Biol Rev* 78(3):438–468.
4. Kerfeld CA, et al. (2005) Protein structures forming the shell of primitive bacterial organelles. *Science* 309(5736):936–938.
5. Yeates TO, Jorda J, Bobik TA (2013) The shells of BMC-type microcompartment organelles in bacteria. *J Mol Microbiol Biotechnol* 23(4-5):290–299.
6. Bobik TA (2006) Polyhedral organelles compartmenting bacterial metabolic processes. *Appl Microbiol Biotechnol* 70(5):517–525.
7. Abdul-Rahman F, Petit E, Blanchard JL (2013) The distribution of polyhedral bacterial microcompartments suggests frequent horizontal transfer and operon reassembly. *J Phylogen Evolution Biol* 1:118.
8. Jorda J, Lopez D, Wheatley NM, Yeates TO (2013) Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci* 22(2):179–195.
9. Rae BD, Long BM, Badger MR, Price GD (2013) Functions, compositions, and evolution of the two types of carboxysomes: Polyhedral microcompartments that facilitate CO_2 fixation in cyanobacteria and some proteobacteria. *Microbiol Mol Biol Rev* 77(3):357–379.
10. Yeates TO, Kerfeld CA, Heinhorst S, Cannon GC, Shively JM (2008) Protein-based organelles in bacteria: Carboxysomes and related microcompartments. *Nat Rev Microbiol* 6(9):681–691.
11. Kerfeld CA, Heinhorst S, Cannon GC (2010) Bacterial microcompartments. *Annu Rev Microbiol* 64:391–408.
12. Cheng S, Liu Y, Crowley CS, Yeates TO, Bobik TA (2008) Bacterial microcompartments: Their properties and paradoxes. *BioEssays* 30(11-12):1084–1095.
13. Bobik TA, Havemann GD, Busch RJ, Williams DS, Aldrich HC (1999) The propanediol utilization (pdu) operon of Salmonella enterica serovar Typhimurium LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B_{12}-dependent 1,2-propanediol degradation. *J Bacteriol* 181(19):5967–5975.
14. Penrod JT, Roth JR (2006) Conserving a volatile metabolite: A role for carboxysome-like organelles in Salmonella enterica. *J Bacteriol* 188(8):2865–2874.
15. Brinsmade SR, Paldon T, Escalante-Semerena JC (2005) Minimal functions and physiological conditions required for growth of Salmonella enterica on ethanolamine in the absence of the metabolosome. *J Bacteriol* 187(23):8039–8046.
16. Thiennimitr P, et al. (2011) Intestinal inflammation allows Salmonella to use ethanolamine to compete with the microbiota. *Proc Natl Acad Sci USA* 108(42):17480–17485.
17. Garsin DA (2010) Ethanolamine utilization in bacterial pathogens: Roles and regulation. *Nat Rev Microbiol* 8(4):290–295.
18. Yeates TO, Thompson MC, Bobik TA (2011) The protein shells of bacterial microcompartment organelles. *Curr Opin Struct Biol* 21(2):223–231.
19. Tanaka S, et al. (2008) Atomic-level models of the bacterial carboxysome shell. *Science* 319(5866):1083–1086.
20. Tanaka S, Sawaya MR, Phillips M, Yeates TO (2009) Insights from multiple structures of the shell proteins from the beta-carboxysome. *Protein Sci* 18(1):108–120.
21. Crowley CS, et al. (2010) Structural insight into the mechanisms of transport across the Salmonella enterica Pdu microcompartment shell. *J Biol Chem* 285(48):37838–37846.
22. Klein MG, et al. (2009) Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. *J Mol Biol* 392(2):319–333.
23. Tanaka S, Sawaya MR, Yeates TO (2010) Structure and mechanisms of a protein-based organelle in Escherichia coli. *Science* 327(5961):81–84.
24. Huseby DL, Roth JR (2013) Evidence that a metabolic microcompartment contains and recycles private cofactor pools. *J Bacteriol* 195(12):2864–2879.
25. Havemann GD, Bobik TA (2003) Protein content of polyhedral organelles involved in coenzyme B_{12}-dependent degradation of 1,2-propanediol in Salmonella enterica serovar Typhimurium LT2. *J Bacteriol* 185(17):5086–5095.
26. Sinha S, et al. (2014) Alanine scanning mutagenesis identifies an asparagine-arginine-lysine triad essential to assembly of the shell of the Pdu microcompartment. *J Mol Biol* 426(12):2328–2345.
27. Cheng S, Sinha S, Fan C, Liu Y, Bobik TA (2011) Genetic analysis of the protein shell of the microcompartments involved in coenzyme B_{12}-dependent 1,2-propanediol degradation by Salmonella. *J Bacteriol* 193(6):1385–1392.
28. Havemann GD, Sampson EM, Bobik TA (2002) PduA is a shell protein of polyhedral organelles involved in coenzyme B_{12}-dependent degradation of 1,2-propanediol in Salmonella enterica serovar Typhimurium LT2. *J Bacteriol* 184(5):1253–1261.
29. Sampson EM, Bobik TA (2008) Microcompartments for B_{12}-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. *J Bacteriol* 190(8):2966–2971.
30. Eisenberg D, McLachlan AD (1986) Solvation energy in protein folding and binding. *Nature* 319(6050):199–203.
31. Leal NA, Havemann GD, Bobik TA (2003) PduP is a coenzyme-A-acylating propionaldehyde dehydrogenase associated with the polyhedral bodies involved in B_{12}-dependent 1,2-propanediol degradation by Salmonella enterica serovar Typhimurium LT2. *Arch Microbiol* 180(5):353–361.
32. 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(22):6056–6063.
33. Sriramulu DD, et al. (2008) Lactobacillus reuteri DSM 20016 produces cobalamin-dependent diol dehydratase in metabolosomes and metabolizes 1,2-propanediol by disproportionation. *J Bacteriol* 190(13):4559–4567.
34. Talarico TL, Axelsson LT, Novotny J, Fiuzat M, Dobrogosz WJ (1990) Utilization of glycerol as a hydrogen acceptor by Lactobacillus reuteri: Purification of 1,3-propanediol:NAD oxidoreductase. *Appl Environ Microbiol* 56(4):943–948.
35. Toraya T (2000) Radical catalysis of B_{12} enzymes: Structure, mechanism, inactivation, and reactivation of diol and glycerol dehydratases. *Cell Mol Life Sci* 57(1):106–127.
36. Cai F, Sutter M, Bernstein SL, Kinney JN, Kerfeld CA (August 12, 2014) Engineering bacterial microcompartment shells: Chimeric shell proteins and chimeric carboxysome shells. *ACS Synth Biol*, 10.1021/sb500226.
37. Sinha S, Cheng S, Fan C, Bobik TA (2012) The PduM protein is a structural component of the microcompartments involved in coenzyme B_{12}-dependent 1,2-propanediol degradation by Salmonella enterica. *J Bacteriol* 194(8):1912–1918.

Supplementary information

Materials and Methods

Chemicals and Reagents

Antibiotics, vitamin B_{12} (CN-B_{12}, AdoB_{12}), NAD^+, NADH, $NADP^+$, and NADPH were from Sigma Aldrich (St. Louis, MO). *KOD* DNA polymerase and restriction enzymes and T4 ligase were respectively from Novagen (Cambridge, MA) and New England Biolabs (Beverly, MA). Bacterial protein extraction reagent (B-PER II) was purchased from Pierce (Rockford, IL). Other chemicals were from Fisher Scientific (Pittsburgh, PA).

Bacterial strains, media, and growth conditions

The bacterial strains used are listed in Table S5. All strains are derivatives of *Salmonella enterica* serovar Typhimurium strain LT2. The rich medium used was Luria-Bertani/Lennox (LB) medium (Difco, Detroit, MI) (1). TYE media was used was prepared with bacto-tryptone, yeast extract and agar were from Difco Laboratories (Difco, Detroit, MI). The minimal medium used was no carbon-E (NCE) medium supplemented with 1 mM $MgSO_4$, 0.3 mM each of valine, isoleucine, leucine, and threonine and of ferric citrate (50 μM), as described previously (2, 3).

Three dimensional model building and visualization

Initial selection of amino acid residues that might obstruct the PduA pore was based on three dimensional models built with the Swiss Model server (4) and all the visualization and graphical presentations were carried out in Pymol. PduA-S40A was modeled on PduA (PDBID: 3NGK) using the Coot software.

Construction of chromosomal mutations

Scarless chromosomal *pduA* deletion mutations as well as missense mutations surrounding their pore regions were made by recombineering as described earlier (5, 6). All mutations were confirmed by DNA sequencing.

Growth studies

Growth studies were performed as previously described using a Synergy HT Microplate reader (BioTek, Winooski, VT) (3).

MCP purification

Pdu microcompartment was isolated according to the prior protocols with minor modifications (7). Cells were harvested by centrifugation and lysed in buffer A (50 mM Tris-HCl, 500 mM KCl, 12.5 mM $MgCl_2$, 1.5% 1,2-PD, pH 8.0) supplemented with lysozyme (1 mg ml^{-1}) and BPERII (60% v/v) (6). After removing cellular debris MCPs were pelleted by centrifugation at 20, 000 x g for 30 min. The translucent MCPs were resuspended in buffer B (50 mM Tris-HCl pH 8.0, 50 mM KCl, 5 mM $MgCl_2$, 1% 1,2-PD) and kept at 4 ^oC for future use.

Electron microscopy

Transmission electron microscopy was performed as previously described (7). Immediately after purification, MCPs were diluted to a concentration of 0.1 mg ml^{-1} in buffer B and 10 µl of was applied to carbon coated copper grids (200 mesh). Buffer salts were removed after washing with ultrapure water and the excess liquid was wicked off with filter paper. The samples were fixed with 2% glutaraldehyde. Finally, MCPs were negatively stained with uranyl acetate (2%) and visualized under JEOL 2100 scanning transmission EM (JEOL USA, Inc., Peabody, MA) (7).

Diol dehydratase purification and assay

DDH was purified from crude membrane as described Tobimatsu et al. (8). DDH activity was measured as described earlier with minor modifications (9) The one ml reaction mixture contained 5 to 10 µg MCPs (or purified DDH), 200 mM 1, 2-PD, 50 mM HEPES buffer (pH 7.5), 50 mM KCl, 0.4 mM NADH and excess alcohol dehydrogenase (ADH) (*Saccharomyces cerevisiae*, Sigma-Aldrich, St Louis, MO). The reaction was started with 20 µM Ado-B_{12} and monitored at 340 nm at 37 °C. Reaction rates were quantified using $\Delta\varepsilon_{340} = 6.22\ mM^{-1}\ cm^{-1}$.

Glycerol inhibition of diol dehydratase activity

DDH activity assays were performed as described above except that 10 mM 1,2-PD was used. Reactions were performed in triplicate and the IC_{50} was calculated from a nonlinear curve fit, Log [Glycerol] vs Activity, using GraphPad Prism v 5.0 software (GraphPad Software, La Jolla, CA).

Propionaldehyde dehydrogenase activity

The PduP propionaldehyde dehydrogenase activity was measured using a modification of a prior method (10) The 1 ml reaction contained 50 mM HEPES, pH7.5, 50 mM KCl, 1 mM dithiothreitol, 75 µM NAD^+, 100 µM HS-CoA and 5 µg (or 10 µg) purified MCPs. The reaction was initiated with addition of 10 mM propionaldehyde and the conversion of NAD^+ to NADH was monitored at 340 nm at 37 °C.

Determination of propionaldehyde and 1,2-PD in culture media

An overnight LB culture was harvested, rinsed once and resuspended in NCE minimal medium with 1 mM $MgSO_4$. These cells were used to inoculate 50 ml of NCE minimal media (described above) supplemented with 0.4% 1, 2-PD and 150 nM CN-B_{12} to a final optical density 0.1 at 600 nm. These cultures were grown in 250 ml Erlenmeyer flasks at 37 °C with continuous shaking at 275 rpm (3), and sampled at timed intervals. Cells were removed by centrifugation and samples were filtered using 0.22 µm Millex- GV syringe filters (Millipore, Darmstadt, Germany). Propionaldehyde and 1,2-PD was then determined by high-performance liquid chromatography (HPLC) (11).

Growth on 1-propanol

An overnight culture was pelleted by centrifugation, rinsed, and resuspended in NCE minimal medium with no carbon source. This suspension was used to inoculate NCE minimal media supplemented with 1-propanol (0.1%, w/v) and 1, 2-PD (0.2%, w/w), to a final OD_{600} to 0.15. Growth was monitored in a Synergy HT Microplate reader (BioTek, Winooski, VT) as described before.

Cloning, Expression, and Purification of PduA mutants for crystallography

Site-directed mutagenesis (QuikChange, Stratagene, La Jolla, CA) was performed on full-length PduA (acquired from PCR amplification of chromosomal DNA from *S. enterica* serovar Typhimurium LT2 and inserted in a pET22b(+) (Novagen, Darmstadt, Germany) to generate the mutants of pore residue S40: S40L, S40Q, S40H, S40C and the insertion of an additional glycine on either side of the native sequence GS40G (referred to as S40GSG) (Table S2) in combination with the K26A mutation which was shown to increase the solubility of PduA. All constructs were confirmed by sequencing reactions (Laragene, GeneWiz). PduA mutant constructs were transformed in BL21 CodonPlus(DE3)-RIL *E. coli* cells (Stratagene), grown in Luria-Bertani (LB) media with ampicillin (100 μg ml^{-1}) and chloramphenicol (34 μg ml^{-1}) at 37°C, shaking at 225 rpm. When cell cultures reached an optical density at wavelength 600 nm (OD_{600}) of 0.4, temperature was dropped to 30 °C; and at OD_{600} of 0.6, protein expression was induced with 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) for 4 h. Cells were pelleted at 4,000 rpm for 15 min and resuspended (15 g cell pellet per 35 ml total) in 50 mM Tris-HCl pH 8.0 and 200 mM NaCl (Buffer A). Resuspended cells were stored at -80°C until purification. Cells were thawed and homogenized in the presence of lysozyme, protease inhibitor cocktail, DNase I and 1 mM $MgCl_2$. Cells were lysed by sonication in 3 cycles, increasing amplitude from 40-60% per cycle. Cell lysate was centrifuged at 17,000 x g for 30 min and the supernatant was clarified by 0.45μm filtration. The clarified supernatant was loaded on a HisTrap column (CV=5ml, GE Healthcare, Pittsburgh, PA). The column was

washed with 10% Buffer B (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 500 mM Imidazole) and eluted with a 9 CV linear gradient to 90% Buffer B. PduA constructs eluted between 30-50% Buffer B. Fractions containing the fairly pure and large quantities of protein migrating at 10 kDa were pooled and dialyzed into crystallization buffer (30 mM Tris-HCl pH 8.0, 50 mM NaCl, 1% v/v glycerol) overnight at 4 °C. PduA mutant constructs were then concentrated to 10-42 mg ml^{-1} each, flash frozen with liquid N_2, and stored at -80 °C.

Crystallization

96-well crystallization screening trials by hanging drop method were set up for PduA mutant constructs, including screens from Emerald Biosystems, Hampton Research, and Qiagen. PduA-K26A S40L (19.9 mg ml^{-1}) crystallized in hexagonal plate and trigonal pyramid crystals in multiple conditions. PduA-K26A S40L crystals that were used for structure determination were optimized from 0.1 M MES sodium salt pH 6.5, 2.0 M ammonium sulfate, and 5% w/v PEG 400. PduA-K26A S40Q (42.0 mg ml^{-1}) crystallized in hexagonal plate crystals in multiple conditions; crystals were optimized from 2.0 M ammonium sulfate, 0.1 M cacodylate buffer pH 6.5, 0.2 M NaCl. PduA-K26A S40GSG (22.0 mg ml^{-1}) crystallized in hexagonal plate and trigonal pyramid crystals in multiple conditions; crystals were optimized from 0.2 M potassium sulfate and 2.2 M ammonium sulfate. PduA-K26A S40H (10.0 mg ml^{-1}) crystallized in hexagonal plate crystals in multiple conditions; crystals were optimized from 0.2M sodium/potassium tartrate and 2.2 M ammonium sulfate. PduA-K26A S40C (10.0 mg ml^{-1}) crystallized in hexagonal plate crystals in multiple conditions; crystals were optimized from 0.5 M ammonium sulfate, 0.1 M HEPES pH 7.5, 30% MPD.

Data collection and structure determination

Protein crystals were cryoprotected with 35% glycerol to 65% well solution. Diffraction data were collected at the Advanced Photon Source, beamline 24-ID-C, at Argonne National Laboratories, IL. Data sets were integrated, merged, and scaled using the XDS suite programs for crystallography. Phases

for diffraction intensities were obtained by molecular replacement using wild type PduA (PDB ID: 3NGK) or PduA-K26A (PDB ID: 4PPD) as a search model using PHASER. Structures were refined using PHENIX and manual manipulation using COOT. For PduA-K26A S40L, protein geometry analysis revealed no Ramachandran outliers, with 97.9% residues in favored regions; Molprobity clash score after adding hydrogens is 6.11. For PduA-K26A S40GSG, protein geometry analysis revealed 1.6% residues as Ramachandran outliers, with 95.7% residues in favored regions; Molprobity clash score after adding hydrogens is 10.01. For PduA-K26A S40H, protein geometry analysis revealed no Ramachandran outliers, with 98.7% residues in favored regions; Molprobity clash score after adding hydrogens is 8.2. For PduA-K26A S40C, protein geometry analysis revealed no Ramachandran outliers, with 97.5% residues in favored regions; Molprobity clash score after adding hydrogens is 11.13. For PduA-K26A S40Q, protein geometry analysis revealed 0.4% residues as Ramachandran outliers, with 96.5% residues in favored regions; Molprobity clash score after adding hydrogens is 13.73. Data collection statistics are shown in Table S1.

Calculation of pore properties

The electrostatic potential along the central axis of the pore was calculated using in-house programs to evaluate the Poisson-Boltzmann equations on a 0.75 Å grid, with physiological ionic strength and a dielectric value for the protein interior approximated at 20. Solvation free energies were calculated for analogous pore segments in each hexamer assembly as indicated in Table 2, using solvent accessible surface area estimated by AREAIMOL in the CCP4 suite of crystallography programs and atomic solvation parameters from Eisenberg & McLachlan (12).

Supplementary Information

Table S1: Data collection and refinement statistics (Molecular replacement)

	PduA-K26A S40L	PduA-K26A S40GSG	PduA-K26A S40H	PduA-K26A S40C	PduA-K26A S40Q
Data collection					
Space group	P 6 2 2	F 2 3	P 6(5) 2 2	F 2 3	P 6(5) 2 2
Cell dimensions					
a, *b*, *c* (Å)	116.8, 116.8, 64.1	236.8, 236.8, 236.8	108.8, 108.8, 334.3	235.4, 235.4, 235.4	108.6, 108.6, 336.3
α, *β*, *γ* (°)	90, 90, 120	90, 90, 90	90, 90, 120	90, 90, 90	90, 90, 120
Wavelength (Å)	0.9792	0.9791	0.9793	0.9789	0.9789
Resolution (Å)	58.4-2.30 (2.36-2.30) *	19.7-3.10 (3.18-3.10) *	94.2-2.00 (2.05-2.00) *	135.9-3.30 (3.38-3.30) *	20.0-2.79 (2.86-2.79) *
R_{sym} or R_{merge}	0.102(1.38)	0.076(0.694)	0.161(2.27)	0.163(1.32)	0.342 (2.34)
$I/\sigma I$	15.6(1.08)	11.6 (1.73)	11.8(1.05)	32.5(4.24)	11.0 (1.84)
Completeness (%)	98.6(94.2)	97.4(98.8)	99.4(93.3)	99.7(99.8)	99.6(95.2)
Redundancy	13.6(5.7)	2.9(2.90)	10.8(10.4)	39.9(38.9)	36.7(25.6)
Refinement					
Resolution (Å)	54.1-2.30	19.5-3.10	94.2-2.00	135.9-3.30	94.1-2.79
No. of unique reflections	11771	19473	79885	14728	30231
R_{work}/ R_{free}	.229/.268	.194/.218	.181/.219	.192/.224	.229/.253
No. atoms	1750	4172	6077	4301	5487
Protein	1738	4148	5612	4291	5436

Ions	5	15	164	10	25
Water	4	2	301		26
B-factors					
Protein	80.2	95.1	45.3	95.5	60.5
Ions	134.3	156.2	67.2	151.0	117.0
Water	53.5	66.4	48.8		47.9
R.m.s deviations					
Bond lengths (Å)	.014	.010	.018	.027	.012
Bond angles (°)	1.9	1.3	1.7	1.8	1.8

Each structure was determined from one crystal.

*Highest resolution shell is shown in parenthesis.

Table S2: Summary of PduA pore mutant structures by pore symmetry and pore diameter

PduA Mutant	Structure	Space group	Copies in ASU	Pore symmetry	Pore diameter* (Å)
WT	3NGK	P 6 2 2	1	6-fold from 1 monomer in asu	5.6
-S40L	solved	P 6 2 2	3	(1) 3-fold symmetry from 2 distinct monomer types, chains A and B	5.9
				(2) 6-fold from 1 monomer type, chain C	5.1
-S40A	modeled				5.9
-S40Q	solved	P 6(5) 2 2	9	(1) 2-fold from 3 monomer types, chains ABC, symmetry broken by translation	0.2
				(2) 6 monomer types, chains DEFGHI, no symmetry about the pore	0.9
-S40GSG	solved	F 2 3	7	(1) 3-fold from 2 monomer types, chains AB	5.0
				(2) 3-fold from 2 monomer types, chains CD	0.6
				(3) 2-fold from 3 monomer types, chains EFG	Occluded
-S40H	solved	P 6(5) 2 2	9	(1) 6 monomer types, chains ABCDEF, no symmetry about the pore	3.0 × 6.1
				(2) 2-fold from 3 monomer types, chains GHI, symmetry broken by translation	3.0 × 5.6
-S40C	solved	F 2 3	7	(1) 3-fold from 2 monomer types, chains AB which formed disulfide bridges between adjacent C40	2.3

	(2) 3-fold from 2 monomer types, chains CD	2.2
	(3) 2-fold from 3 monomer types, chains EFG, symmetry slightly broken by translation	3.4

*Pore diameter was calculated as the distance between two opposite-facing atoms at the narrowest point through the pore subtracting their corresponding van der Waals radii

Table S3: DDH activity of purified PduA-S40C MCPs in presence of DTT and $FeCl_2$.

Pdu mutants	Specific activity (μmol min^{-1} mg^{-1})
WT	28.8 ± 1.5
PduA-S40C	27.1 ± 0.2
WT + Fe^{2+}	28.9 ± 0.6
PduA-S40C + Fe^{2+}	27.9 ± 0.5
WT + DTT	25.2 ± 0.9
PduA-S40C + DTT	25.5 ± 0.9
WT + Fe^{2+} + DTT	25.2 ± 0.9
PduA-S40C + Fe^{2+} + DTT	18.3 ± 0.3

Table S4: PduP and DDH activity in intact and broken MCPs

Pdu mutants	PduP activity ($\mu mol\ min^{-1}\ mg^{-1}$)	DDH activity ($\mu mol\ min^{-1}\ mg^{-}$)
WT	2.8 ± 0.2	27.8
WT (broken)*	2.9 ± 0.6	34.2 ± 1.5
PduA-S40L	2.6 ± 0.1	15.4
PduA-S40L (broken)	2.8 ± 0.1	33.8 ± 0.3
PduA-S40GSG	2.8 ± 0.8	18.1
PduA-S40GSG (broken)	ND**	34.5 ± 0.5
PduA-S40Q	ND	22.5
PduA-S40Q (broken)	ND	35.7 ± 0.8
ΔpduA	ND	39.1 ± 0.2

*MCPs were broken by dialysis and sonication. **Not determined

Table S5: Bacterial Strains used in this study

Strain	Genotype/variant	Source
LT2	*Salmonella enterica* serovar Typhimurium strain LT2	Lab collection
BE182	LT2, *ΔpduAB652*	Lab collection
BE213	LT2, *ΔpduBB'675*	Lab collection
BE293	LT2/pKD46	Lab collection
SS71	LT2, *ΔpduA689*	Lab collection
CC3	LT2, PduA-S40L	This study
CC12	LT2, PduA-S40M	This study
CC13	LT2, PduA-S40GSG	This study
CC14	LT2, PduA-S40T	This study
CC15	LT2, PduA-S40Q	This study
CC43	LT2, PduA-S40N	This study
SS59	LT2, PduA-S40A	This study
SS60	LT2, PduA-S40C	This study
CC44	LT2, PduA-S40H	This study
CS471	*E. coli* NEB-T7-express/pTA925-	Lab collection

	pduCDE	
BE903	LT2, *ΔpduQ688::frt*	Lab collection
BE2158	LT2, *adhP1*::*cam*(sw)	gift from Prof. John Roth

Table S6: Summary of the characteristics of PduA pore mutants

Mutant	Pore size	Polarity at the pore region	Growth rate at limiting B_{12}	Aldehyde efflux from the MCP	DDH activity of purified MCPs
WT	Normal	Normal	Normal	Normal	28.4 ± 0.6
-S40L	Normal	Hydrophobic	Slower	Normal	14.8 ± 0.8
-S40GSG	Occluded	Normal	Slower	Normal	17.5 ± 0.5
-S40Q	Reduced	Polar	Slightly slower	ND*	21.2 ± 0.7
-S40C	Slightly reduced	Normal	Slower	ND	27.1 ± 0.2**
-S40A	Normal	Normal	Normal	Faster	26.1 ± 0.8
-S40H	Slightly reduced	Polar	Normal	ND	28.5 ± 0.4
-S40T	Normal	Normal	Normal	ND	29.4 ± 0.4
-S40M	Normal	Normal	Normal	ND	29.2 ± 0.2
-S40N	Normal	Normal	Normal	ND	27.8 ± 0.8

*ND: Not determined

**The PduA-S40C mutant has slower activity depending on the presence of iron and DTT (see Table S3).

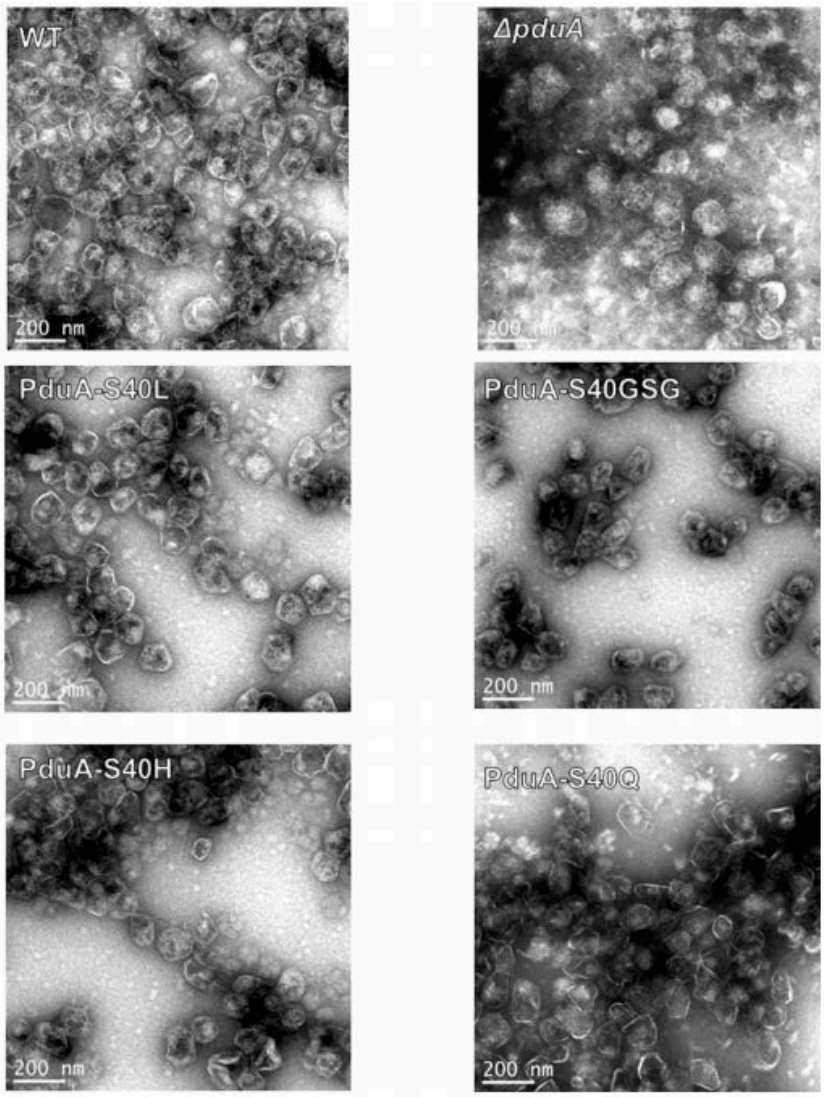


Figure S1: Transmission electron micrographs of purified MCPs from wild-type and PduA pore mutants. All the pore mutants investigated in this study produced intact microcompartments with no visible morphological deformities as compared to wild-type. Only the pore mutants most important to this study are shown. A *ΔpduA* mutant was used as a control known to produce MCPs with a more porous shell with enlarged morphology.

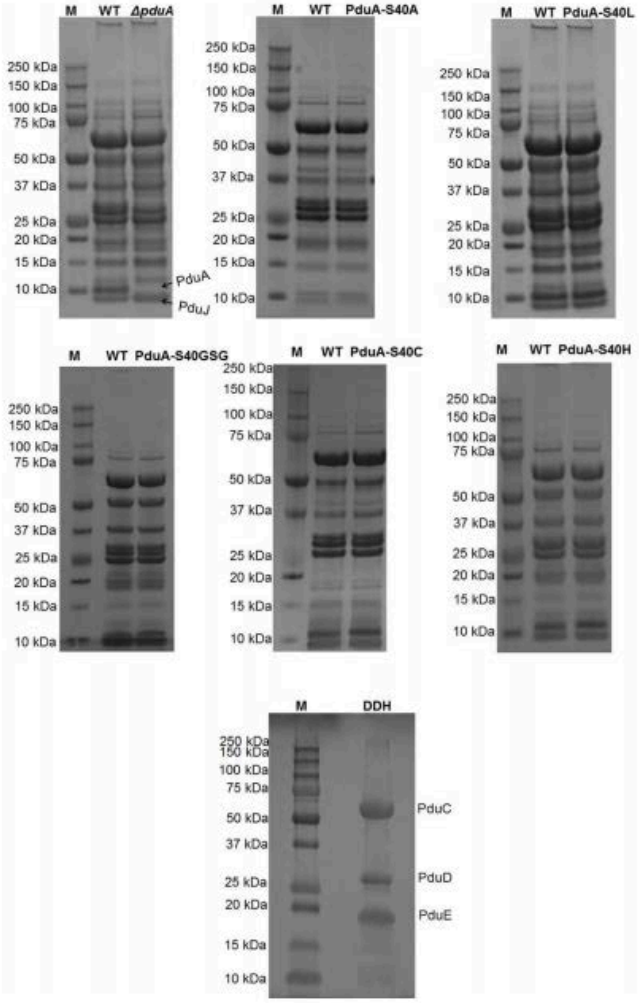
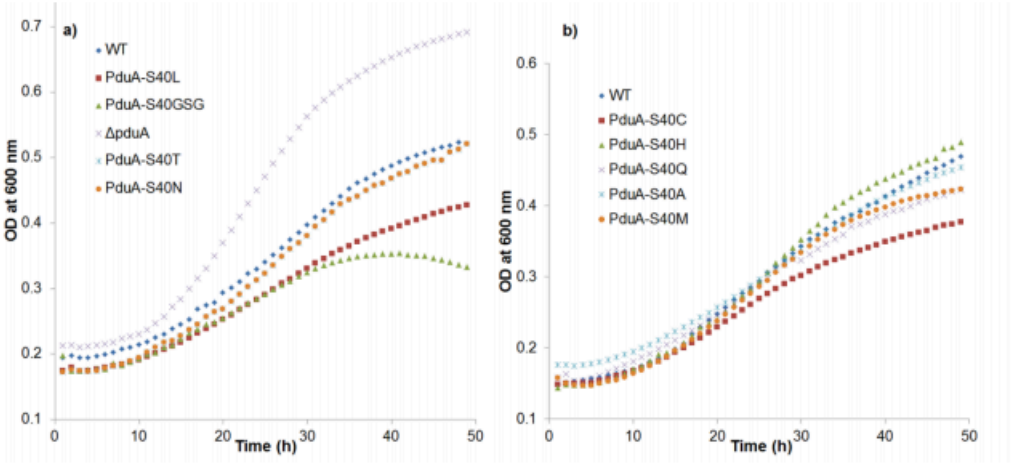


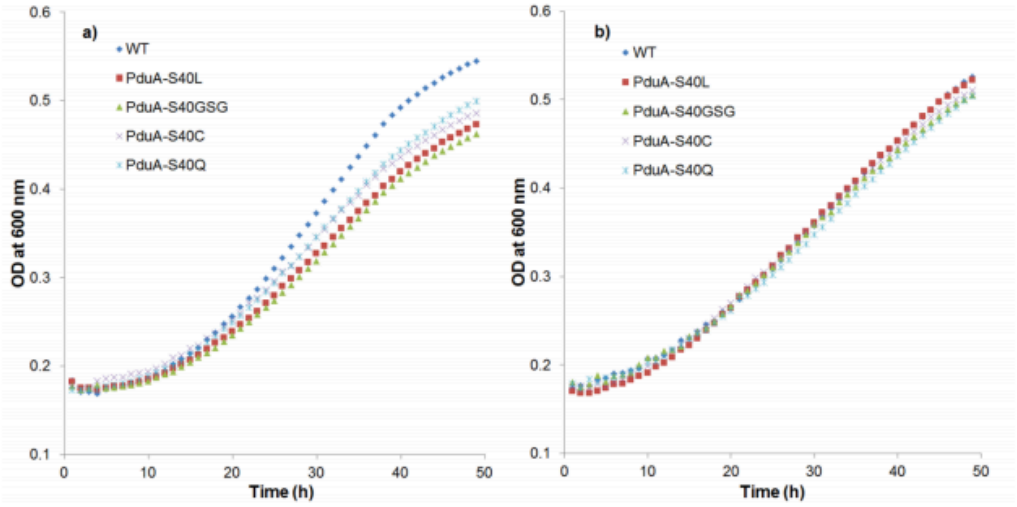
Figure S2: SDS-PAGE of purified MCPs from several PduA pore mutants and wild-type *S. enterica* (LT2) and purified diol dehydratase. Approximately 15-20 μg purified MCPs were loaded in each well of 4-12% polyacrylamide gels. In the case of purified DDH, approximately 15 μg purified protein was loaded to check the purity.



c.

Strain	Doubling time (h)	Strain	Doubling time (h)
WT	16.5 ± 1.2	PduA-S40M	16.4 ± 0.2
PduA-S40GSG	24.03 ± 0.7	PduA-S40T	16.2 ± 0.7
PduA-S40L	23.9 ± 0.4	PduA-S40N	16.1 ± 0.5
PduA-S40C	21.5 ± 0.7	PduA-S40H	15.8 ± 0.6
PduA-S40Q	18.2 ± 0.6	*ΔpduA*	8 ± 0.4
PduA-S40A	17.9 ± 0.2		

Figure S3: Growth of PduA pore mutants on 1,2-propanediol minimal medium with limiting B_{12}. Panels a and b are similar experiments done on different days with different mutants. Panel c shows the doubling times and standard error of the wild-type then the mutants descending order. The *ΔpduA* mutant was used as a control for broken MCPs which have been shown to have an increased growth rate on 1,2-PD (11). No propionaldehyde toxicity is seen in these experiments because they were performed using limiting levels (20 nM) of added B_{12}, the required cofactor for diol dehydratase (the enzyme that forms propionaldehyde). Growth assays were repeated at least three times in triplicate or more and representative curves are shown



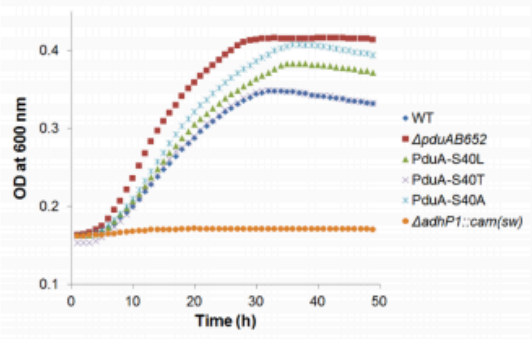
c)

Strain	Doubling time (h)	Strain	Doubling time (h)
O.4% 1,2-PD		**2.4% 1,2-PD**	
WT	17.8 ± 1.1	WT	18.83 ± 0.43
PduA-S40L	23.02 ± 0.92	PduA-S40L	20.1 ± 0.7
PduA-S40GSG	22.5 ± 0.84	PduA-S40GSG	19.4 ± 0.75
PduA-S40C	21.10 ± 0.82	PduA-S40C	18.4 ± 0.8
PduA-S40Q	21.48 ± 0.4	PduA-S40Q	18.9 ± 0.67

Figure S4: Growth of rescue of certain PduA pore mutants at higher levels of 1,2-PD. Panel a is low levels of 1,2-PD (0.4%) and Panel b is higher levels of B_{12} (2.4%). Panel c lists the growth rates and standard errors. These experiments were repeated three times with similar results. Limiting levels (30 nM) of B_{12} were used.

Growth of PduA pore mutants on 1-propanol.

Results showed that growth of *Salmonella* on 1-propanol minimal medium required induction of the *pdu* operon indicating that *pdu* genes were needed for 1-propanol metabolism. No B_{12} was added to cultures which prevented growth on 1,2-PD under these conditions. We also found that deletion of the AdhP alcohol dehydrogenase (a cytoplasmic enzyme) prevented growth on 1-propanol and that deletion of the PduP propionaldehyde dehydrogenase (a MCP lumen enzyme (10)) increased the doubling time on 1-propaniol from 22.14 ± 0.4 h for wild-type to 55.6 ± 0.5h. In contrast, deletion of the PduQ alcohol dehydrogenase had no effect on growth of *Salmonella* on 1-propanol showing that it was not involved in this process. These results support a model in which growth of *Salmonella* on 1-propanol requires conversion 1-propanol to propionaldehyde in the cytoplasm by AdhP followed by diffusion of propionaldehyde across the MCP shell and into the lumen where it is converted to propionyl-CoA by the PduP enzyme (10). The propionyl-CoA then exits the MCP and enters the central metabolism via the methylcitrate pathway allowing growth (13). Growth tests also showed that genetic disruption of the Pdu MCP shell resulted in faster growth on 1-propanol suggesting that the diffusion of propionaldehyde into the Pdu MCP limited growth. Importantly, extensive genetic studies of *Salmonella* growing on ethanol using enzymes of the ethanolamine utilization MCP gave analogous results, and indicated that growth was limited by diffusion of acetaldehyde across the shell of the Eut MCP (14). This further supported the idea that growth of *Salmonella* on 1-propanol is limited by diffusion of propionaldehyde from the cytoplasm into the Pdu MCP. Hence, we used growth of *Salmonella* on 1-propanol as a measure of thrate of propionaldehyde diffusion across the shell and into Pdu MCP (Fig. S5). These studies indicated that the relative rate of propionaldehyde movement from the cytoplasm into the Pdu MCP were PduAS40A > PduS40L > wild-type ≈ PduAS40T. No propionaldehyde toxicity was observed in these experiments indicating that this aldehyde did not accumulate to toxic levels when produced from 1-propanol by AdhP under the conditions used



Strain	Doubling time (h)
WT	21.4 ± 0.4
*ΔpduAB*652	10.5 ± 0.2
PduA-S40L	18.9 ± 0.9
PduA-S40A	16.3 ± 0.7
PduA-S40T	22.3 ± 0.4
ΔadhP1::cam(sw)	-

Figure S5: Growth of PduA pore mutants on 1-propanol. Under the conditions used, growth rate is a measure of the rate at which propionaldehyde moves across the MCP shell (see above). WT is wild-type *S. enterica*. The *ΔpduAB* mutation breaks the shell of the Pdu MCP allowing propionaldehyde to reach MCP lumen enzymes by free diffusion. The Δ*adh*P1::cam (BE2158) lacks the cytoplasmic AdhP alcohol dehydrogenase which converts 1-propanol to propionaldehyde. This experiment was repeated three times in triplicate and the representative data are shown.

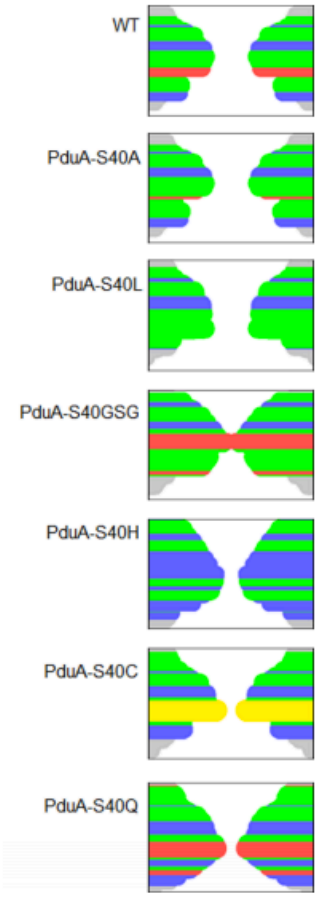


Figure S6: Polarity at the PduA pore viewed as a cross section centered at z=0 and r=0. Atom type plotted through the PduA hexamer pore oriented with cytosolic face at the top and MCP lumenal face at the bottom is shown (from top image to bottom) for wild-type PduA (PDBID: 3NGK), PduA-S40A (model), PduA-S40L (PDBID: 4RBT), PduA-S40GSG (PDBID: 4RBV), PduA-S40H (PDBID: 4QIF), PduA-S40C (PDBID: 4QIG), PduA-S40Q (PDBID: 4RBU), and KCSA (PDBID: 1BL8, potassium channel from *Streptomyces lividans*). Atom types are represented by color: C, green; N, blue; O, red; and S, yellow. The shape is based on van der Waals radii (where radii >15Å are represented in grey).

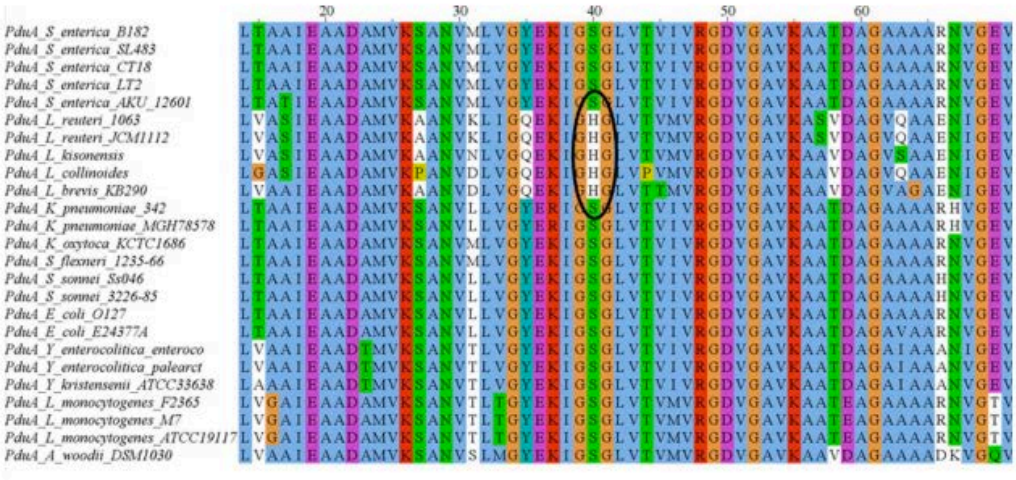


Figure S7: Multiple sequence alignment of PduA shell homologs. In *Lactobacillus* sp. H40 is present rather than S40.

Supplementary References

1. Miller JH (1972) *Experiments in molecular genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).
2. Berkowitz D, Hushon JM, Whitfield HJ, Jr., Roth J, & Ames BN (1968) Procedure for identifying nonsense mutations. *J. Bacteriol.* 96(1):215-220.
3. Cheng S, Sinha S, Fan C, Liu Y, & Bobik TA (2011) Genetic analysis of the protein shell of the microcompartments involved in coenzyme B_{12}-dependent 1,2-propanediol degradation by *Salmonella*. *J Bacteriol* 193(6):1385-1392.
4. Guex N & Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18(15):2714-2723.
5. Datsenko KA & Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* 97(12):6640-6645.
6. Sinha S, *et al.* (2014) Alanine scanning mutagenesis identifies an asparagine-arginine-lysine triad essential to assembly of the shell of the Pdu microcompartment. *J. Mol. Biol.* 426(12):2328-2345.
7. Sinha S, Cheng S, Fan C, & Bobik TA (2012) The PduM protein is a structural component of the microcompartments involved in coenzyme B_{12}-dependent 1,2-propanediol degradation by *Salmonella enterica*. *J Bacteriol* 194(8):1912-1918.
8. Tobimatsu T, *et al.* (1997) Heterologous expression, purification, and properties of diol dehydratase, an adenosylcobalamin-dependent enzyme of *Klebsiella oxytoca*. *Arch Biochem Biophys* 347(1):132-140.
9. Fan C & Bobik TA (2011) The N-terminal region of the medium subunit (PduD) packages adenosylcobalamin-dependent diol dehydratase (PduCDE) into the Pdu microcompartment. *J Bacteriol* 193(20):5623-5628.
10. Leal NA, Havemann GD, & Bobik TA (2003) PduP is a coenzyme-A-acylating propionaldehyde dehydrogenase associated with the polyhedral bodies involved in B_{12}-dependent 1,2-propanediol degradation by *Salmonella enterica* serovar Typhimurium LT2. *Arch Microbiol* 180(5):353-361.
11. Sampson EM & Bobik TA (2008) Microcompartments for B_{12}-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. *J Bacteriol* 190(8):2966-2971.
12. Eisenberg D & McLachlan AD (1986) Solvation energy in protein folding and binding. *Nature* 319(6050):199-203.
13. Horswill AR & Escalante-Semerena JC (1999) *Salmonella typhimurium* LT2 catabolizes propionate via the 2-methylcitric acid cycle. *J. Bacteriol.* 181(18):5615-5623.
14. Huseby DL & Roth JR (2013) Evidence that a metabolic microcompartment contains and recycles private cofactor pools. *J Bacteriol* 195(12):2864-2879.

Chapter 4

The free energy barrier across the PduA pore

is higher for the toxic aldehyde than for propanediol

Introduction

Why would encapsulating a metabolic intermediate be beneficial? The intermediate can be highly reactive or toxic to the cell and encapsulating them can prevent unwanted reactions[11]. Also, the rate of diffusion is often greater than rates of reactions so MCPs can hinder diffusion across concentration gradients and concentrate the intermediate near enzyme scaffolds, which have been shown to increase reaction efficiencies.

The results from biochemical study showed that the PduA hexamer pore is a route of small molecule transport and that a key pore residue S40 may use its hydroxyl side chain to limit the efflux of propionaldehyde[23]. The PduA hexamer pore is relatively thin compared to other transemembrane pore channel proteins and has not been observed in other conformations (such as with a wider pore). We do not expect a transport mechanism that involves multiple conformational changes. Instead, we conjectured that the mechanism of small molecule transport is passive but also selective by the presence of the S40 hydroxyl side chains. In this aim, we determined to calculate a difference in free energy to penetration of the pore by the substrate molecule as compared to the toxic intermediate molecule.

In order to calculate the free energy barrier to pore penetration by small molecules, we employed the molecular dynamics technique metadynamics. In metadynamics, a history-dependent bias potential V_{meta} accelerates sampling in the form of small Gaussian "hills"

composed of height W centered at some reaction coordinate ε_i for the ith of N_{CV} collective variables, accumulated over time t as shown below[24]:

$$V_{meta}(\varepsilon) = \sum_{t'=\delta t,\, 2\delta t,\, \ldots}^{t'<t} W \prod_{i=1}^{N_{CV}} e^{\left(\frac{-\left(\varepsilon_i - \varepsilon_i(t')\right)^2}{2\delta^2}\right)}$$

The hills accumulate in the energy landscape over the course of the metadynamics simulation, enabling a system to overcome local energy barriers and explore other local minima (Figure 4.1). Accounting for V_{meta} can recover the underlying free energy profile A.

This method converges more smoothly in the well-tempered version of the program where V_{meta} changes based on the bias temperature ΔT, the hill height ω, and the normalized histogram of collective variables visiting reaction coordinate ε at time t:

$$V_{meta}(\varepsilon, t) = \Delta T ln\left(1 + \frac{\omega N(\varepsilon, t)}{\Delta T}\right)$$

Thus, ΔT limits the exploration of the free energy surface to an energy range of $T+\Delta T$. Tuning ΔT to lower values can increase barrier crossing and facilitate efficient exploration of CVs based on how V changes[25,26]:

$$V_{n+1}(\varepsilon) = V_n(\varepsilon) + G(\varepsilon, \varepsilon_{n+1}) e^{-V_n(\varepsilon_{n+1})/\Delta T}$$

In the PduA hexamer system, the collective variables were chosen to represent the exploration of small molecules in the hexamer pore region. The two collective variables were defined as distXY or radius of the small molecule from the z-axis through the hexamer pore and distZ or the z-

coordinate (see Figure 4.2). The results of this study are described in the following article that is a version of a manuscript in submission.

Figure 4.1. A schematic of the Metadynamics technique in which a history-dependent bias potential , $V_{meta}(\varepsilon)$, is applied to a collective variable, ε, to facilitate calculate of the underlying free energy surface, $A(\varepsilon)$.

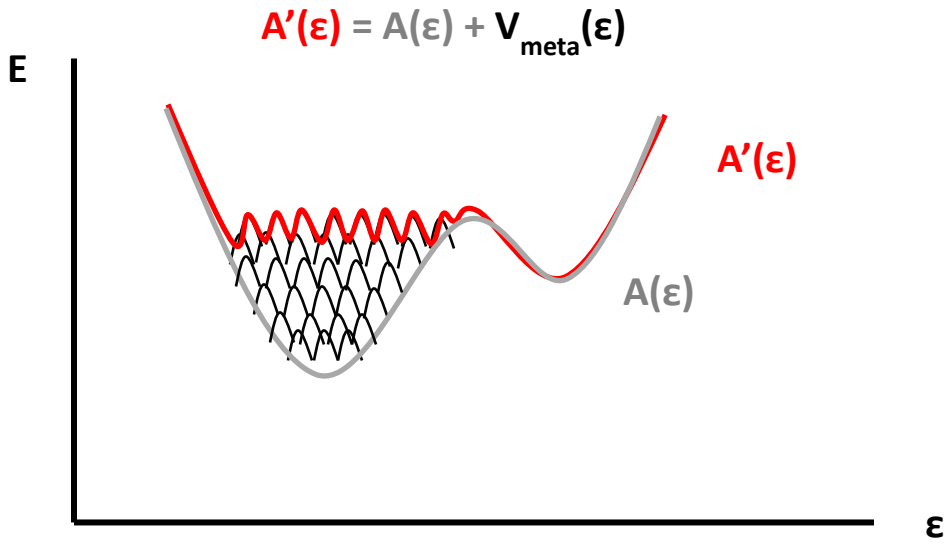


Figure 4.2. Side view of the PduA hexamer with Z-axis through the pore center marked as a bold line and maximum radius away from the Z-axis marked in dashed lines.

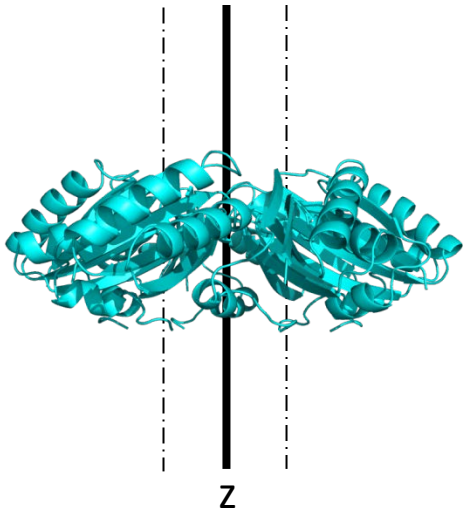
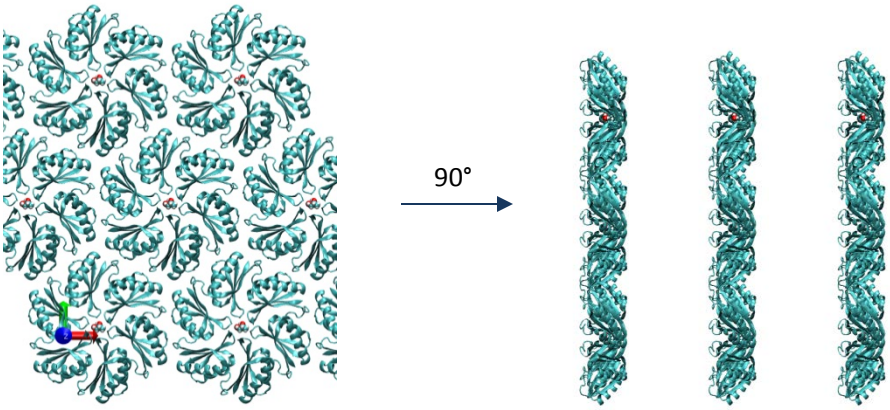


Figure 4.3. The PduA hexamers are tiled in a hexagonal periodic box and simulated as flat layers in explicit TIP3P water solvent (not shown).



Molecular Dynamics Simulations of Selective Metabolite Transport Across the Propanediol Bacterial Microcompartment Shell

S. Chun,* J. Park,† T. A. Bobik,% K. N. Houk,† T. O. Yeates*‡

* Molecular Biology Institute, University of California, Los Angeles; † Department of Chemistry and Biochemistry, University of California, Los Angeles; ‡ UCLA-DOE Institute for Genomics and Proteomics; % Roy J. Carver Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University

ABSTRACT Bacterial microcompartments are giant protein-based organelles that encapsulate certain metabolic pathways in phylogenetically diverse bacteria. Structural and genetic studies indicate that metabolic substrates enter these microcompartments by passing through the central pores in hexameric assemblies of shell proteins. Limiting the escape of toxic metabolic intermediates created inside the microcompartments would likely confer a selective advantage. Here, we report the first molecular dynamics (MD) simulations to analyze small-molecule transport across the BMC shell. PduA is a major shell protein in a bacterial microcompartment that metabolizes 1,2-propanediol via a toxic aldehyde intermediate, propionaldehyde. Using both metadynamics and replica-exchange umbrella sampling, we find that the pore of the PduA hexamer has a lower energy barrier for passage of (and is therefore more permeable to) the 1,2-propanediol substrate, versus the toxic propionaldehyde generated within the microcompartment. The energetic effect is consistent with a lower capacity of a serine residue at the point of constriction in the pore to hydrogen bond to propionaldehyde, relative to the more freely permeable 1,2-propanediol. These results highlight the importance of molecular diffusion and transport in a new biological context, separate from the already well-studied cases of transport through transmembrane protein channels and pumps.

Received for publication "Staff will complete" and in final form "Staff will complete"

Address reprint requests and inquiries to

Diverse biological processes rely on the regulated transport of small molecules across barriers between cells or between different subcellular compartments. Membrane channels and pumps enable the regulated movement of molecules across lipid bilayers; molecular transport in transmembrane protein systems has been studied at the functional and structural level for over 50 years[1,2]. Recently, an entirely distinct class of proteins has come under investigation in the context of molecular transport into and out of giant protein-only organelles known as bacterial microcompartment. These structures act in diverse bacterial species to sequester key enzymatic pathways that involve toxic or volatile chemical intermediates[3–5]. A key aspect of microcompartment function is that metabolic substrates must enter and products must exit through diffusive processes.

The outer shells of bacterial microcompartments are formed primarily by a tightly packed layer of hexameric proteins that belong to the so-called BMC family. BMC proteins assemble into cyclic hexamers with a narrow central pore. The pore has been shown experimentally to be the route for diffusive molecular transport in these systems (Figure 1)[6,7]. The starting substrate for the encapsulated pathway must pass efficiently from the cytosol into the microcompartment interior. Furthermore, it has been argued that a physiological advantage of the microcompartment would be conferred by a pore that is less permeable to volatile or cytotoxic intermediates produced in the interior, which would otherwise escape the microcompartment and damage the cell's DNA[8]. However, to date it has not been possible to measure experimentally the relative permeabilities of the starting substrate and the toxic intermediate. This is due to the complexity of bacterial microcompartments, which are comprised of thousands of shell proteins and enzyme molecules, and the difficulties of direct transport studies in vitro. Crystal structures of the relevant BMC shell proteins[6,7] offer opportunities to examine their transport properties in atomistic detail via molecular dynamics (MD) simulations.

In this first study of molecular transport through BMC shell proteins, we examine the hexameric BMC shell protein known as PduA, whose structure we determined in an earlier study[6]. The PduA pore is the route for the diffusive influx of 1,2-propanediol into the Pdu microcompartment. The encapsulated enzymes in the Pdu microcompartment transform the substrate into the cytotoxic intermediate propionaldehyde and then into non-toxic metabolites. These last steps must occur before the aldehyde can diffuse outward across the shell and into the cytosol[7–9]. We tested the hypothesis that the PduA pore is more permeable to the propanediol substrate compared to the propionaldehyde intermediate by computing free energy profiles of the two small molecules along the diffusion pathway across the pore.

To validate our results by independent methods, we employed both metadynamics and replica exchange

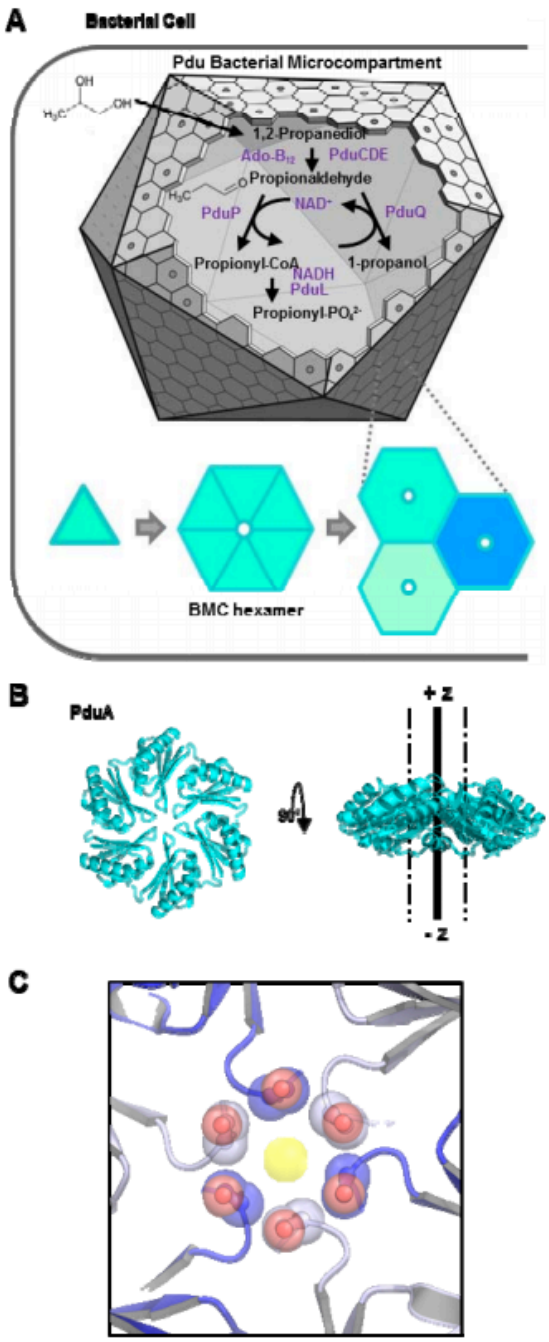


Figure 1. Molecular transport across the Pdu bacterial microcompartment shell. (A) The 1,2-propanediol utilization (Pdu) pathway converts 1,2-propanediol to propionate and propanol via a toxic aldehyde intermediate. The bacterial microcompartment shell is built from hexameric BMC shell proteins that bear central pores. (B) PduA is a BMC shell protein of the Pdu microcompartment and is the route of substrate (1,2-propanediol) entry. The thick line, running from $-z$ to $+z$, indicates the molecular six-fold axis; the pair of broken lines parallel to this axis demarcate the radial limits (in our cylindrical coordinate system) within which the diffusing substrate was confined in our simulations. (C) A zoomed in view of the PduA hexamer pore (cartoon) showing residue S40 (spheres) lining the pore. A yellow sphere highlights the space in the pore (radius 2.8Å).

umbrella sampling (REUS)[10,11] as two complementary approaches for computing free energy barriers via all-atom simulations. In the metadynamics approach,[10] the Hamiltonian of the solvated system is augmented with a history-dependent bias potential (v_{meta}), which gradually flattens the local minima and maxima of the underlying energy landscape for a given set of reaction coordinates (i.e., collective variables), thereby decreasing barrier heights and accelerating sampling of the system's phase space. Without a biasing potential, conventional unbiased MD simulations are often too slow in exploring important but rare events in complex systems with rugged energy landscapes. Upon convergence of a metadynamics simulation, the biasing potential reveals the underlying (unbiased) free energy landscape defined by the collective variables[12].

In the PduA hexamer system, the coordinates of the center of mass of a small molecule able to move within the hexamer pore region (Figure 1) were chosen as the collective variables describing the reaction coordinate for diffusion. We defined a cylindrical coordinate system relative to the pore center, where the z-axis passes through the center of the pore and positive values of z correspond to the exterior of the compartment. The radial distance from the z-axis (r) and the vertical position (z) of the small molecule were analyzed in all simulations. In order to model an intact two-dimension layer of protein molecules, the hexameric PduA shell protein was constructed as a tiled layer, as would be found on a flat facet of the Pdu microcompartment shell[6,13]. In addition to the PduA hexamer, our simulation systems contain explicit solvent and a small molecule, either 1,2-propanediol or propionaldehyde (see Extended Data for method details.)

The diffusing small molecules—propanediol and propionaldehyde—were simulated at the PduA pore for total times of 640 ns and 320 ns of metadynamics, respectively, for $0Å < r < 15Å$ and $-30Å < z < 30Å$. Inside the PduA protein ($-8Å < z < 8Å$), each substrate experiences a higher energy than in the bulk solvent region ($z > 8Å$ or $z < -8Å$). The path of lowest energy passes through the open center of the pore ($r=0Å$) as expected, and exhibits a potential energy barrier near $z=0Å$. A comparison of the two potential of mean force (PMF) profiles through the pore reveals that propionaldehyde

experiences a higher free-energy barrier, by about 1.8 kcal/mol (Figure 2). In a simplified treatment where the rate of transport is dictated only by the free energy peak height according to the Boltzmann distribution, then the peak height energy difference between the two cases provides an indication of the relative permeability,

$$P_{rel} = \exp(-\beta(G^T_{PDO} - G^T_{PPN})) \quad \text{(Eq 1)}$$

where P_{rel} is the relative permeability, $\beta = 1/k_BT$ is the Boltzmann factor at room temperature, and G^T_{PDO} and G^T_{PPN} are the energy barrier heights of propanediol and propionaldehyde, respectively. With an energy difference in the present case of 1.8 kcal/mol, the relative permeability would be about 21. However, a more correct treatment of diffusive rates must account for the full free energy profiles as a function of the axial position, z. Under simplifying assumptions (see Extended Data), an expression of the following form can be obtained for the diffusive permeability:

$$P \propto D / \int \exp(-\beta G(z))dz \quad \text{(Eq 2)}$$

where P is the permeability, D is the diffusion coefficient, and $G(z)$ is the free energy as a function of position, z. Our derivation of permeability of small molecules diffusing through a channel parallels the study by Bauer and Nadler (2006) on the transport of particles that can bind in a channel[14]. We used the expression above to analyze the free energy profiles (Fig. 2); in so doing, we ignored small differences in the diffusion coefficients for the two molecules in question (i.e. we accounted only for the denominator term), and evaluated the integral only over the range of z values corresponding to the narrowest constriction of the pore (-3.5Å < z < 1.5Å). Using this somewhat more complete approach, we estimate the permeability for 1,2-propanediol to be ≈ 19 times higher than for propionaldehyde.

As a complementary approach to estimating free-energy barriers and relative permeabilities, we also used replica exchange umbrella sampling (REUS) to examine the PMF of the two small molecules through the PduA pore[11]. The REUS simulations are intended to obtain an accurate energy height at the center of the pore, relative to the bulk solvent. The PMF profiles in the bulk solvent region (|z| > 8Å) computed by the metadynamics MD simulations fluctuate significantly, a drawback of metadynamics MD simulations reported previously[12]. In REUS, the small molecule was restrained, with a 0.5 kcal/mol/Å^2 spring constant, at 59 overlapping windows of width 0.5 Å across the z-axis through the PduA pore. The REUS simulations were computed for 1003 ns and 590 ns for PDO and PPN, respectively, and evaluated via the weighted histogram analysis method (WHAM)[16]. The PMF profiles for propanediol and propionaldehyde through the WT PduA pore indicate a barrier for the aldehyde that is ≈ 0.8 kcal/mol higher than for the diol near the center of the PduA pore (Figure 3). Our error estimates from the REUS

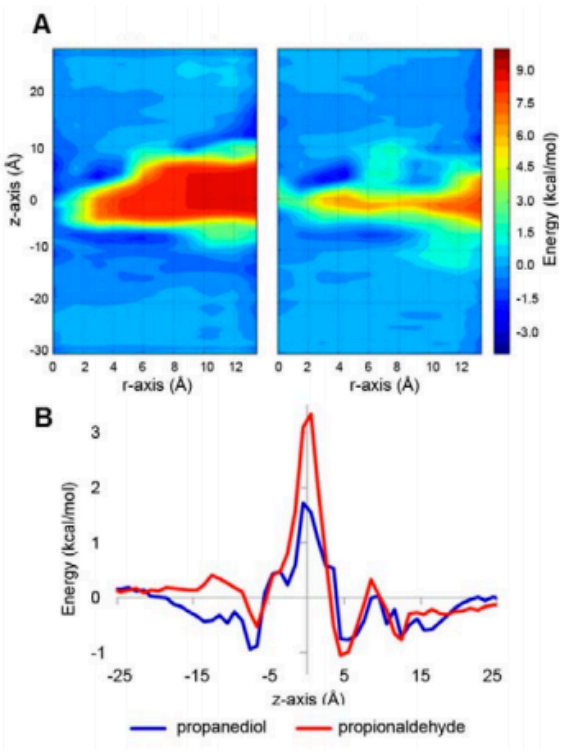


Figure 2. Metadynamics potential of mean force profiles for small molecules diffusing through the PduA hexamer pore represented as (A) 2-dimensional heatmaps for propanediol (left) and propionaldehyde (right), and as (B) 1-dimensional line plots through the pore (r=0) for propanediol (blue) and propionaldehyde (red).

calculation are relatively large, yielding greater uncertainty in the analysis of relative permeabilities. Nevertheless, following the same treatment as described above for analyzing the effects of the free energy profile, the permeability of the propanediol substrate is calculated to be ≈ 2.4 times higher than that of the propionaldehyde intermediate. The difference in numerical values of permeabilities obtained by the two enhanced sampling approaches illustrates the degree of uncertainty in calculating energetic properties for such complex systems. This difference notwithstanding, the metadynamics and REUS results are both consistent with one another (i.e., propanediol has the higher permeability by both computational approaches) and with our overall biochemical model (i.e., the shell is less permeable to the toxic aldehyde species).

To elucidate a possible structural basis for differential permeabilities of the two small molecules, we considereded whether it might be possible to attribute the higher energy

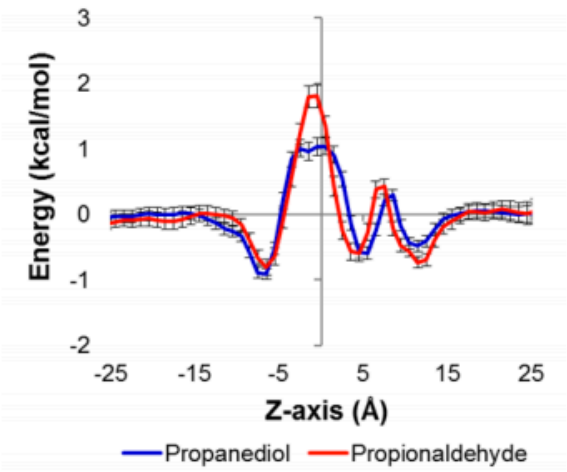


Figure 3. REUS PMF profile of propanediol (blue) and propionaldehyde (red) through the PduA hexamer pore. Consistent with the metadynamics-based PMF (Figure 2), note that the propionaldehyde energy barrier (red trace) exceeds the propanediol peak near z = 0.

barrier encountered by the aldehyde to a particular molecular feature or interaction with the PduA hexamer. It has been suggested that differences in the hydrogen bonding capacities of the aldehyde vs. the diol might be important, particularly in light of the presence of six copies of a serine residue (Ser40) at the point of narrowest constriction in the PduA pore[6]. Therefore, we computed and examined 100 ns of conventional (equilibrium, unbiased, unconstrained) MD trajectories. In these simulations, the small molecule was confined, by a limited range of z-values, to a region near the Ser40 position. We evaluated the number of hydrogen bonds that the Ser40 side-chains made with either water molecules or diffusing solute (aldehyde or diol) molecules (see Extended Data). In the conventional MD runs with the small molecule corralled near the PduA hexamer pore, the six Ser40 residues exhibited an average of 7.8±0.4 hydrogen bonds to waters or small molecule per frame in the presence of propanediol and 6.9±0.6 hydrogen bonds to waters or small molecule per frame with propionaldehyde. This result shows that the serine side chains hydrogen bond more readily with propanediol. This is an intuitive finding, given the additional hydrogen bond donating capacity of propanediol versus propionaldehyde. Hydrogen bonds and other energetic interactions in the pore are expected to enhance the flux of favorably-interacting species. Theoretical analysis has shown that this phenomenon—enhanced translocation, despite tighter binding—is explained by a large increase in the probability of a diffusing molecule occupying the pore in the presence of energetically favorable interactions; this surmounts the associated increase in mean first passage times[14].

The metadynamics simulations and REUS results from our studies of 1,2-propanediol and propionaldehyde in PduA are the first all-atom MD simulations to probe small-molecule transport in BMC shell proteins. Our MD results provide biophysical evidence to support the hypothesis that the shells of bacterial microcompartments have pores that are evolved to be selective for uptake of the necessary substrates, and are less permeable to the metabolic intermediate produced within the microcompartment. The generality of this result is, however, an open question. In the carboxysome—a different type of microcompartment, comprised of shell proteins that are homologous to PduA—structural features (such as a positively charged pore) support the idea that that shell could be more permeable to the negatively charged carboxysome substrate (bicarbonate) versus the CO_2 intermediate produce inside[15]. However, recent metabolic modeling calculations (Mangan et al., 2016, in press) argue that the carboxysome could operate efficiently without having to be selectively permeable to bicarbonate versus CO_2.

Further study of molecular transport phenomena in diverse bacterial microcompartments will be necessary to more fully understand these extraordinary systems. Future MD studies of BMC shell pores could be extended to the alpha and beta-type carboxysomes, ethanolamine utilizing bacterial microcompartments, and more recently discovered microcompartment types that sequester glycyl radical-based enzymatic pathways[15–17]. A clearer view of molecular transport would enlighten the design of more efficient systems, or even novel compartments for the biosynthesis of drugs, therapeutics, and biofuels[18–20].

Author Contributions

SC and JP conducted the simulations and analyzed the results. TOY, KNK and TAB conceived the study and interpreted the findings. SC, JP, and TOY wrote the manuscript.

Acknowledgements

The authors thank Cameron Mura for a critical reading of the manuscript and members of the Andrew McCammon laboratory for early discussions. This work was supported by NIH grant R01AI081146 (TOY and TAB) and NIH grant GM097200 (KNH). SC was supported by the CBI NIH training grant (T32-GM008496) and by the UCLA Graduate Division. Computations from this work were run and analyzed on the UCLA IDRE Hoffman2 cluster, the Keeneland and Stampede supercomputers through the XSEDE program (TG-CHE040013N), and the UCLA-DOE Cassini cluster, supported by the BER program of the DOE Office of Science under award # DE-FC02-02ER63421.

References

1. Doyle, D. A. *et al.* The Structure of the Potassium Channel: Molecular Basis of K+ Conduction and Selectivity. *Science* **280,** 69–77 (1998).
2. MacKinnon, R. Potassium channels. *FEBS Lett.* **555,** 62–65 (2003).
3. Chowdhury, C., Sinha, S., Chun, S., Yeates, T. O. & Bobik, T. a. Diverse Bacterial Microcompartment Organelles. *Microbiol. Mol. Biol. Rev.* **78,** 438–468 (2014).
4. Yeates, T. O., Kerfeld, C. a, Heinhorst, S., Cannon, G. C. & Shively, J. M. Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat. Rev. Microbiol.* **6,** 681–91 (2008).
5. Cheng, S., Liu, Y., Crowley, C. S., Yeates, T. O. & Bobik, T. a. Bacterial microcompartments: their properties and paradoxes. *Bioessays* **30,** 1084–95 (2008).
6. Crowley, C. S. *et al.* Structural insight into the mechanisms of transport across the Salmonella enterica Pdu microcompartment shell. *J. Biol. Chem.* **285,** 37838–46 (2010).
7. Chowdhury, C. *et al.* Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proc. Natl. Acad. Sci.* **112,** 2990–2995 (2015).
8. Sampson, E. M. & Bobik, T. a. Microcompartments for B12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. *J. Bacteriol.* **190,** 2966–71 (2008).
9. Havemann, G. D. & Bobik, T. A. Protein Content of Polyhedral Organelles Involved in Coenzyme B 12 -Dependent Degradation of 1 , 2-Propanediol in Salmonella enterica Serovar Typhimurium LT2 †. *J. Bacteriol.* **185,** 5086–5095 (2003).
10. Iannuzzi, M., Laio, A. & Parrinello, M. Efficient Exploration of Reactive Potential Energy Surfaces Using Car-Parrinello Molecular Dynamics. *Phys. Rev. Lett.* **90,** 238302 (2003).
11. Sugita, Y., Kitao, A. & Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **113,** 6042 (2000).
12. Cavalli, A., Spitaleri, A., Saladino, G. & Gervasio, F. L. Investigating Drug − Target Association and Dissociation Mechanisms Using Metadynamics-Based Algorithms. *Acc. Chem. Res.* **48,** 277–285 (2015).
13. Dryden, K. a, Crowley, C. S., Tanaka, S., Yeates, T. O. & Yeager, M. Two-dimensional crystals of carboxysome shell proteins recapitulate the hexagonal packing of three-dimensional crystals. *Protein Sci.* **18,** 2629–35 (2009).
14. Bauer, W. R. & Nadler, W. Molecular transport through channels and pores : Effects of in-channel interactions and blocking. *Proc. Natl. Acad. Sci.* **103,** 1–6 (2006).
15. Kerfeld, C. a *et al.* Protein structures forming the shell of primitive bacterial organelles. *Science* **309,** 936–8 (2005).
16. Tanaka, S., Sawaya, M. R. & Yeates, T. O. Structure and mechanisms of a protein-based organelle in Escherichia coli. *Science* **327,** 81–4 (2010).
17. Jorda, J., Lopez, D., Wheatley, N. M. & Yeates, T. O. Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci.* **22,** 179–95 (2013).
18. Frank, S., Lawrence, A. D., Prentice, M. B. & Warren, M. J. Bacterial microcompartments moving into a synthetic biological world. *J. Biotechnol.* **163,** 273–9 (2013).
19. Kim, E. Y. & Tullman-Ercek, D. Engineering nanoscale protein compartments for synthetic organelles. *Curr. Opin. Biotechnol.* **24,** 627–32 (2013).
20. Cai, F., Sutter, M., Bernstein, S. L., Kinney, J. N. & Kerfeld, C. a. Engineering Bacterial Microcompartment Shells: Chimeric Shell Proteins and Chimeric Carboxysome Shells. *ACS Synth. Biol.* (2014). doi:10.1021/sb500226j

Supporting Material for:

Molecular Dynamics Simulations of Selective Metabolite Transport Across the Propanediol Bacterial Microcompartment Shell

S. Chun, J. Park,T. A. Bobik, K. N. Houk, T. O. Yeates

I. Permeability through a pore in the presence of a free energy barrier

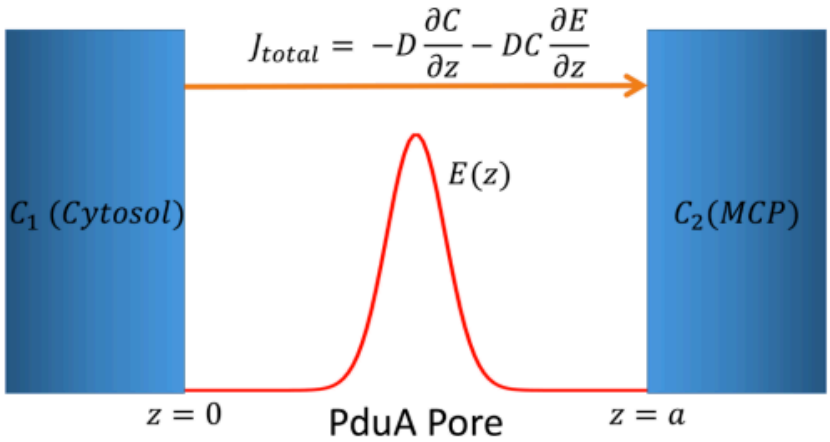


Figure S 1: Mathematical description of the selective permeation through the PduA pore. We assume the initial substrate or the intermediate has constant concentrations at both the outside (C1) and the inside (C2) of the bacterial microcompartment (MCP). Following Fick's law of diffusion, the total flux (J_{total}) is the sum of the contributions from the concentration gradient ($-D\frac{\partial C}{\partial z}$) and the external energy gradient ($-DC\frac{\partial E}{\partial z}$).

We consider the diffusion of a small molecule through the PduA pore where Fick's laws of diffusion are obeyed. In this system, 1-D diffusion of the small molecule along an axis (z) through the PduA pore is considered (Figure S1). By Fick's 1st law, in the absence of an energy term, diffusion in a concentration gradient gives rise to a net flux, J:

$$J_1 = -D\frac{\partial C}{\partial z}$$

, where D is the diffusion coefficient. In the presence of an external energy term, E, an additional transport velocity is caused by the external force:

$$v_E = -\frac{1}{f}\frac{\partial E}{\partial z}$$

, where f is the frictional coefficient and is inversely related to D by the Einstein-Smoluchowski relation ($D = k_B T/f$), such that

$$v_E = -D\frac{\partial E}{\partial z}$$

, for E expressed in units of k_BT. The transport velocity due to the energy term gives rise to a flux,

$$J_2 = v_E C = -DC\frac{\partial E}{\partial z}.$$

Thus, in the case of molecular diffusion on a non-uniform energy landscape, the total flux can be written as

$$J_{total} = J_1 + J_2 = -D\frac{\partial C}{\partial z} - DC\frac{\partial E}{\partial z}$$

, which is one form of the Smoluchowski equation for diffusion across an energy barrier[1].

We consider the relative permeation at a steady state, where the two substrates are in the same constant concentrations at both sides of the PduA pore. This is a hypothetical situation where both the initial and the intermediate substrates are consumed at the same rate inside the MCP. At steady state, the total flux is constant over time ($\frac{\partial J_{total}}{\partial t} = 0$) and we can consider the spatial dependence of the flux. We are left to solve an ordinary differential equation of the form

$$-DC' - DCE' = c$$

, where the primes indicate the derivative with respect to position.

Let b stand in for the constant $-c/D$, so

$$C' + CE' = b.$$

For the one-dimensional system considered along z, where $z = 0$ on the lumenal side of the bacterial microcompartment pore, the general solution to this first-order linear differential equation is

$$C(z) = be^{-E(z)}\int_0^z e^{E(z)}dz.$$

[Note that the coordinate choice for z=0 here differs from the main text, where z=0 is chosen near the point of narrowest restriction in the pore.] We drop the additive constant from the integrated expression under the assumption that $C(0) = 0$. Then, expressing the permeability p as a ratio of flux J to concentration difference ΔC from positions z=0 to z=a,

$$p = \frac{J}{C(a)}$$

Noting from substitutions above that the constant flux is equal to $-bD$, substituting the equation for C into the equation above and taking $E(a)$ to be zero, gives the following expression for permeability

$$p = \frac{-D}{\int_0^a e^{E(z)}dz}$$

A similar method is used by Bauer and Nadler (2006) to derive flux of particles that are non-interacting with a channel, interacting with a channel, or blocked from passage through a channel[2].

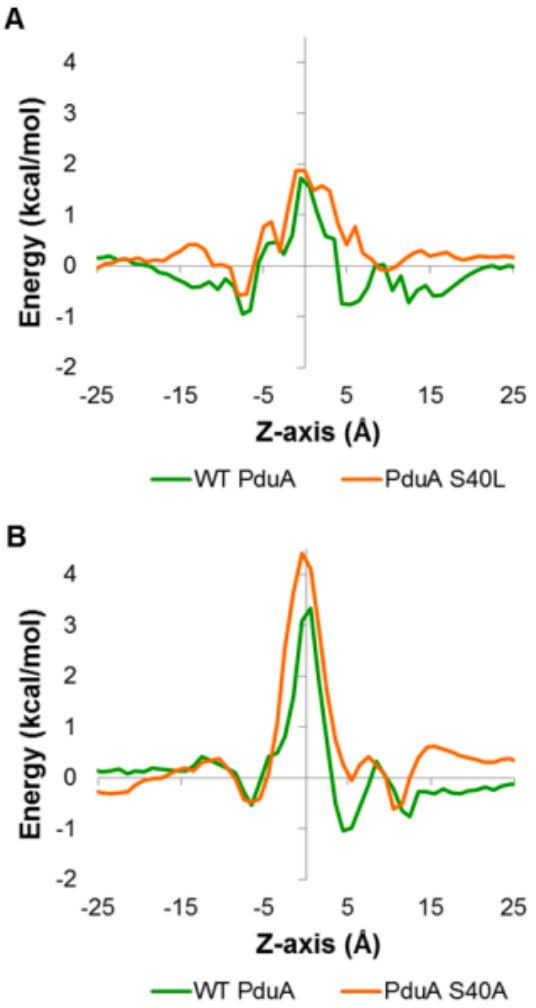
Thus, the permeabilities of the small molecules 1,2-propanediol and propionaldehyde through the PduA pore are proportional to expressions of the form $\frac{1}{\int_0^a e^{E(z)}dz}$, where the energy profiles are different for the

two molecules, and those energy profiles are obtained in the present work from MD-based enhanced sampling simulations.

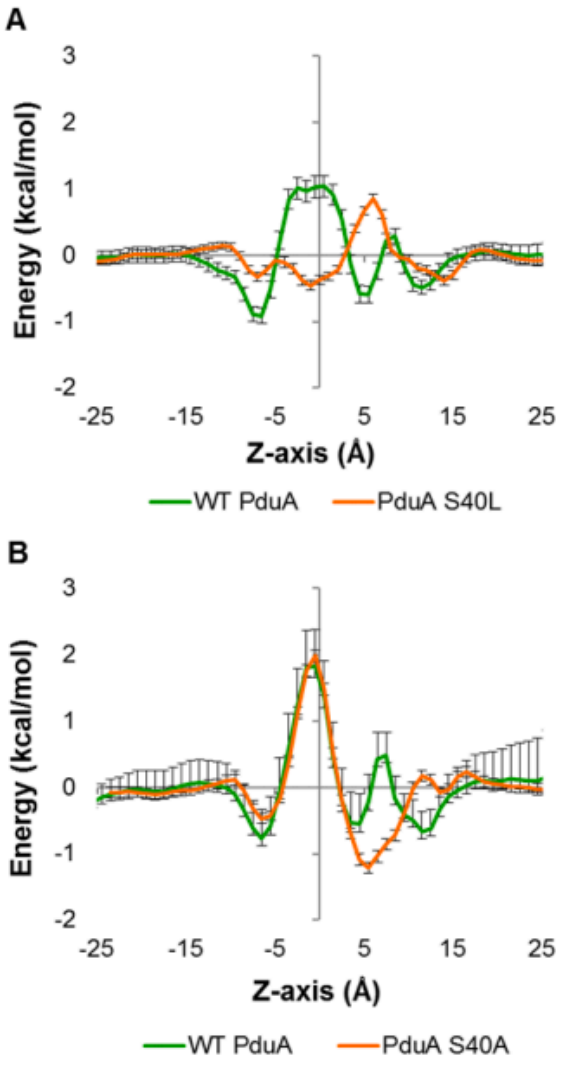
II. Free energy profiles through the pores of mutant PduA shell proteins

Motivated by the insight gained from our molecular dynamics comparison of propanediol vs propionaldehyde transport through the pore of PduA (see main text), we considered whether recent data on PduA mutants might be similarly analyzed. A recent experimental study concluded that a serine to alanine mutant (S40A) in PduA increased the leakage of propionaldehyde out of the microcompartment, and that a serine to leucine mutant (S40L) impeded the influx of propionaldehyde[3].

In the free energy profiles of the PduA pore mutants (shown below), there were only small differences compared to WT PduA with the corresponding small molecules, whether simulated by metadynamics or REUS. Indeed, the small observed differences in propanediol uptake and propionaldehyde efflux in *in vivo* experiments comparing WT PduA and PduA pore mutants anticipated the difficulty of interpreting MD simulations of these particular cases because the free energy differences would likely be within the uncertainty range of MD[3]. Given the general lack of consistency between the results from the two approaches – metadynamics, and REUS – firm conclusions could not be reached regarding the behavior of these PduA pore mutants.



Extended Data Figure 1. META PMF profiles of small molecules through mutant PduA pores. (A) PMF profiles of 1,2-propanediol through the hexamer pore in WT PduA (green) and PduA S40L (orange). (B) PMF profiles of propionaldehyde through the hexamer pore in WT PduA (green) and PduA S40A (orange).



Extended Data Figure 2. REUS PMF profiles of small molecules through mutant PduA pores. (A) PMF profiles of 1,2-propanediol through the hexamer pore in WT PduA (green) and PduA S40L (orange). (B) PMF profiles of propionaldehyde through the hexamer pore in WT PduA (green) and PduA S40A (orange).

III. Molecular mechanics parameterizations for MD simulations

The protein coordinates for PduA are from *Salmonella enterica* serovar Typhimurium LT2, PDB ID 3NGK. An infinite 2-D layer of PduA hexamer proteins was effectively constructed via periodic boundary conditions, using a hexagonal prism solvation box with ~20Å between two PduA layers. The system was protonated at pH 7.0 and solvated with approximately 40,000 TIP3P[4] waters. Na+ or Cl- ions were added to neutralize the charge on the PduA hexamers. The Amber FF99SB force-field[5] was used to describe the energetics of the protein. For the parameterizations of the small molecules, we followed the recommended procedure for General Amber Force Field (GAFF), as described in Ref 7.[7] The geometries of the small molecules were optimized with at the B3LYP/6-31G(d) level of theory using the Gaussian 09 code.[6] The covalent energy parameters (bond, angle, dihedral, improper) as well as Lennard-Jones radii of the constituent atoms were parametrized using AmberTools14. Atomic partial charges were assigned via the Restrained Electrostatic Potential (RESP) method[8]. A single point energy was computed at the HF/6-31G(d)//B3LYP/6-31G(d) level of theory. We used Antechamber software that is included in the AmberTools14 to deduce atomic partial charges that best fit the spatial charge distribution from the single point energies computed at the HF/6-31G(d) level theory.

Each solvated system was (i) energy-minimized over 500 steps with no restraints imposed to Cα atoms (ii) heated from 0 to 300 K over 1 ns of dynamics, (iii) equilibrated in the NVT ensemble, and then (iv) propagated as a production run in the NPT ensemble for 10 ns. For the metadynamics simulations, 9 replicas of each system were simulated in parallel, following the general approach of Limongelli et al. (2013)[9]; this number of replicas was chosen to maximize computation speed and efficiency. The pressure of the solvated system was controlled by scaling the three periodic unit cell dimensions independently. The well-tempered metadynamics algorithm was used in order to achieve smoothly converging free energy profiles[10,11].

IV. Methods for conventional molecular dynamics simulations of small molecules constrained at WT PduA pore

The systems for conventional MD (cMD) were set up identical to those for Metadynamics and REUS (as described in the main text). PduA was simulated as an infinite 2-D layer was effectively imposed using a periodic solvation box in the shape of a hexagonal prism. The system was protonated at pH 7.0 and solvated with a TIP3P water model. The AmberFF99SB force field was used to describe the energetics of the protein. The geometries of the small molecules were optimized in B3LYP/6-31G(d) level theory using Gaussian 09 software. The General Amber Force Field (GAFF) was used to parameterize covalent interactions of the small molecules. Atomic partial charges of the small molecules were assigned using the Restrained Electrostatic Potential (RESP) method. A solvated system contains a PduA hexamer and a small molecule, 1,2-propanediol or propionaldehyde. The solvated systems were energy minimized in 500 steps, heated from 0K to 300K for 1 ns, equilibrated in NVT ensemble for 1ns, then in NPT ensemble for 10 ns.

Production MD was extended for 100ns (for 3 replicas) with the small molecules restrained (spring constant 2.0 kcal·mol^{-1}·Å^2) at z=0.5Å (width 1Å) and $0 \leq r \leq 5$Å in the PduA pore. The cMD trajectories were analyzed using VMD and VMD Extension tool Hydrogen Bonds to calculate number of hydrogen bonds involving Ser40 residues and waters or small molecule 1,2-propanediol or propionaldehyde[12].

References

1. Ray, D. S. Notes on Brownian motion and related phenomena. *arXiv:physics/9903033 [physics.ed-ph]* (2008).

2. Bauer, W. R. & Nadler, W. Molecular transport through channels and pores : Effects of in-channel interactions and blocking. *Proc. Natl. Acad. Sci.* **103,** 1–6 (2006).

3. Chowdhury, C. *et al.* Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proc. Natl. Acad. Sci.* **112,** 2990–2995 (2015).

4. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926 (1983).

5. Hornak, V. *et al.* Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins Struct. Funct. Bioinforma.* **725,** 712–725 (2006).

6. Frisch, M. J. *et al.* Gaussian09. (2009).

7. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **25,** 1157–1174 (2004).

8. Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **97,** 10269–10280 (1993).

9. Limongelli, V., Bonomi, M. & Parrinello, M. Funnel metadynamics as accurate binding free-energy method. *Proc. Natl. Acad. Sci.* **110,** 6358–6363 (2013).

10. Barducci, A., Bussi, G. & Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **100,** 020603 (2008).

11. Bonomi, M. & Parrinello, M. Enhanced Sampling in the Well-Tempered Ensemble. *Phys. Rev. Lett.* **104,** 190601 (2010).

12. Humphrey, W., Dalke, A. & Schulten, K. VMD : Visual Molecular Dynamics. *J. Mol. Graph.* **7855,** 33–38 (1996).

Chapter 5

A study of Tandem-BMC domain proteins with ligand-gated pores

Introduction

In the Ethanolamine utilization (Eut) MCP, the tandem-BMC domain shell protein EutL has been solved in both "open" and "closed" conformations by X-ray crystallography (*E. coli*, PDB IDs: 3I87, 3I82 respectively)[27]. By soaking EutL crystals in ethanolamine or various metal ion solutions, Takenoya et al. found that Zn^{2+} bound to EutL in the open conformation[28]. Therefore, they concluded that binding of Zn^{2+} triggers significant conformational change that opens the central pore (12Å in diameter) of the EutL trimer assembly. However, the crystallization condition at 1M $ZnCl_2$ is not physiologically relevant. The large pore of EutL can be a possible route for passage of small molecules and larger co-factors. A quick APBS (A Poisson-Boltzmann Solver) surface electrostatics calculation shows a highly negatively charged surface surrounding and continuing through the large pore (Figure 5.1).

Here, I used two molecular dynamics simulation approaches, targeted MD and accelerated MD (reviewed by Schlick[29]), to understand this conformational shift.

Results

Conformational shifts (on the millisecond timescale) in proteins can be observed at timescales of hundreds of nanoseconds when using accelerated MD[30]. Accelerated MD adds a boost potential, $\Delta V(r)$), to the true energy potential, $V(r)$, giving a modified potential, $V^*(r)$,

$$V^*(r) = \begin{cases} V(r), V(r) \geq E \\ V(r) + \Delta V(r), V(r) < E \end{cases}$$

where E is the boost energy chosen for the system[31]. The added boost potential preserves the energy landscape by smoothly raising minima while keeping barriers constant:

$$\Delta V(r) = \frac{(E - V(r))^2}{\alpha + (E - V(r))}$$

where α is a tuning parameter that controls how close the boost potential will approach the limit of $\big(E - V(r)\big)$ (Figure 5.2). This biases the system to sample higher energy states, accelerating the sampling speed of the simulation. In addition, a dual boost on dihedral terms can improve the sampling in accelerated MD and can be implemented separately:

$$\Delta V(r) = \frac{(E_P - V(r))^2}{\alpha_P + (E_P - V(r))} + \frac{(E_D - V(r))^2}{\alpha_D + (E_D - V(r))}$$

where subscript P stands for Potential energy and subscript D stands for Dihedral energy.

Starting from either open or closed conformations for EutL, we prepared systems for accelerated MD. Protein coordinates were obtained from the RCSB PDB. Missing chains were modeled in using homology modeling in Schrodinger's Maestro program. The protein models were then protonated (because the structures were solved using X-ray crystallography) using the PDB2PQR server at pH 7.0 and solvated with TIP3P water model in an octahedral box with 12Å distance from box edge to protein. Systems were neutralized by adding counterions Na^+ or Cl^- using xleap software. Using the AMBER99SB force field, these periodic systems were first energy minimized for solvent, then all atoms, heated from 0K to 300K, equilibrated in NVT ensemble (constant molecule numbers, volume, and temperature) for 4ns, and equilibrated in NPT ensemble(constant molecule numbers, pressure, and temperature) for 50ns. Parameters for accelerated MD were extracted from the NPT simulation production:

	Open EutL	**Closed EutL**
Avg Dihedral Energy	6710.90655	6706.2157
Avg Total Potential Energy	-303913.502016667	-159588.5744
Total Atoms	54102	54102
Total Residues	648	648
Energy contribution per Degree of Freedom (3.5kcal/mol/residue)	2268	2268

Parameters were calculated as described the Amber12 Reference Manual:

alphaD= α*Energy contribution per degree of freedom (EC)
EthreshD= Avg Dihedral + EC
alphaP= α*total Atoms
EthreshP= Avg Eptot + alphaP

where α=0.175.

Accelerated MD implemented in Amber 12 was produced for 50ns with these parameters and with tuning parameters set to α, αx2, αx3, and αx4. No conversion from open to closed or closed to open conformations was observed even at the smallest value of tuning parameter, α. Results were analyzed with POVME analysis[32], which measures the change in volume of a pocket in a protein structure in an aligned MD simulation (Figure 5.3A). POVME analysis showed volume fluctuations in the pore region (Figure 5.3B,C). However, it would take a more detailed look specific to the EutL pore to see if the pore loops would be able to change.

True conformational change is defined by and would require the movement of the central pore loops in EutL (Figure 5.4A,B). Measuring the triangular area formed by the tyrosine residues at the tips of the central pore loops is a better way to gauge movement of the central pore loops than measuring the pore region volume. Triangle area analysis showed that the pore loops were in a stable conformation throughout all the accelerated MD simulations (Figure 5.4C,D).

A structure of EutL from *Clostridium perfringens* with bound ethanolamine molecules revealed that ethanolamine binds in a conserved hydrophobic pocket of the closed EutL structure that would normally be occupied by the central loop tyrosines in the open EutL structure (Figure 5.5A)[33]. Interestingly, in the open EutL MD simulations the tyrosines' preferred orientation is to be sandwiched by the other hydrophobic residues in the binding pocket (Figure 5.5B). The tyrosines cannot escape the hydrophobic pocket, which may be due to solvation in water, representing the natural state (Figure 5.5C).

From these results, we can conclude that accelerated MD could not overcome the favorable hydrophobic interactions of the central loop tyrosines in open or closed EutL

conformations. Also, binding of the substrate molecule ethanolamine to EutL in high concentrations of substrate obstructs the hydrophobic binding pocket where the central loop tyrosines would bind in the open EutL conformation.

A second and independent approach to the problem of observing the opening and closing of the EutL pore was to use a method called Targeted Molecular Dynamics. Targeted MD trajectory can evolve a system from one set of coordinates to another, showing the transitional states necessary for the change[34]. The method uses r.m.s.d. as a measure for determining the rate of change from the initial to the final states:

$$\rho = |\mathbf{x} - \mathbf{x}_f| = [\sum(x_i - x_{Fi})^2]^{1/2}$$

Targeted MD was implemented using NAMD 2.9 code with the AMBER99SB force field on systems of closed or open EutL, prepared as above. Molecular dynamics were simulated for 50ns each and r.m.s.d. trajectory followed the plots in Figure 5.6. Targeted MD pulled the 'open' conformation to the 'closed' conformation in two protocols: (1) by targeting all α-Carbons (Figure 6A) and (2) targeting all "heavy" (non-proton) atoms (Figure 5.6B) in the protein. However, the converged upon coordinates in either case did not fully exhibit the closed conformation in secondary structure or inter-loop contacts found in the crystal structure of the closed EutL coordinates (Figure 5.6C).

Discussion

In this work, I attempted to perturb the EutL system using accelerated MD and targeted MD but was not able to observe the path of conformational changes and loop rearrangements leading to closed or open pore structures. Thus, I concluded that the scale of the structural

rearrangements was too great for the molecular dynamics methods we used or would require much larger time scales for simulation. Further studies may include extending these molecular dynamics simulations to the hundreds or thousands of nanoseconds timescale in order to observe these events.

Figure 5.1. Surface electrostatic representation of EutL from *E. coli* in (A) closed (PDB ID: 3I87) and (B) open conformations (PDB ID: 3I82) generated by APBS calculator

A B

Figure 5.2. A schematic diagram of accelerated MD applied to a system with reaction coordinate x and energy V(x), re-produced from Hamelberg et al., 2004[31].

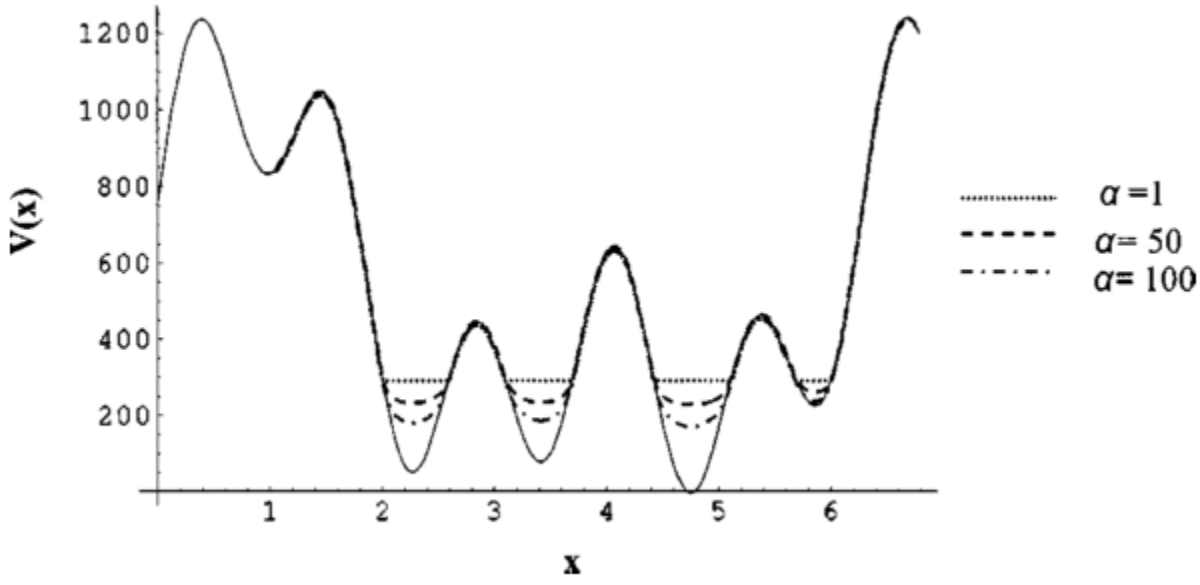
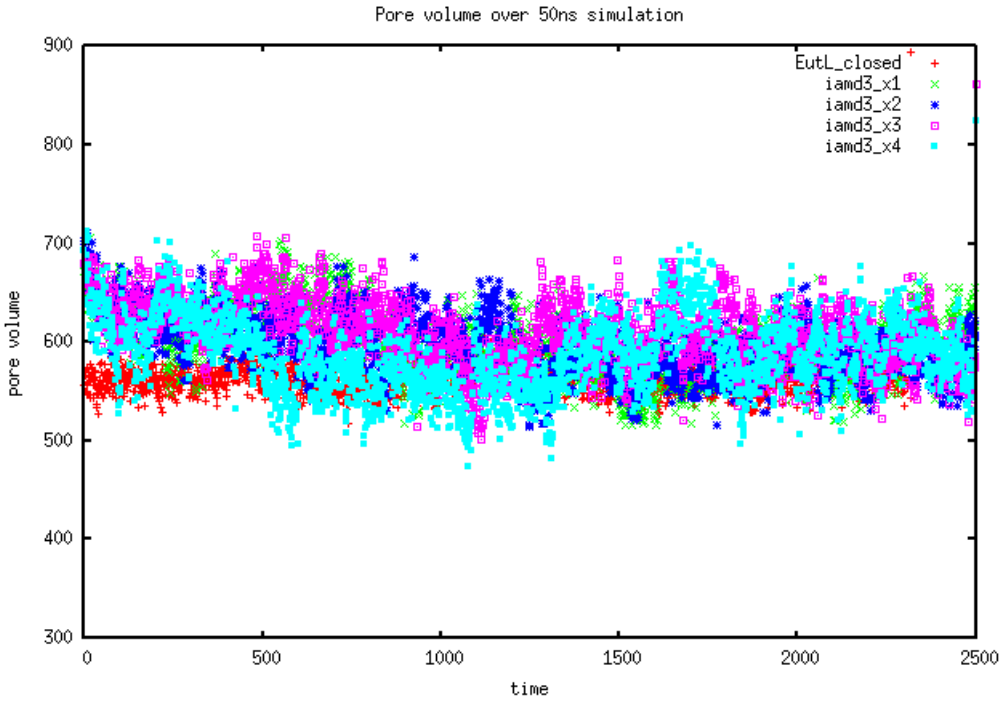


Figure 5.3. POVME analysis (A) POVME analysis measures the volume (pink spheres) in a pocket of a protein (the EutL pore region) over the course of an MD simulation. (B) POVME analysis of open EutL (iamd) at tuning parameters α, αx2, αx3, αx4. (C) POVME analysis of closed EutL (clamd) at tuning parameters α, αx2, αx3, αx4.

A

B



C

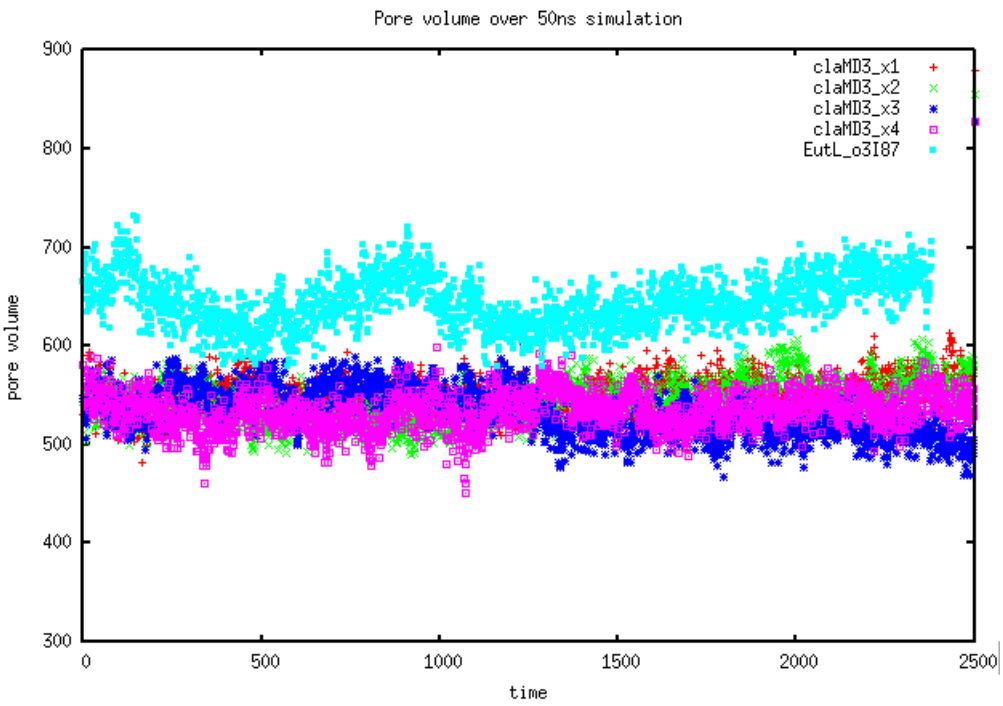
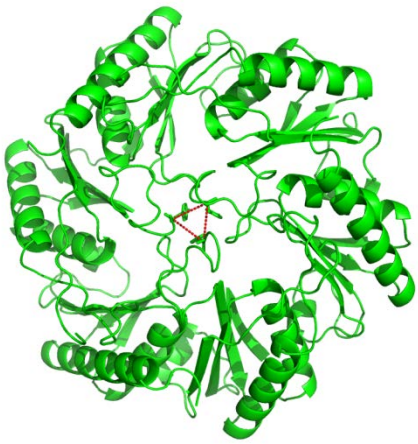
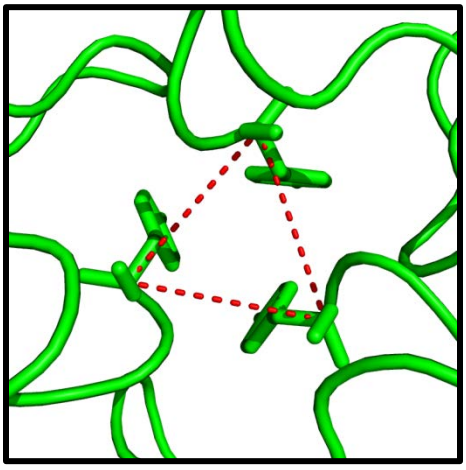

Pore volume over 50ns simulation
claMD3_x1
claMD3_x2
claMD3_x3
claMD3_x4
EutL_o3I87
pore volume
time
900
800
700
600
500
400
300
0
500
1000
1500
2000
2500

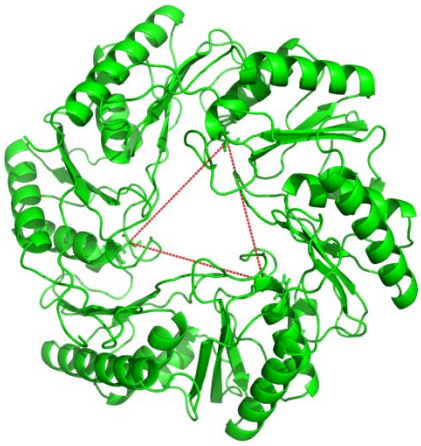
Figure 5.4. Triangle area analysis to monitor orientation and conformation of central pore loops in EutL in transition from (A) closed to (B) open pore states. Triangle area analysis for aMD of (C) open (iamd) and (D) closed EutL (clamd) at tuning parameters α, αx2, αx3, αx4.

A

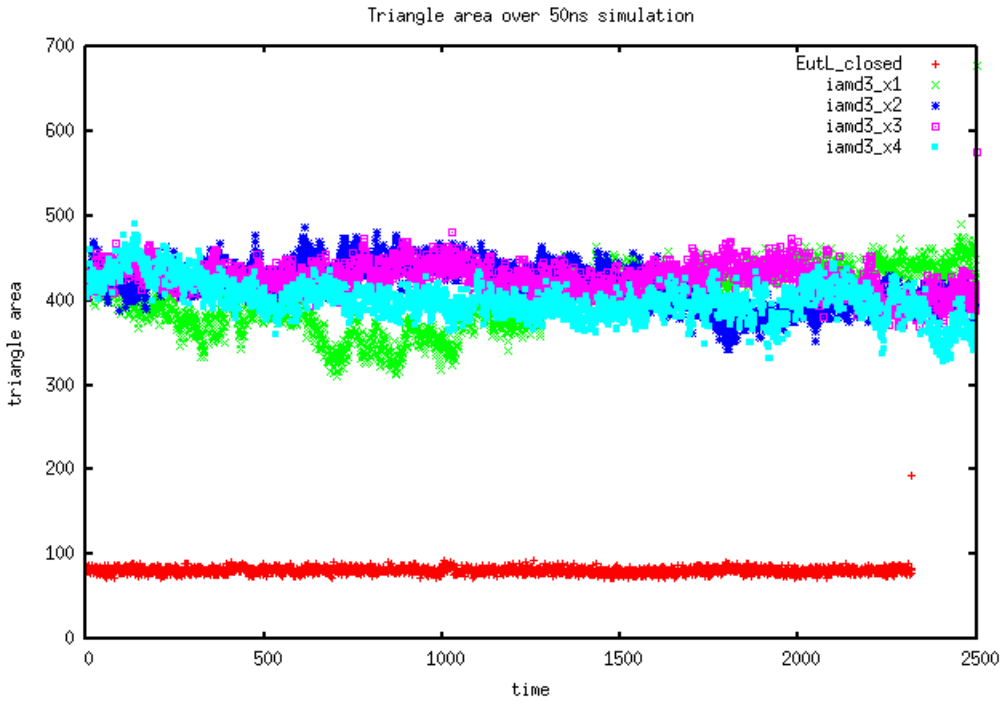




B



C


Triangle area over 50ns simulation
EutL_closed
iamd3_x1
iamd3_x2
iamd3_x3
iamd3_x4
triangle area
time
0
100
200
300
400
500
600
700
0
500
1000
1500
2000
2500

D

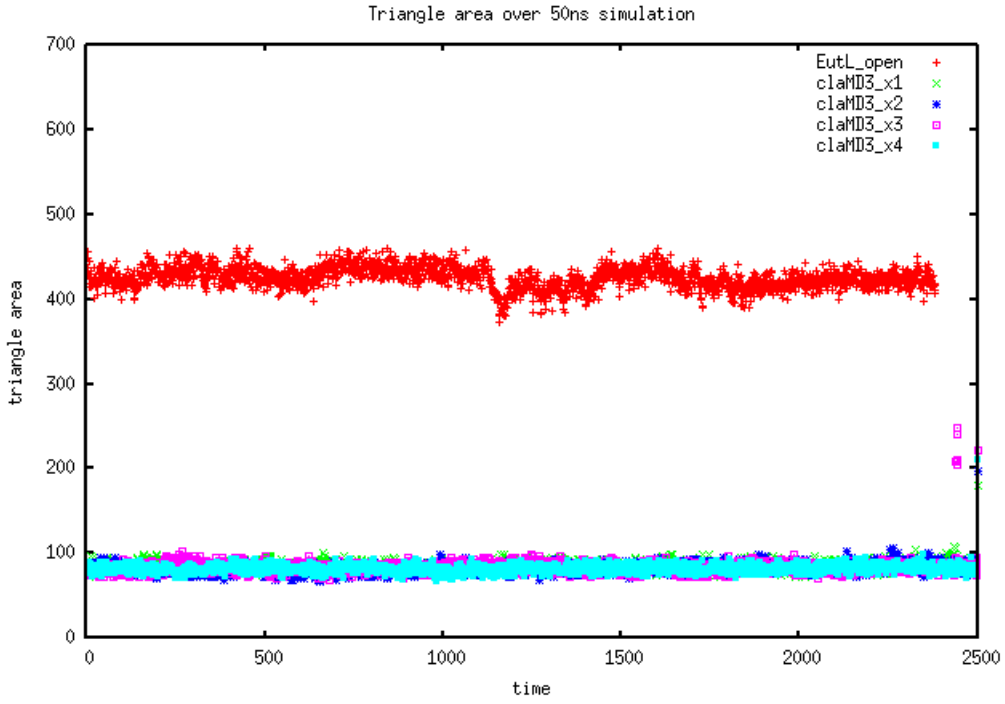
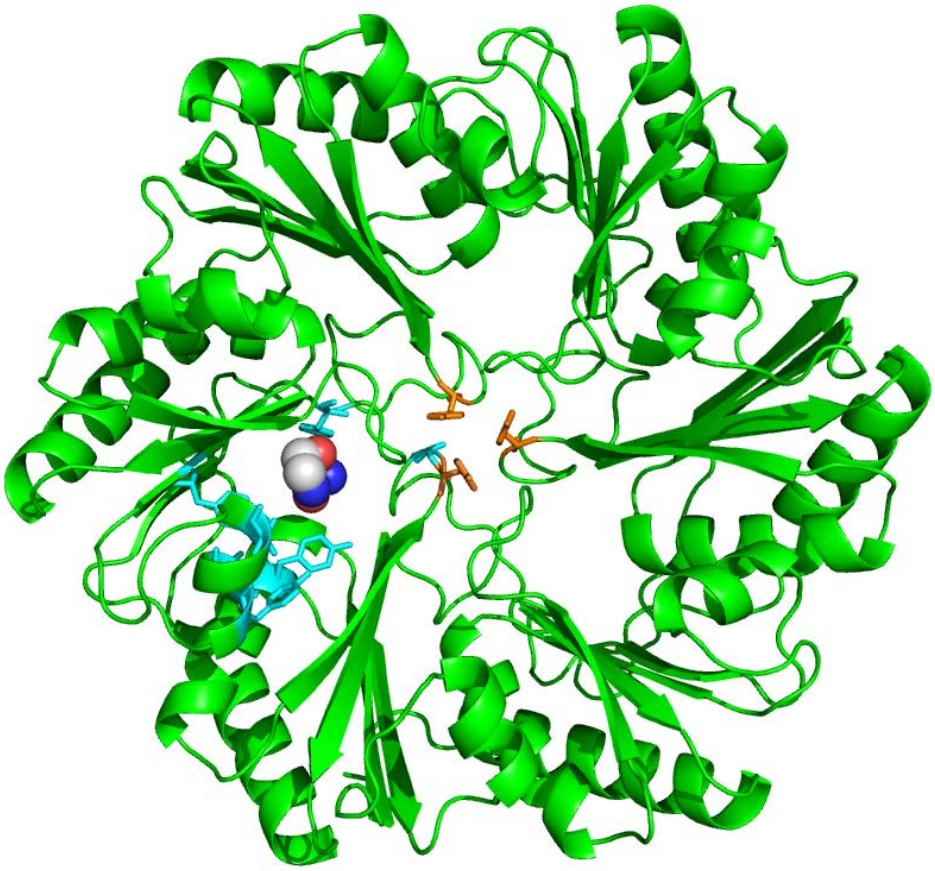

Triangle area over 50ns simulation
EutL_open
claMD3_x1
claMD3_x2
claMD3_x3
claMD3_x4
triangle area
time
0
100
200
300
400
500
600
700
0
500
1000
1500
2000
2500

Figure 5.5. (A) Ethanolamine obstructs access to the hydrophobic pocket for central tyrosines in *Clostridium perfringens* EutL. In the open conformation of EutL, (B) central pore loop tyrosines remained buried in a hydrophobic pocket and (C) do not escape the hydrophobic pocket during the course of MD simulation.

A



B

C

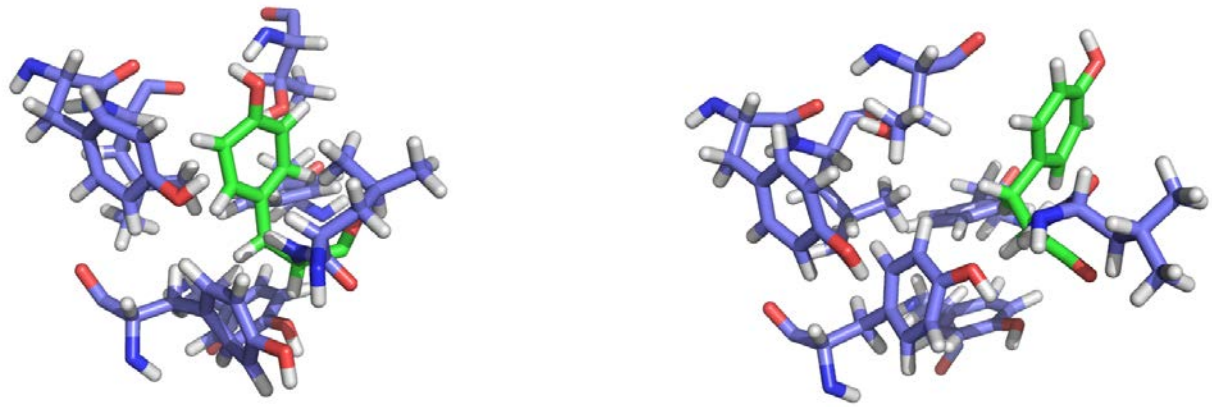
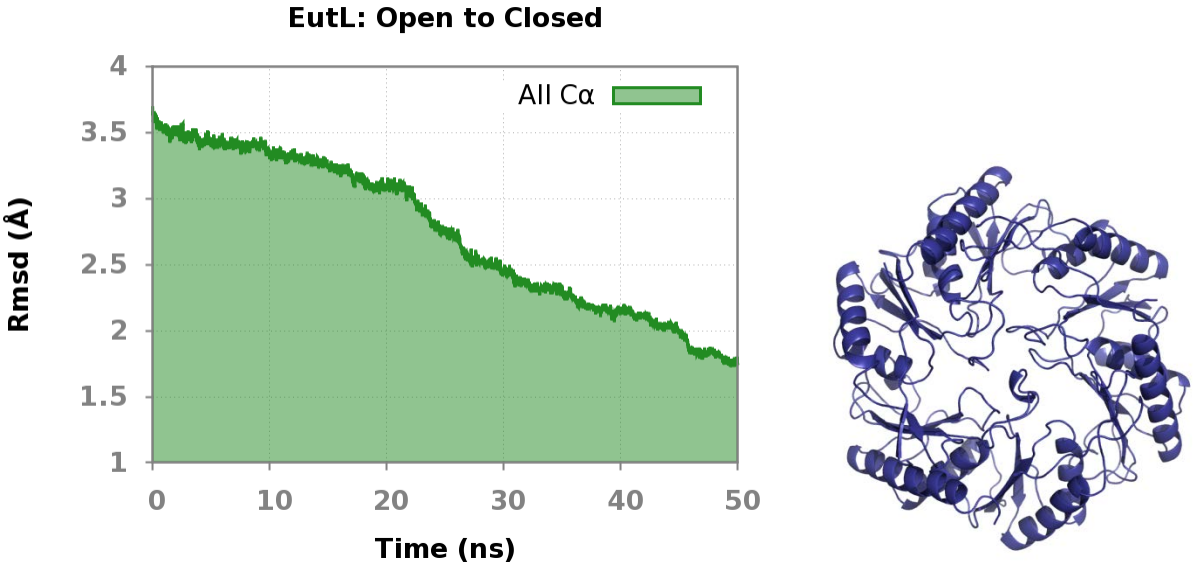
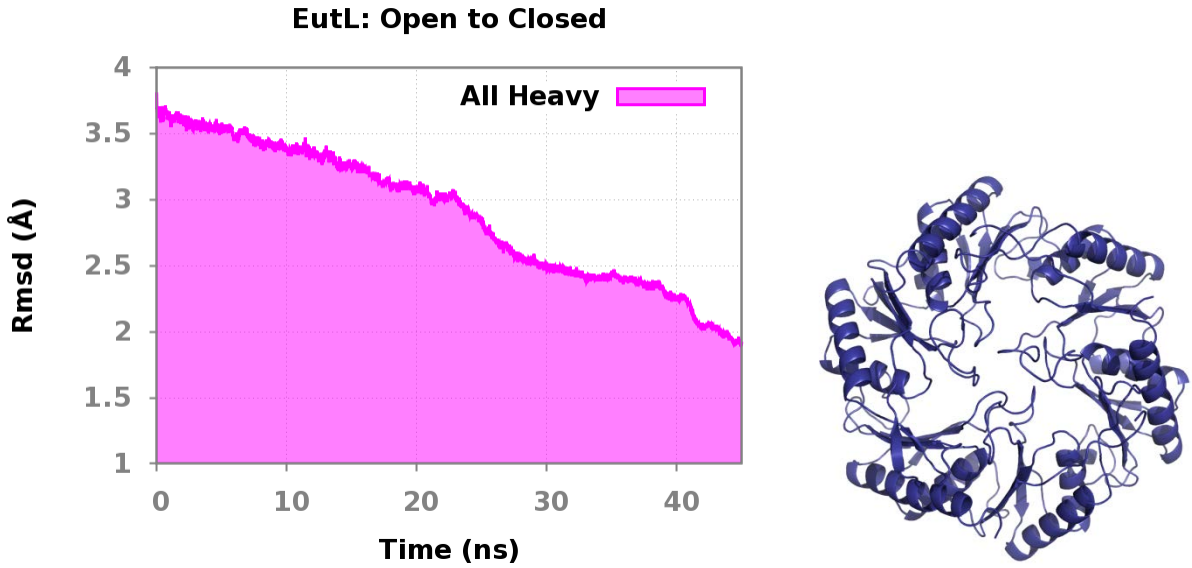


Figure 5.6. Targeted MD results from 50ns simulations of targeting EutL in the 'open' conformation to the 'closed' conformation (shown in C) using (A) only alpha-carbons or (B) all atoms.

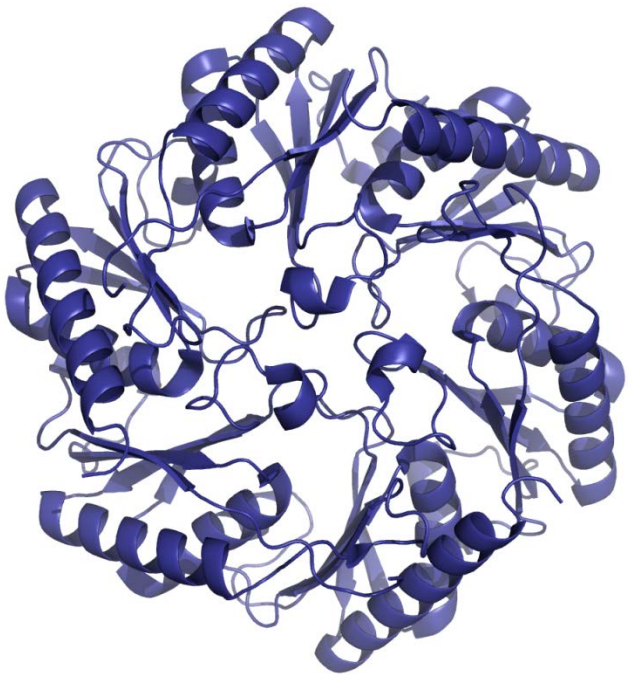
A



B


EutL: Open to Closed
All Heavy
Rmsd (Å)
Time (ns)
4
3.5
3
2.5
2
1.5
1
0
10
20
30
40

C



Chapter 6

Studies on PduB' to observe putative open pore conformation

Introduction

The PduB' protein from the Pdu MCP is a homolog of CsoS1D and EutL—tandem BMC domain proteins from the carboxysome and Eut MCPs, respectively. CsoS1D was solved in the open and closed configurations revealing a 15 Å radius pore through the center[35]. Similarly, a possibly gated pore was observed in EutL, suggesting that gated pore transport could be a shared mechanism in BMC shells[36]. We hypothesize that PduB' also has an open conformation though it has not yet been observed in crystal structures[37]. Our collaborators, Bobik et al., tried to destabilize the pore region. By making the D79A mutant of PduB' F188W, they observed a blue shift in tryptophan emission peak wavelength in an *in vitro* bulk fluorescence assay, which may suggest this mutant has an open pore (unpublished). In this chapter, I describe efforts to pursue the crystal structure of PduB' D79A and the stabilization of the PduB' open pore conformation by developing a ligand binding screen. These projects were aided by the contributions of undergraduate research fellow Qiaojun Chen and graduate rotation student Jeannette T. Bowler. Later, Bobik et al. asserted that the mutants A53F and A53R were better candidates for an open pore conformation in PduB'. With improved conditions for expressing PduB', I was able to solve the structure of PduB' A53F. However, the crystal structure did not exhibit the open pore conformation.

PduB' expression and purification testing for crystallization trials

The gene of interest, PduB' D79A, was cloned into an expression strain of *E. coli* bacteria BL21(DE3) (donated by Bobik et al.) with a 6xHistidine tag for pull-down purification at the N-terminus of the sequence. The bacteria was cultured with 34 μg/mL kanamycin (KAN) to select for the plasmid carrying the gene. The culture was grown in ZYP-5052 media for 36h, shaking at 20°C for autoinduction of protein expression, unless otherwise noted. Cells were harvested at 5000rpm for 10 minutes. Cell pellets were resuspended in lysis buffer before lysis by Emulsiflex or sonication. Lysis buffer was Buffer A (50mM TRIS pH 8.0, 150mM NaCl, 10mM Imidazole, 10% glycerol, 1mM DTT, plus additives as described below) complemented with protease inhibitors, 50 μg/mL DNase I, 50 μg/mL lysozyme, and 1 M $MgCl_2$ for lysis. Cell lysate was then pelleted at 17000xg for 30min. Clarified cell lysate was loaded on a pre-equilibrated HisTrap column and eluted with a step gradient of 10%, 30%, 50% and 100% Buffer B (Buffer A with 500mM Imidazole). PduB' D79A eluted at 30% Buffer B (or 150mM Imidazole). Eluted fractions and purified protein were tested by SDS-PAGE. Eluted fractions containing pure PduB' D79A were combined and dialyzed into Buffer C (30mM TRIS pH 8.0, 50mM NaCl, 5% glycerol, 1mM DTT). Purified protein was concentrated to 16.5mg/mL and set up for crystallization screens.

Initial attempts at purification demonstrated the insolubility of the PduB' F188W D79A and PduB' D79A mutants. Protein was trapped in the insoluble fraction of the lysate and could not be isolated by simple re-suspension of the lysate pellet in Buffer A (Figure 6.1). Once the insoluble protein was confirmed, 1L of PduB' F188W D79A was grown in autoinduction media for additive testing to solubilize the protein (Table 6.1). After 36 hours incubation at 20°C, the cells were spun down by centrifuging at 5,000 rpm for 10 min, and 0.8 g aliquots of cells were resuspended in 5 mL of Buffer A complemented with different additives. After addition of

lysozyme, the cells were lysed with rocking for 1.5 hours, and DNase 1 was added to reduce viscosity. 1 mL of the cell lysate was transferred to microcentrifuge tubes. The membranes and unlysed cells were pelleted by centrifuging at 20,000 xg for 20 min. The supernatants were carefully transferred to new tubes; the pellets were resuspended in 1 mL of Buffer A. To analyze the improvement of solubility by the additives, the supernatant and resuspended pellet for each additive were loaded side by side in SDS-PAGE gel. However no additive improved the solubility of the protein of interest. Bands at 25kD in all of the re-suspended pellets were fat compared to the extremely weak bands in the supernatants (Figure 6.2). In Figure 6.2, although lane 10 (200mM Arginine) and lane 12 (0.5% CHAPS) showed a slight increase in protein solubility, the amount of solubilized proteins was not enough to pursue protein purification and protein crystallization.

We carried out a second additive testing (Table 6.2), in which protein expression was induced by adding isopropyl β-D-thiogalactopyranoside (IPTG, 0.5mM final concentration) for 2 L of cells grown in Luria Bertani (LB) media, shaking at 37°C to an optical density at 600nm (OD_{600}) of 0.6 . Protein was expressed for at 30°C for 4 hours. The cells were spun down, and 0.3 g aliquot was suspended in 2 mL Buffer A with additives. The cells were lysed with rocking at 4°C for 1.5 hours, then incubated with DNase I, and rocked for 10 more min. 1 mL of cell lysate in the non-additive buffer and various additive containing buffers were transferred to microcentrifuge tubes. The pellets were collected at 17,000 xg for 20 min, and resuspended in 1 mL of Buffer A after separation from supernatant. The pellets and supernatants were subjected to SDS-PAGE for analysis of protein solubility. The fat bands in resuspended pellets showed the protein was still trapped in pellet fractions (Figure 6.3).

Unfortunately, none of the above additives were able to solubilize the protein. Thus, PduB' D79A without F188W mutation was used for analysis of protein expression and purification because success had been reported for the wildtype PduB' structure (Bobik et al., unpublished). However, PduB' D79A had proven insoluble by resuspension of the pellet of clarified lysate so the protein was still queued for additive testing (Table 6.3). The protein expression was induced by IPTG induction, cells were harvested and resuspended in Buffer A with the various additives, and analyzed by SDS-PAGE in the same manner as the second additivie testing of PduB' D79A F188W. We found that the combination of 200mM Arginine and 0.5% CHAPS buffer was the best additive to the lysis buffer for solubilization of PduB' D79A (Figure 6.4).

After the success of additive testing for solubilizing PduB' D79A, cells were grown up in 4x1 L of autoinduction media at 20°C for 36 hours. Protein was purified as recorded in the Methods section with 200mM Arginine and 0.5% CHAPS buffer complemented to Buffer A. The clarified supernatant was passed through 0.45 um filter and loaded on a pre-equilibrated HisTrap column. Then, the protein was eluted with a step gradient using 10%, 30%, 50% and 100% Buffer B (15mL per step for a column volume of 5mL). To resolve the sample, the eluted fractions were subjected to SDS-PAGE analysis. This allowed us to estimate the purity of protein and to identify the protein at the expected molecular weight. The expected molecular weight of the protein of interest is 24.66 kD. Referring to a band of molecular weight of 25 kD in a protein ladder, the expressed protein can be localized by comparing the intensity of bands at 25 kD in pellet and supernatant fractions. In SDS-PAGE analysis of PduB' D79A purification, the intensive bands in elution fractions showed that the purified proteins were eluted at fractions 13-19 (Figure 6.5). The identified protein fractions were combined and dialyzed into Buffer C. After

the protein was concentrated to 16.5 mg/mL, protein was aliquoted at 50μL aliquots. The aliquots were flash-frozen in liquid nitrogen (N_2) and stored at -80°C prior to crystallization trials.

For crystallization, 150μL of purified proteins was thawed under running cool water and then transferred into one tube. Any debris was pelleted out by centrifugation at 14000xg for 5 min at 4°C. The supernatant was carefully transferred to a new tube. A method of vapor diffusion by hanging drop was applied for crystallization conditions. Four crystal screens were used: Index, PACT, $AmSO_4$, JCSG+. Each screen contains 96 conditions that were further varied by 3 protein:well solution ratios. Altogether, this gives 1152 crystal conditions that were screened in this project. The trays were checked periodically with no crystal formation observed.

Insolubility is a major issue and obstacle to structural and functional studies of proteins. The observed insoluble protein expression in our project may be due to sequestration of our protein of interest within inclusion bodies, a home for misfolded proteins. In the additive tests for solubilizing our protein of interest, the highly intense bands in pellet fractions show both proteins (PduB' D79A F188W, PduB' D79A) are possibly sequestered in inclusion bodies. Soluble expression of protein can be achieved by reducing the growth rate and/or the rate of expression, such as lowering growth temperature between 20-30°C and expressing protein by autoinduction. Additive methods also can efficiently improve soluble recombinant proteins at the cell lysis stage[38]. Our analysis by additive testing show that PduB' D79A F188W is poorly soluble. However, since the mutation of F188W on PduB' was mainly for fluorescence emission assays and does not affect the crystal structure, PduB' D79A was used for the remaineder of the study. We successfully solubilized the mutant (PduB' D79A) by using additives (0.5% CHAPS,

200mM Arginine) for cell lysis. No crystal formation has been observed yet. Future directions are to continue checking the crystallization screens for protein crystals and to optimize any micro-crystal forming conditions. Optimized crystals can be screened by X-ray diffraction and solved by crystallography for the 3-D atomic coordinates. If PduB' D79A can be captured in the pore-open form, we may infer that the Pdu MCP, like other BMCs, has dynamically gated and large pores in the BMC shell. Such a finding can give us insight into the mechanisms of small molecule and co-factor transport across the BMC shell which is valuable to our understanding of BMCs and to engineering smart BMCs for the production of biofuels and drugs.

Screening ligands for PduB' open pore conformation

Several constructs of PduB' were designed for ligand screening: PduB' F188W in pET22b, PduB' F188W in DCLIC, full length PduB F225W in pET22b, and full length PduB F225 in DCLIC. DCLIC is a custom vector expressing a TEV-protease cleavable MBP GFP fusion protein to aid in expression and folding of proteins; vector was obtained from Feigon et al. In addition, all constructs also carried K65A and K170A (PduB'), K102A and K202A (full length PduB) mutations to disrupt edge binding of PduB trimers and increase protein solubility for purification and biochemistry. Constructs were cloned into expression strain of *E. coli* bacteria BL21(DE3) (donated by Bobik et al.) with a 6xHistidine tag for pull-down purification at the N-terminus of each construct as well as C-terminus on MBP GFP fusion protein. Protein expression was induced by IPTG in 2L cultures. Cells were harvested at 5000rpm for 10 minutes. Cell pellets were resuspended in lysis buffer before lysis by Emulsiflex or sonication. Lysis buffer was Buffer A (50mM TRIS pH 8.0, 150mM NaCl, 10mM Imidazole, 5% glycerol, 1mM DTT) complemented with protease inhibitors, 50 μg/mL DNase I, 50 μg/mL lysozyme,

and 1 M $MgCl_2$ for lysis. Cell lysate was then pelleted at 17000xg for 30min. Protein was purified by Ni^{2+} affinity chromatography and eluted with a gradient of Buffer B (Buffer A with 500mM Imidazole). Eluted fractions were analyzed by SDS-PAGE and dialyzed into Buffer C (30mM TRIS pH 8.0, 50mM NaCl, 5% glycerol, 1mM DTT). Only PduB' F188W MBP GFP fusion yielded highly soluble protein. MBP and GFP were removed by TEV cleavage and subsequent Ni^{2+} affinity chromatography for MBP and GFP removal. Purified PduB' F188W was isolated in the flow-through, concentrated to 8.5 mg/mL and stored at -80°C.

The ligand screen for stabilizing the open conformation of the hypothetical PduB' pore was designed by selecting 12 small molecules that were known cofactors or metabolites in the Pdu pathway: B_{12} (cyanocobalamin), NAD, ATP, acetyl-coA, 1,2-propanediol, propionaldehyde, 1-propanol; and divalent metals $ZnCl_2$, $CoCl_2$, $MgCl_2$, $NiCl_2$, and $FeCl_3$. In addition, 6 buffer conditions were included to vary salt (5mM, 50mM, 250mM NaCl), pH (7, 8, 9) and non-reducing environment (without 1mM DTT). Thus, the resulting ligand screen was comprised of a total of 72 conditions set up in a black opaque bottom 96-well plate. Each well contained 100uL of condition with 10mM PduB' F188W. Fluorescence emission of F188W was measured on a SpectraMax M5 plate reader with excitation at 295 nm and emission meaured from 320-390 nm (10 nm intervals). All measurements reported were measured at room temperature and reference subtracted where reference constituted the condition without protein added.

From our screen, we identified significant changes in the PduB' F188W emission spectrum for cyanocobalamin (1mM), 1,2-propanediol (1mM), $FeCl_3$ (1mM, 10mM), $ZnCl_2$ (1mM, 10mM), and $CoCl_2$ (1mM, 10mM) (Figure 6.6). The largest effect was observed for divalent metals $ZnCl_2$ (1mM, 10mM) and $CoCl_2$ (1mM, 10mM). These metals were further

tested in triplicate at 3 timepoints (15 min, 2h, 22h) (Figure 6.7). A marked decrease in fluorescence intensity may correspond to a conformational change, thus crystallization screens were set up for PduB' F188W with metals $ZnCl_2$ and $CoCl_2$. Crystal trays were checked periodically but no crystals were observed.

In this study, we identified a better approach to expressing and purifying PduB' F188W via an MBP GFP fusion recombinant protein. The MBP GFP fusion increased expression and purification of the protein construct by more than 100% (Figure 6.8-10). We also identified metals $ZnCl_2$ and $CoCl_2$ in shifting the tryptophan emission spectrum. However, the spectra do not resemble the typical tryptophan emission spectrum. A change in tryptophan emission should shift the peak of the emission spectrum but this is not observed in our experiments.

Future experiments toward observing the open pore conformation of PduB' include binding experiments with PduB' and full length PduB with edge mutations and hits from our screen, crystallization of PduB' and full length PduB with edge mutations with hits from our screen.

Crystal structure of PduB' A53F

The constructs PduB' A53F/R K65A K170A with or without F188W were pursued for crystallization trials upon recommendation from Bobik et al. that these constructs had normal growth and MCP formation but pure protein exhibited blue shift in Trp emission spectra. Another set of media and expression conditions was found to improve PduB' expression and purification: minimal media with 1mM IPTG for a 2h expression at 30°C.

Proteins were purified as described above with more than 10-fold increased yields per liter of culture. Crystallization trials were set up for PduB' F188W, PduB' A53R, PduB' A53F (all constructs contain double edge mutations K65A and K170A unless otherwise noted): Index, Wizard, $AmSO_4$, JCSG+. Only PduB' A53F gave crystal hits that were further optimized for increased crystal size and improved X-ray diffraction (Figure 6.11).

Several data sets were collected for PduB' A53F at the UCLA DOE X-ray facility and solved by molecular replacement using the solution of PduB' K65A K170A from Bobik et al. However, all molecular replacement and initial refinements have given closed pore PduB' structures. Mutations are well described by the electron density maps. Difference map peaks do not show significant rearrangement in the pore for any solutions. Theoretically, to be consistent with the Csos1D and EutL pores, I would expect a triangular pore of 10-15Å per edge which is created by the flipping of the central loops containing F188.

A data set for PduB' A53F was collected at the Advanced Photon Source Beamline I-24-C at Argonne National Laboratory. Data were consistent with data collected at UCLA. The reflections were indexed and scaled using XDS programs, solved with molecular replacement using 3 monomers of PduB' K65A K170A from Bobik et al. in the asymmetric unit. Data and refinement statistics are available in Table 6.4. The structure was solved to 2.3Å resolution in spacegroup P1 and was refined to R_{work}/R_{free} = 0.153/0.204. In the biological assembly of the trimer cartoon model of PduB' A53F, the A53F residues and F188 residues are highlighted in sphere representation (Figure 6.12). Like in the case of EutL, we expected that the F188 containing loops would rearrange by flipping into the large pockets containing A53F upon

increasing hydrophobicity, thereby lowering an energy barrier to flipping, through the A53 to F mutation. However, no opening of the PduB' pore was observed in the crystal structure.

Crystal trays of PduB' A53F F188W have also been set up and may produce crystal hits for X-ray structure determination. Future work on this project include Tryptophan emission spectroscopy assays comparing emission wavelength shifts for PduB' F188W, PduB' F188W A53F/R in the presence or absence of 1,2-propanediol to elucidate whether 1,2-propanediol can shift the equilibrium of open pore PduB' F188W A53F/R constructs toward closed pore conformations.

Table 6.1. Additives for protein purification test 1
1M trehalose
1M betaine
1M mannitol
200mM arginine
0.5% CHAPS
0.5M trehalose + 0.5M betaine
0.5M betaine + 0.5M mannitol
0.5M trehalose + 0.5M mannitol
0.5M trehalose + 0.5% CHAPS
0.5M betaine + 0.5% CHAPS
0.5M mannitol + 0.5% CHAPS

Table 6.2. Additives for protein purification test 2
1M trehalose
1M betaine
1M mannitol
200mM arginine
0.5% CHAPS
0.5M trehalose + 0.5M betaine
0.5M betaine + 0.5M mannitol

0.5M trehalose + 0.5M mannitol
0.5M trehalose + 0.5% CHAPS
0.5M betaine + 0.5% CHAPS
0.5M mannitol + 0.5% CHAPS
200mM arginine + 0.5% CHAPS
1M betaine + 0.6M mannitol

Table 6.3. Additives for protein purification test 3
1M trehalose
1M betaine
1M mannitol
200mM arginine
0.5% CHAPS
0.5M trehalose + 0.5M betaine
0.5M betaine + 0.5M mannitol
0.5M trehalose + 0.5M mannitol
0.5M trehalose + 0.5% CHAPS
0.5M betaine + 0.5% CHAPS
0.5M mannitol + 0.5% CHAPS
200mM arginine + 0.5% CHAPS
1M mannitol + 0.5% CHAPS

Figure 6.1. PduB' D79A F188W is mainly in the pellet fraction. Lanes: (1) resuspended pellet, (2) pellet of the resuspended pellet (3) supernatant of the resuspended pellet.

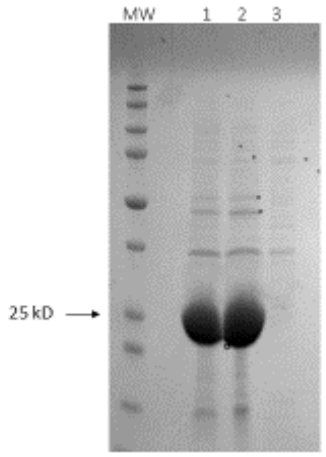


Figure 6.2. Additive test #1 for solubilizing PduB' D79A F188W. Lanes: (1) No additive-pellet, (2) No additive- supernatant, (3) 1M trehalose- pellet, (4) 1M trehalose- sup, (5) 1M betaine-pellet, (6) 1M betaine- sup, (7) 1M mannitol- pellet, (8) 1M mannitol- sup, (9) 200mM arginine- pellet, (10) 200mM arginine-sup, (11) 0.5% CHAPS- pellet, (12) 0.5% CHAPS- sup, (13) 0.5M trehalose + 0.5M betaine- pellet, (14) 0.5M trehalose + 0.5M betaine- sup, (15) 0.5M betaine + 0.5M mannitol- pellet, (16) 0.5M betaine + 0.5M mannitol- sup, (17) 0.5M trehalose + 0.5M mannitol – pellet, (18) 0.5M trehalose + 0.5M mannitol- sup, (19) 1M trehalose + 0.5% CHAPS- sup, (21) 1M betaine + 0.5% CHAPS- pellet, (22) 1M betaine + 0.5% CHAPS- sup, (23) 1M mannitol + 0.5% CHAPS- pellet, (24) 1M mannitol + 0.5% CHAPS- sup

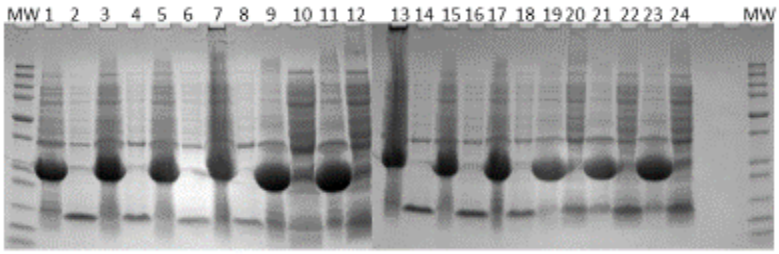


Figure 6.3. Additive test #2 for solubilizing PduB' D79A F188W. Lanes: (1) No additive- pellet, (2) No additive- supernatant, (3) 1M trehalose- pellet, (4) 1M trehalose- sup, (5) 1M betaine-pellet, (6) 1M betaine- sup, (7) 1M mannitol- pellet, (8) 1M mannitol- sup, (9) 200mM arginine- pellet, (10) 200mM arginine-sup, (11) 0.5% CHAPS- pellet, (12) 0.5% CHAPS- sup, (13) 0.5M trehalose + 0.5M betaine- pellet, (14) 0.5M trehalose + 0.5M betaine- sup, (15) 0.5M betaine + 0.5M mannitol- pellet, (16) 0.5M betaine + 0.5M mannitol- sup, (17) 0.5M trehalose + 0.5M mannitol – pellet, (18) 0.5M trehalose + 0.5M mannitol- sup, (19) 1M trehalose + 0.5% CHAPS- sup, (21) 1M betaine + 0.5% CHAPS- pellet, (22) 1M betaine + 0.5% CHAPS- sup, (23) 1M mannitol + 0.5% CHAPS- pellet, (24) 1M mannitol + 0.5% CHAPS- sup, (25) 200mM arginine + 0.5% CHAPS- pellet, (26) 200mM arginine + 0.5% CHAPS- sup, (27) 1M betaine + 0.6M mannitol- pellet, (28) 1M betaine + 0.6M mannitol- sup

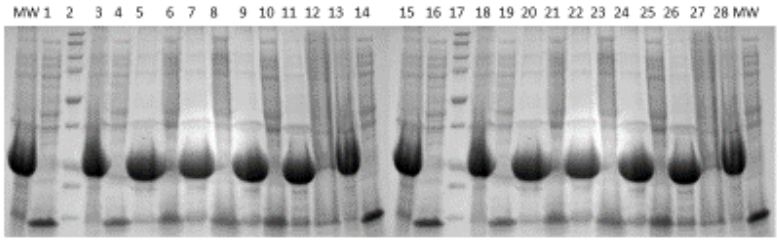


Figure 6.4. Additive test #3 for solubilizing PduB' D79A. Lanes: (1) No additive- pellet, (2) No additive- supernatant, (3) 1M trehalose- pellet, (4) 1M trehalose- sup, (5) 1M betaine-pellet, (6) 1M betaine- sup, (7) 1M mannitol- pellet, (8) 1M mannitol- sup, (9) 200mM arginine- pellet, (10) 200mM arginine-sup, (11) 0.5% CHAPS- pellet, (12) 0.5% CHAPS- sup, (13) 1M betaine + 0.6M mannitol- pellet, (14) 1M betaine + 0.6M mannitol- sup, (15) 0.5M trehalose + 0.5M mannitol- pellet, (16) 0.5M trehalose + 0.5M mannitol- sup, (17) 0.5M betaine + 0.5M mannitol – pellet, (18) 0.5M betaine + 0.5M mannitol- sup, (19) 0.5M trehalose + 0.5M mannitol- pellet, (20) 0.5M trehalose + 0.5M mannitol- sup, (21) 1M trehalose + 0.5% CHAPS- pellet, (22) 1M trehalose + 0.5% CHAPS- sup, (23) 1M betaine + 0.5% CHAPS- pellet, (24) 1M betaine + 0.5% CHAPS- sup, (25) 1M mannitol + 0.5% CHAPS- pellet, (26) 1M mannitol + 0.5% CHAPS- sup, (27) 200mM arginine + 0.5% CHAPS- pellet, (28) 200mM arginine + 0.5% CHAPS- sup

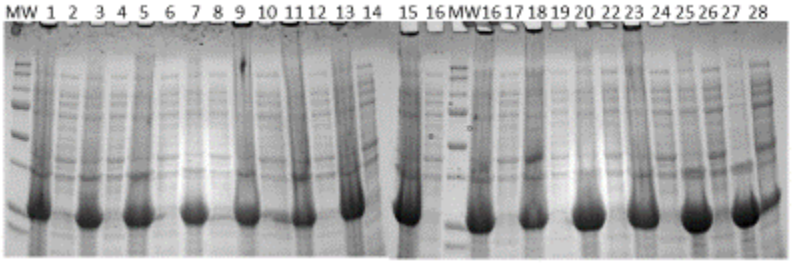


Figure 6.5. HisTrap purification of PduB' D79A. Lanes: (1) Pellet fraction, (2) clarified supernatant, (3) Flowthrough, (4) Wash, (5) Elution fraction #12, (6) #13, (7) #14, (8) #15, (9) #16, (10) #17, (11) #18, (12) #19, (13) #20, (14) #23

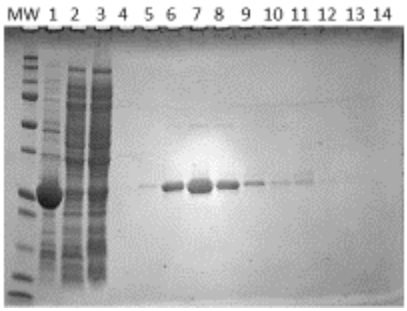
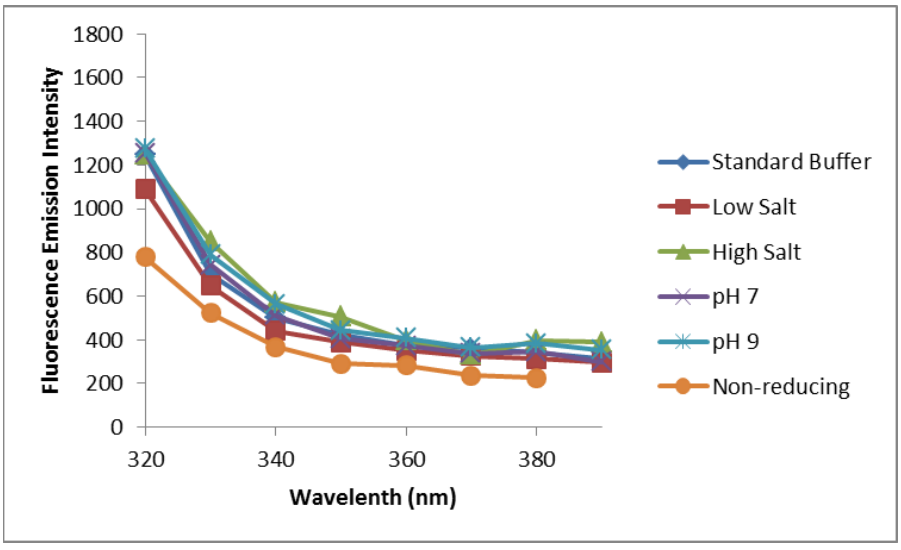
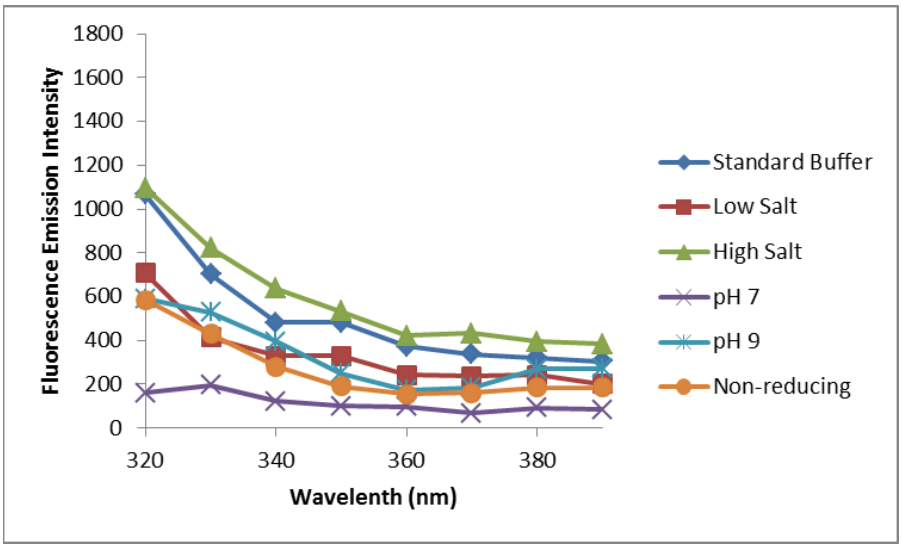


Figure 6.6. Tryptophan fluorescence emission profiles for PduB' F188W in various salt, pH, reducing conditions.

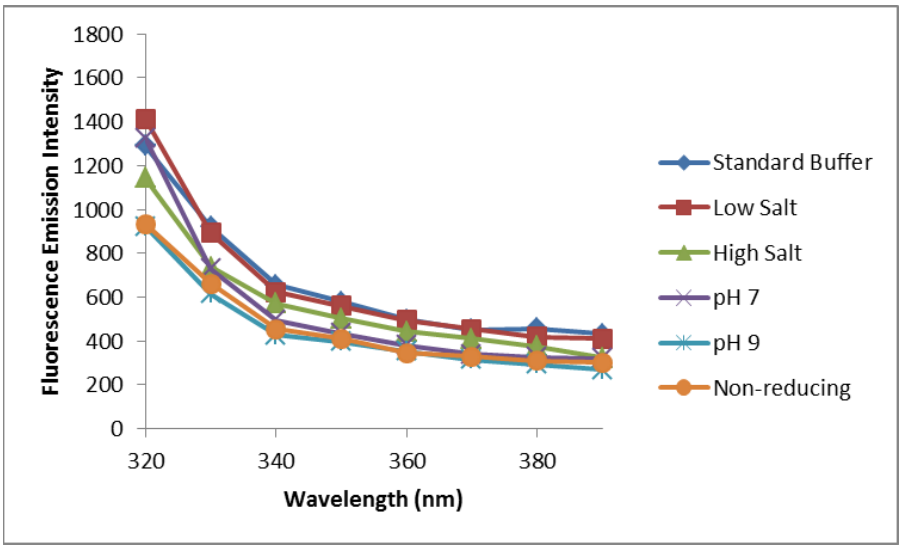
A. 10μM PduB' F188W



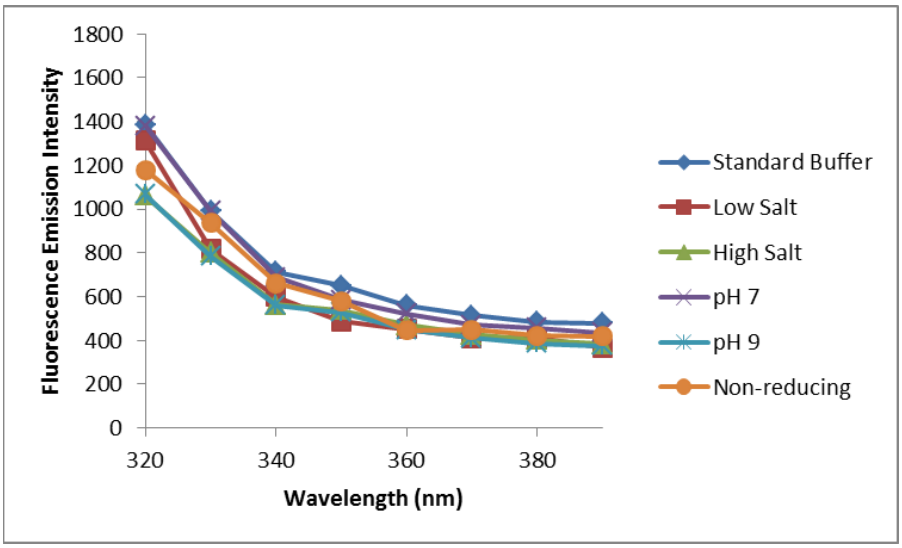
B. 10 µM PduB' F188W with 1mM cyanocobalamin



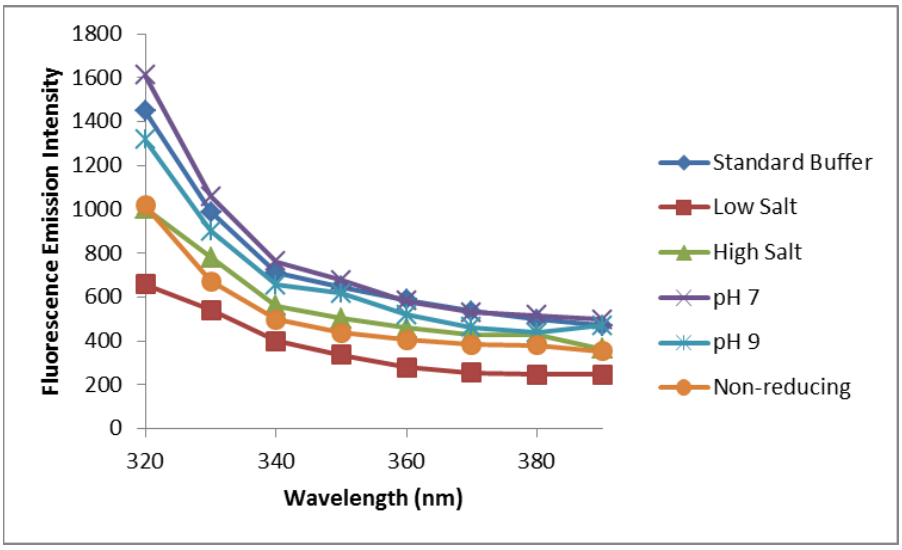
C. 10 µM PduB' F188W with 1mM NAD+



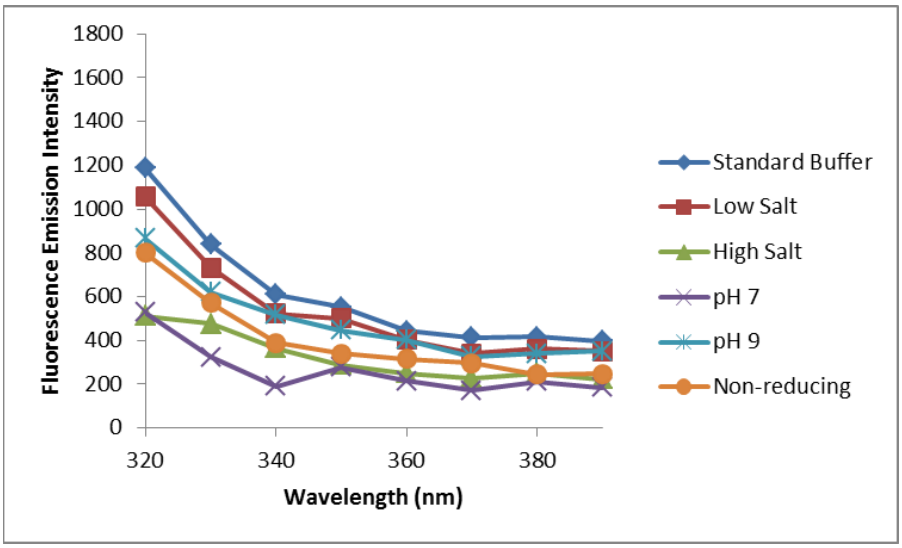
D. 10 µM PduB' F188W with 1mM ATP



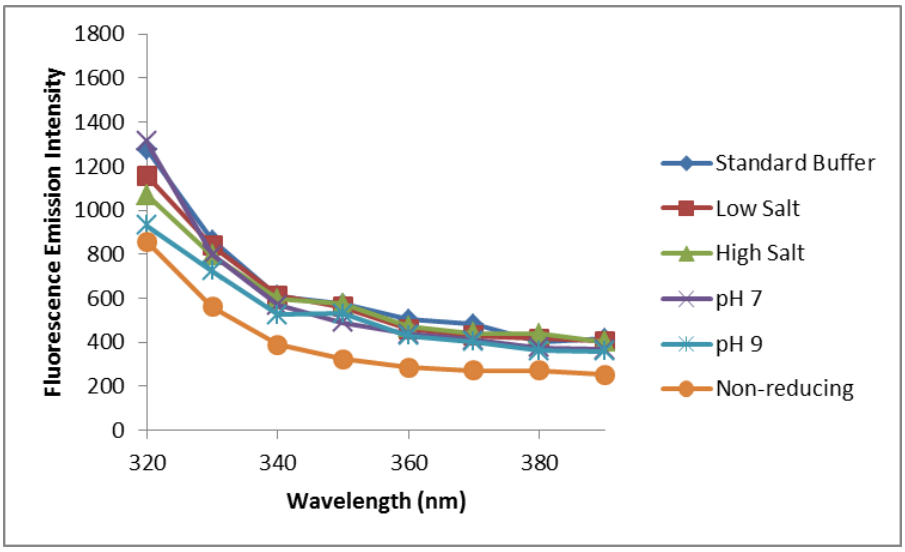
E. 10 µM PduB' F188W with 1mM acetyl-coA



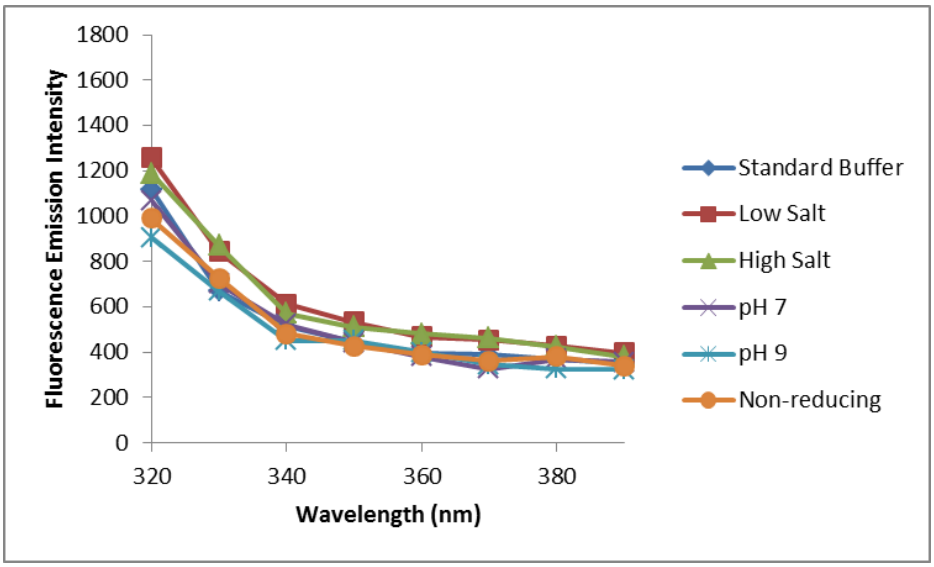
F. 10 µM PduB' F188W with 1mM 1,2-propanediol



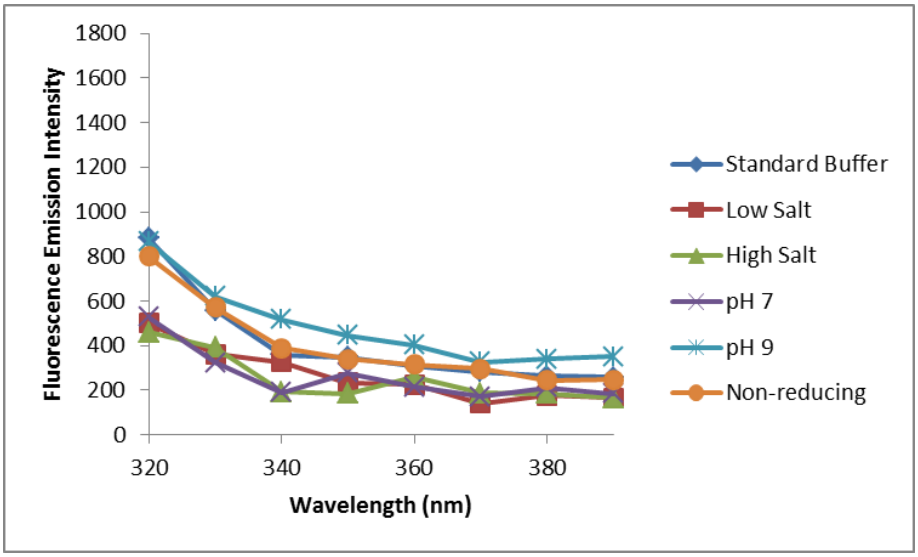
G. 10 µM PduB' F188W with 1mM propionaldehyde



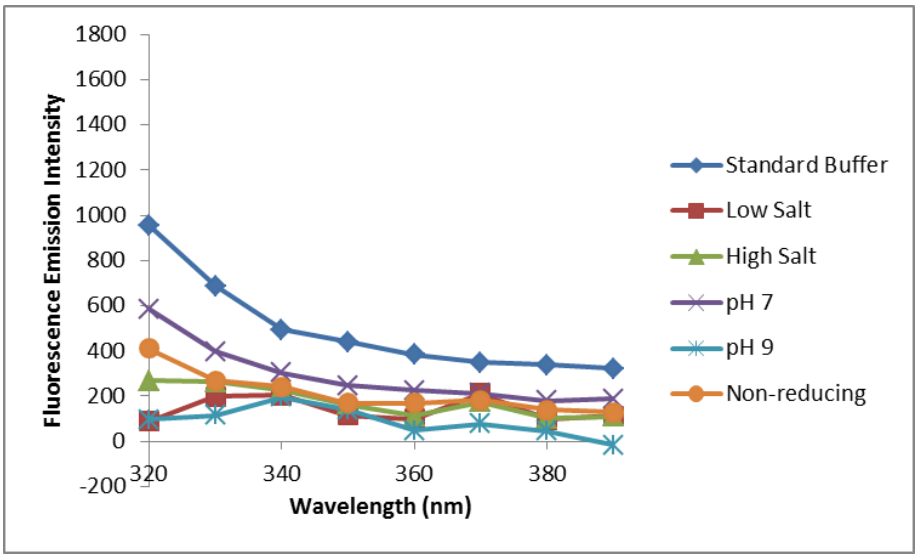
H. 10 µM PduB' F188W with 1mM 1-propanol



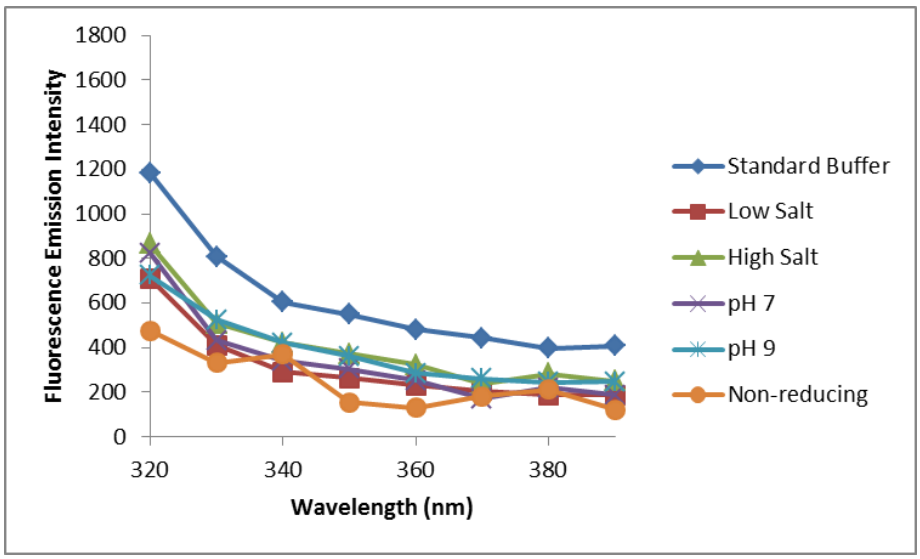
I. 10 µM PduB' F188W with 1mM $FeCl_3$



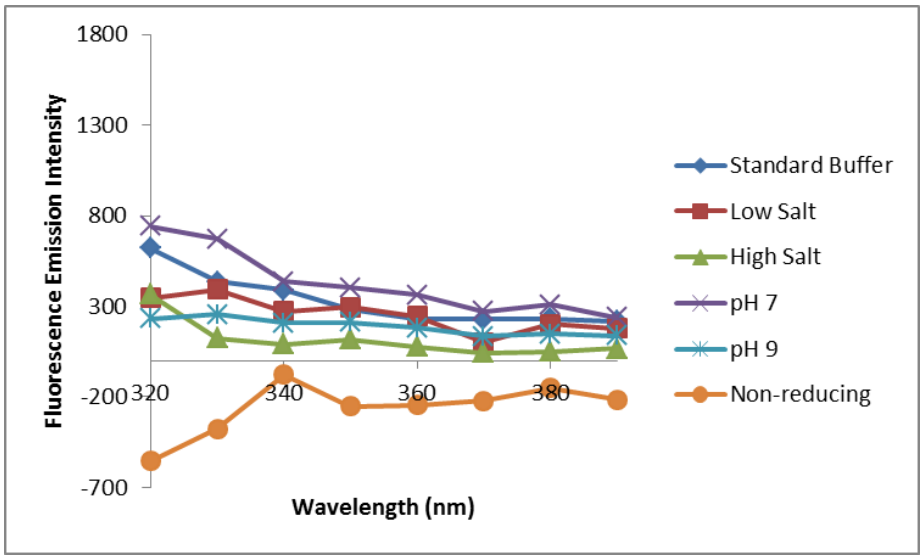
J. 10 µM PduB' F188W with 10mM $FeCl_3$



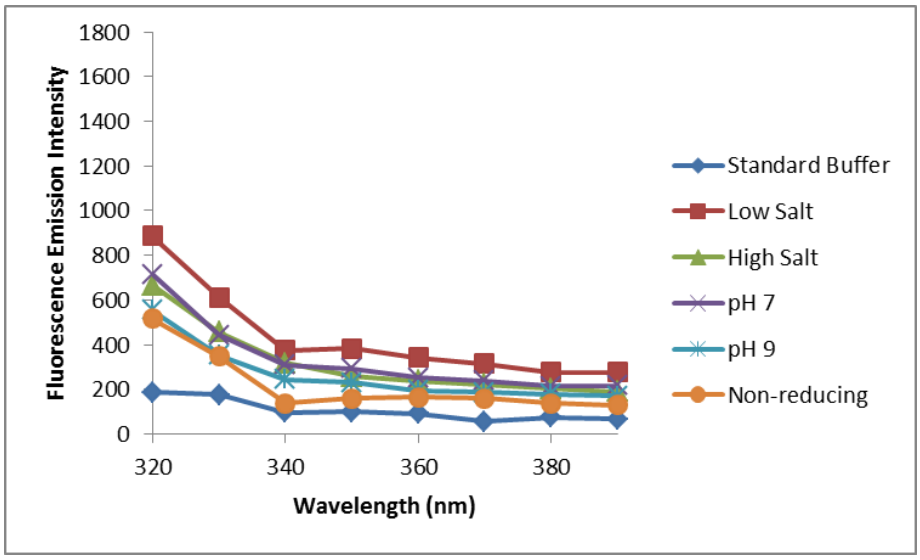
K. 10 µM PduB' F188W with 1mM $ZnCl_2$



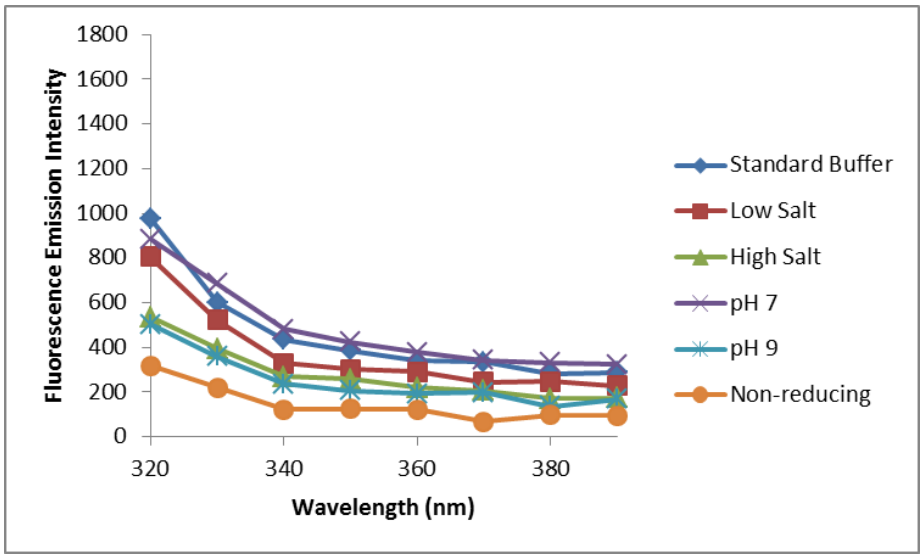
L. 10 µM PduB' F188W with 10mM $ZnCl_2$



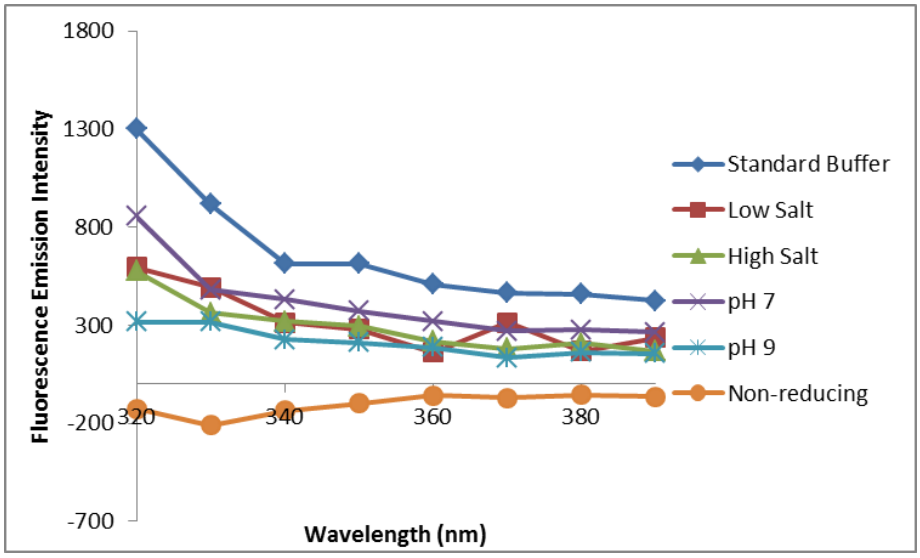
M. 10 µM PduB' F188W with 1mM $MgCl_2$



N. 10 µM PduB' F188W with 10mM $MgCl_2$



O. 10 µM PduB' F188W with 1mM $CoCl_2$



P. 10 µM PduB' F188W with 10mM $CoCl_2$

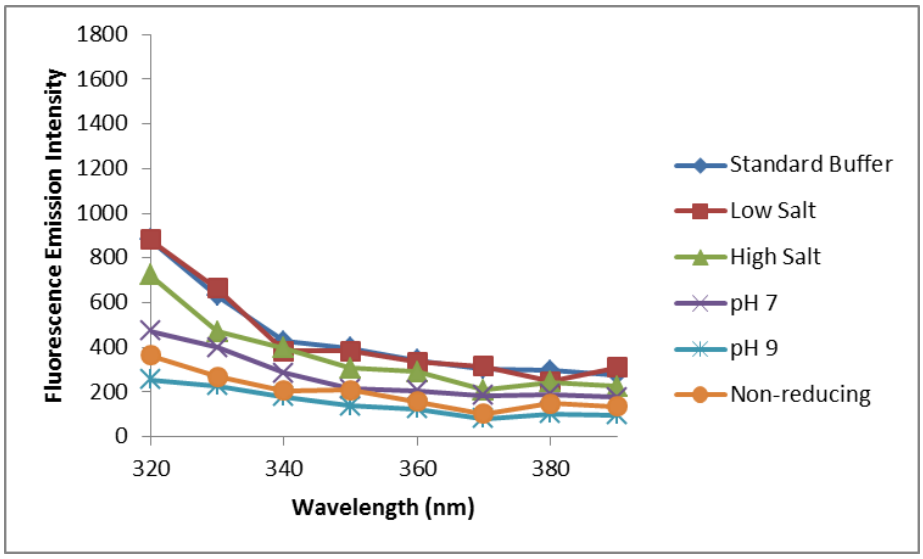
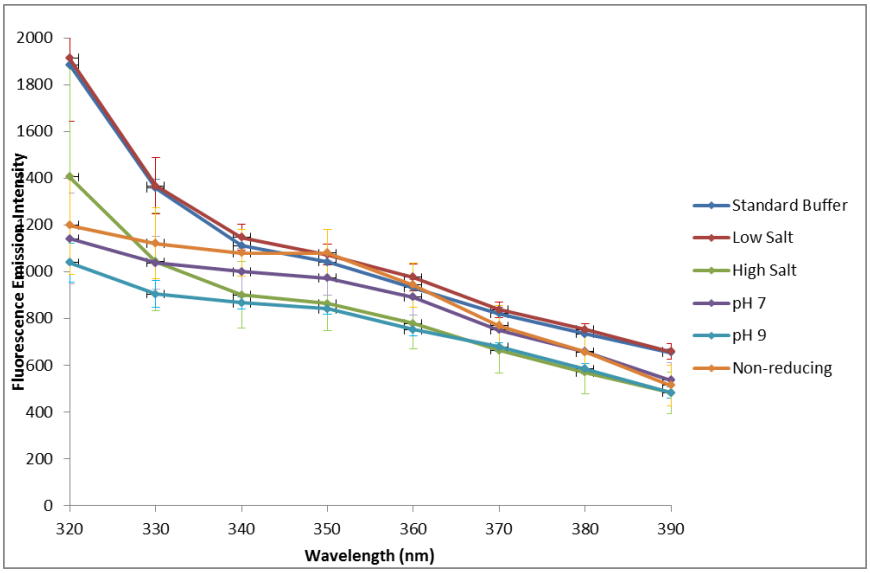
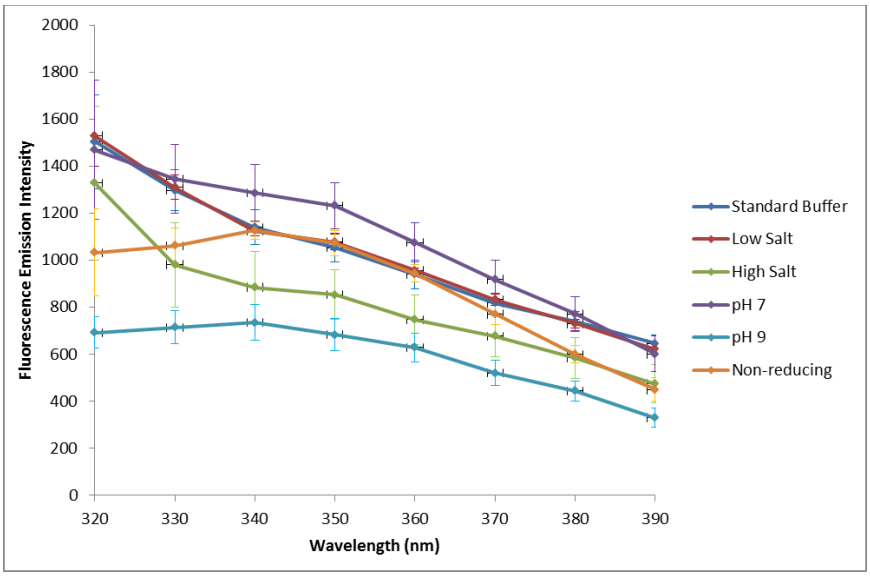


Figure 6.7. Tryptophan fluorescence emission of PduB' F188W with 1mM $ZnCl_2$ and 1mM $CoCl_2$ changing over time.

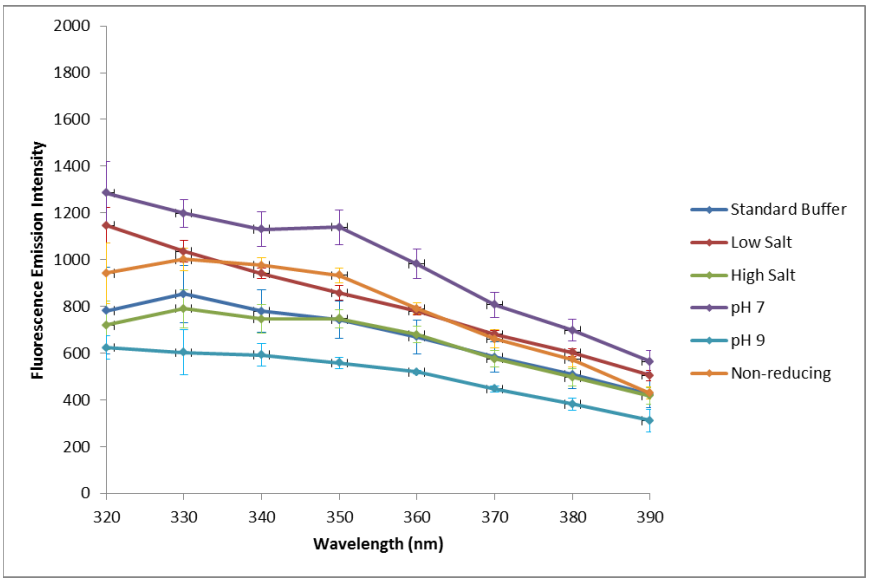
A. 10 µM PduB' F188W with 1mM $ZnCl_2$ (15 min)



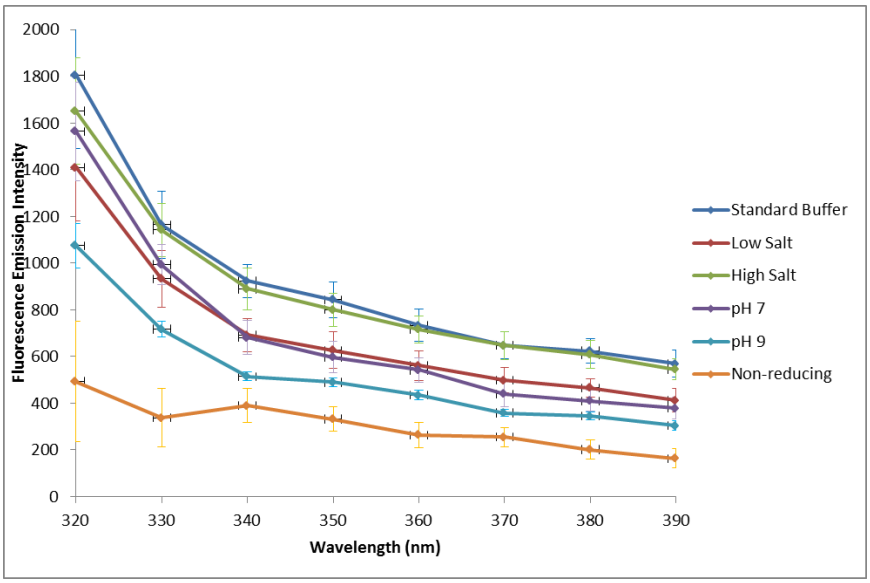
B. 10 µM PduB' F188W with 1mM $ZnCl_2$ (2h)



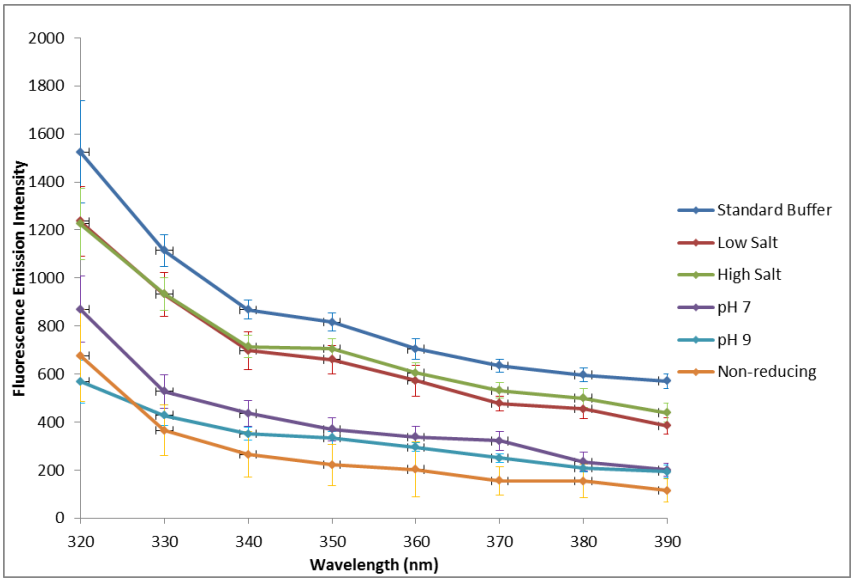
C. 10 µM PduB' F188W with 1mM $ZnCl_2$ (22h)



D. 10 µM PduB' F188W with 1mM $CoCl_2$ (15 min)



E. 10 µM PduB' F188W with 1mM $CoCl_2$ (2h)



F. 10 µM PduB' F188W with 1mM $CoCl_2$ (22h)

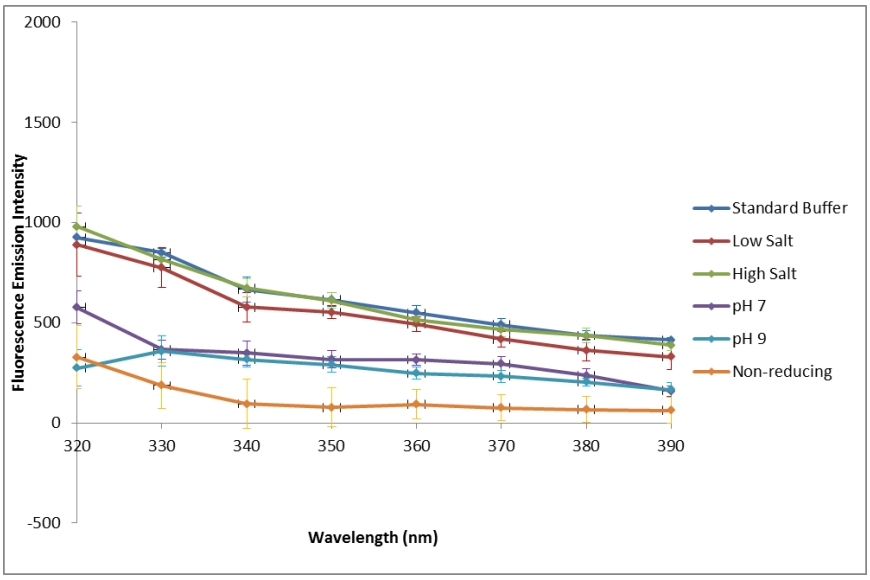

2000
1500
1000
500
0
-500
320
330
340
350
360
370
380
390
Fluorescence Emission Intensity
Wavelength (nm)
Standard Buffer
Low Salt
High Salt
pH 7
pH 9
Non-reducing

Figure 6.8. HisTrap Ni^{2+}-affinity purification of PduB' F188W in pET22b vector: Most protein remained in the insoluble fraction.

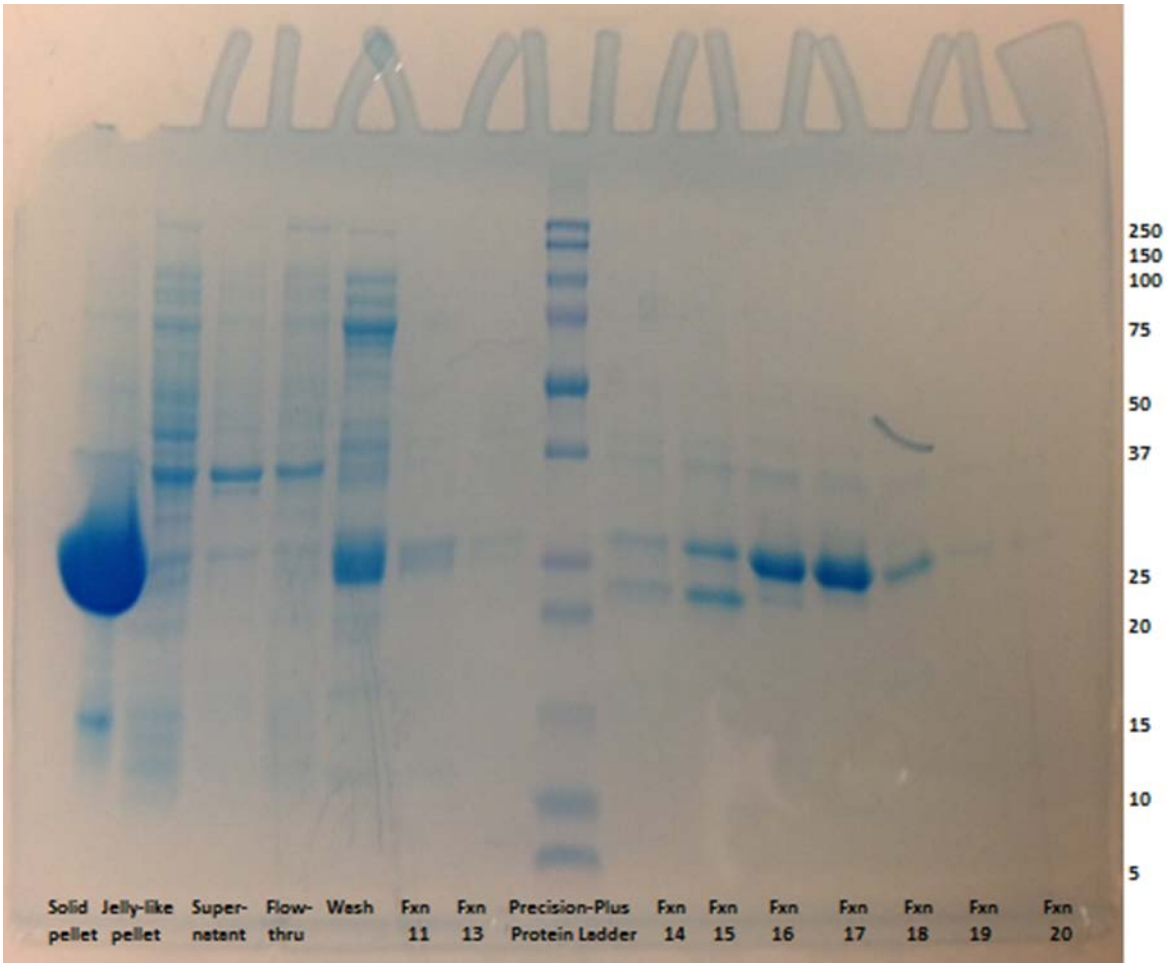


Figure 6.9. HisTrap Ni^{2+}-affinity purification of PduB' F188W MBP GFP fusion in DCLIC vector: Most protein was soluble.

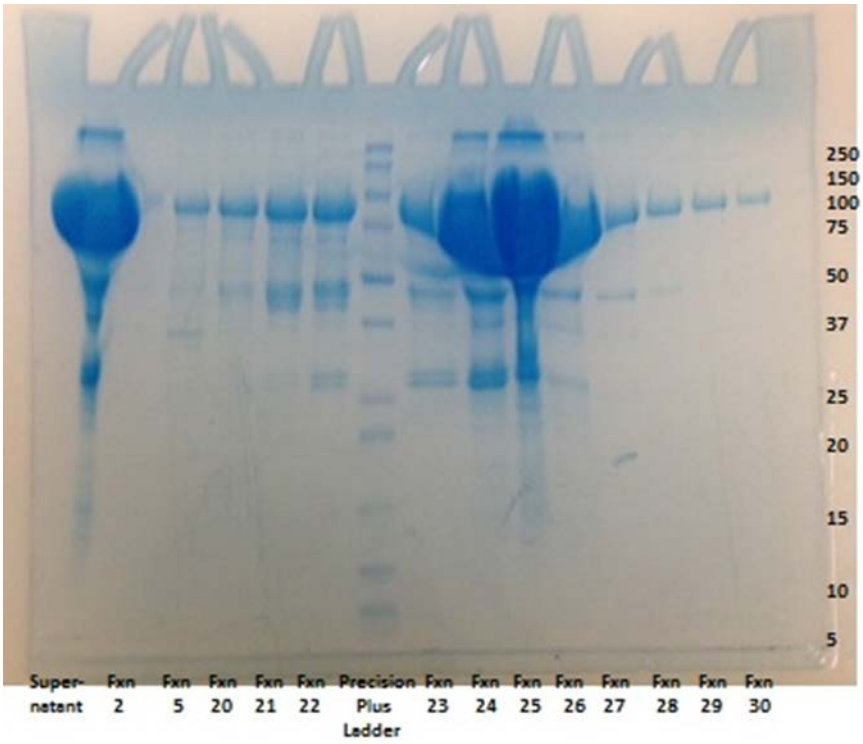


Figure 6.10. Post Ni^{2+} affinity chromatography, PduB' F188W MBP GFP fusion protein was dialyzed and cleaved with TEV protease. A second Ni^{2+} affinity chromatography step removed the 6xHis-tagged MBP and GFP from the protein sample.

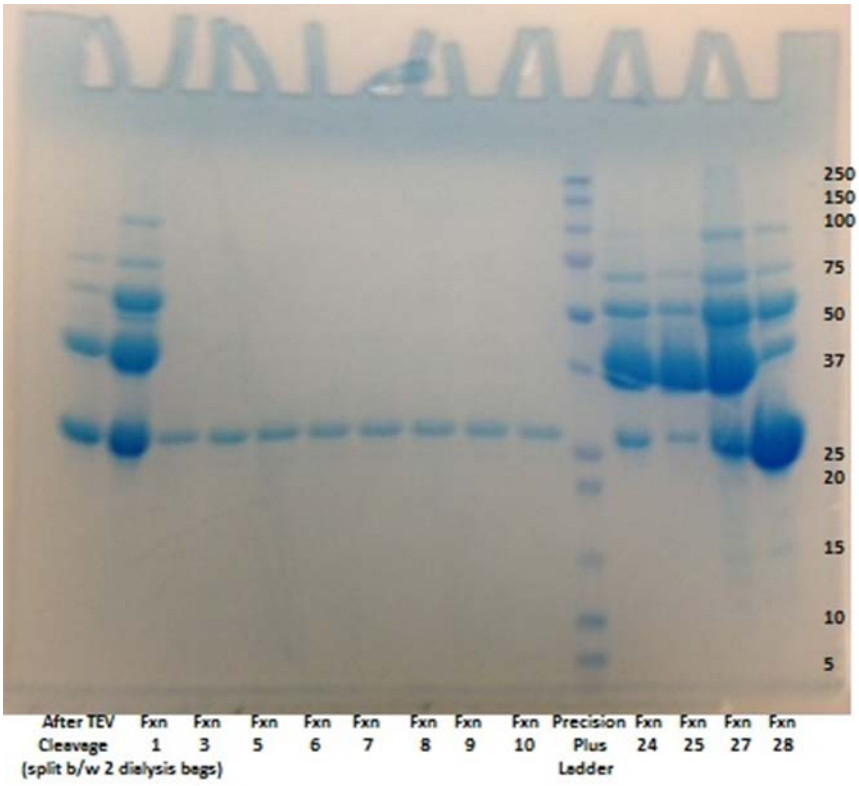


Figure 6.11. PduB' A53F K65A K170A crystals in condition JCSG A12

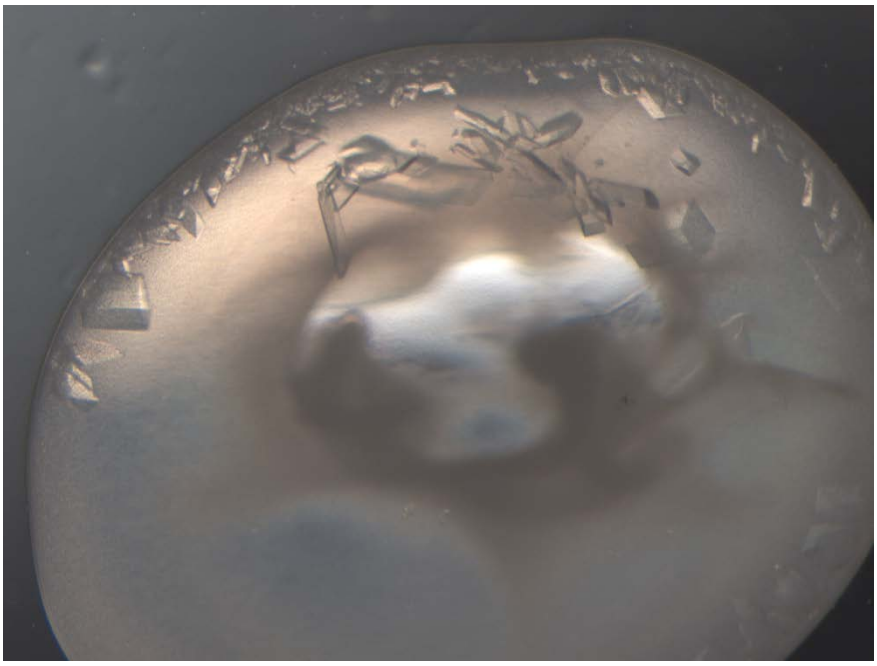
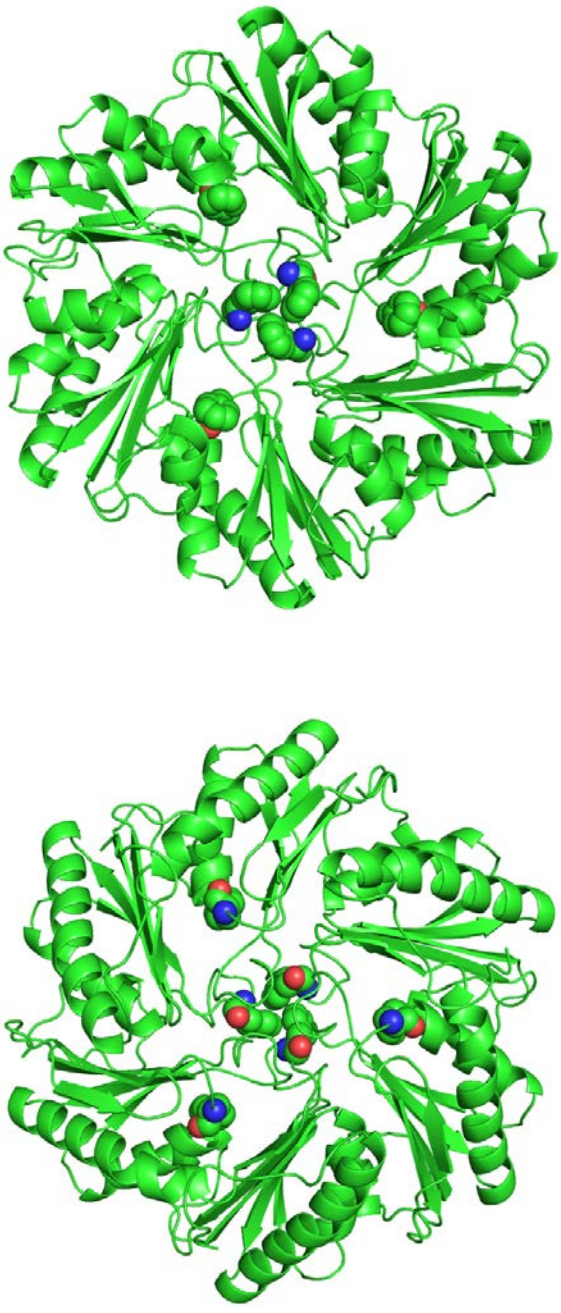


Table 6.4. Data and Refinement statistics for PduB' A53F K65A K170A structure

	PduB' A53F K65A K170A
Data collection	
Space group	P 1
Cell dimensions	
a, *b*, *c* (Å)	43.6, 62.9, 68.9
α, *β*, *γ* (°)	120.4, 94.1, 102.5
Wavelength (Å)	.97910
Resolution (Å)	56.5(2.3) *
R_{sym} or R_{merge}	.055(.075)
$I/\sigma I$	18.2(13.5)
Completeness (%)	92.9(89.4)

Redundancy	3.6(3.7)
Refinement	
Resolution (Å)	2.3
No. reflections	26870
$R_{work/}$ R_{free}	0.153/0.204
No. atoms	5029
Protein	4993
Ions	0
Water	36
B-factors	
Protein	25.5
Ions	0
Water	25.3
R.m.s deviations	
Bond lengths (Å)	.007
Bond angles (º)	0.9

Figure 6.12. PduB' A53F K65A K170A trimer structure solved to 2.3Å resolution. (Top) Back view highlighting central F188 and pocket A53F in spherical representation. (Bottom) front view of same representation.



Chapter 7

The function of PduJ is determined by the genomic position of its encoding gene

Introduction

PduJ is the most abundant protein in the Pdu MCP but the function is yet unknown. It shares 77.7% sequence homology to PduA (Figure 1) however analogous pore mutants (S39L/A/X) introduced by chromosomal mutations do not produce similar phenotypes in *Salmonella enterica* according to Bobik et al. (private discussions). It was suggested that PduJ may serve as an edge structural unit in the Pdu MCP by forming a bent hexamer. Thus, we pursued the crystal structure of PduJ with the K25A mutation which increases protein solubility.

PduJ K25A was crystallized in spacegroup P6 and was solved to 1.5Å resolution (see details below). In the crystal structure, PduJ K25A also forms a hexamer, adopting the canonical BMC hexamer conformation that is not bent. Structural alignment with the PduA hexamer gives a backbone r.m.s.d. of 1Å.

The function of the PduJ microcompartment shell protein is determined by the genomic position of its encoding gene

Chiranjit Chowdhury,[1] Sunny Chun,[2] Michael R. Sawaya,[2] Todd O. Yeates[2,3,4] and Thomas A Bobik[1]*

[1]*Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA 50011.*

[2]*Department of Energy Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, CA 90095.*

[3]*Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA 90095.*

[4]*Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095.*

Summary

Bacterial microcompartments (MCPs) are complex organelles that consist of metabolic enzymes encapsulated within a protein shell. In this study, we investigate the function of the PduJ MCP shell protein. PduJ is 80% identical in amino acid sequence to PduA and both are major shell proteins of the 1,2-propanediol (1,2-PD) utilization (Pdu) MCP of *Salmonella*. Prior studies showed that PduA mediates the transport of 1,2-PD (the substrate) into the Pdu MCP. Surprisingly, however, results presented here establish that PduJ has no role 1,2-PD transport. The crystal structure revealed that PduJ was nearly identical to that of PduA and, hence, offered no explanation for their differential functions. Interestingly, however, when a *pduJ* gene was placed at the *pduA* chromosomal locus, the PduJ protein acquired a new function, the ability to mediate 1,2-PD transport into the Pdu MCP. To our knowledge, these are the first studies to show that that gene location can determine the function of a MCP shell protein. We propose that gene location dictates protein-protein interactions essential to the function of the MCP shell.

Accepted 11 May, 2016. *For correspondence. E-mail bobik@iastate.edu; Tel. (515) 294 8247; Fax (515) 294 0453.

Introduction

Many bacteria use proteinaceous organelles known as microcompartments (MCPs) to optimize metabolic processes (Abdul-Rahman *et al.*, 2013; Rae *et al.*, 2013; Chowdhury *et al.*, 2014; Bobik *et al.*, 2015). MCPs are polyhedral in shape, about the size of a large virus (100–200 nm in diameter), and consist of metabolic enzymes encapsulated within a selectively permeable protein shell. The functions of MCPs are to accelerate catalysis, block wasteful by-reactions and sequester toxic or volatile metabolic intermediates in order to minimize cellular damage and/or loss of valuable carbon (Fig. 1A) (Price and Badger, 1989; Havemann *et al.*, 2002; Penrod and Roth, 2006; Sampson and Bobik, 2008). Bioinformatic analyses indicate that MCPs are produced by hundreds of different bacterial species ranging across 23 different phyla and that they are used to enhance seven or more different metabolic processes (Abdul-Rahman *et al.*, 2013; Jorda *et al.*, 2013). The three best-studied MCPs are carboxysomes, which are used for autotrophic carbon fixation (Rae *et al.*, 2013) and the Pdu and Eut MCPs which are involved in the coenzyme B_{12}-dependent catabolism of 1,2-propanediol (1,2-PD) and ethanolamine, respectively (Chowdhury *et al.*, 2014). Other MCPs are proposed to be involved in ethanol degradation (Seedorf *et al.*, 2008), choline catabolism (Craciun and Balskus, 2012; Kuehl *et al.*, 2014), 1,2-PD degradation initiated by a glycyl radical enzyme (rather that a B_{12}-dependent enzyme) (Scott *et al.*, 2006; Petit *et al.*, 2013) and in the degradation of algal polysaccharides (Erbilgin *et al.*, 2014).

Although MCPs are functionally distinct in terms of metabolism due to variation in their encapsulated enzymes, all have related protein shells (Yeates *et al.*, 2013; Chowdhury *et al.*, 2014). MCP shells are primarily composed of a family of proteins known as bacterial microcompartment (BMC) domain proteins (Kerfeld *et al.*, 2005; Tanaka *et al.*, 2008; 2009; Yeates *et al.*, 2013). Most MCP shells are built from several thousand BMC-domain proteins of 5–10 different types (Chowdhury *et al.*, 2014). A number of BMC-domain proteins

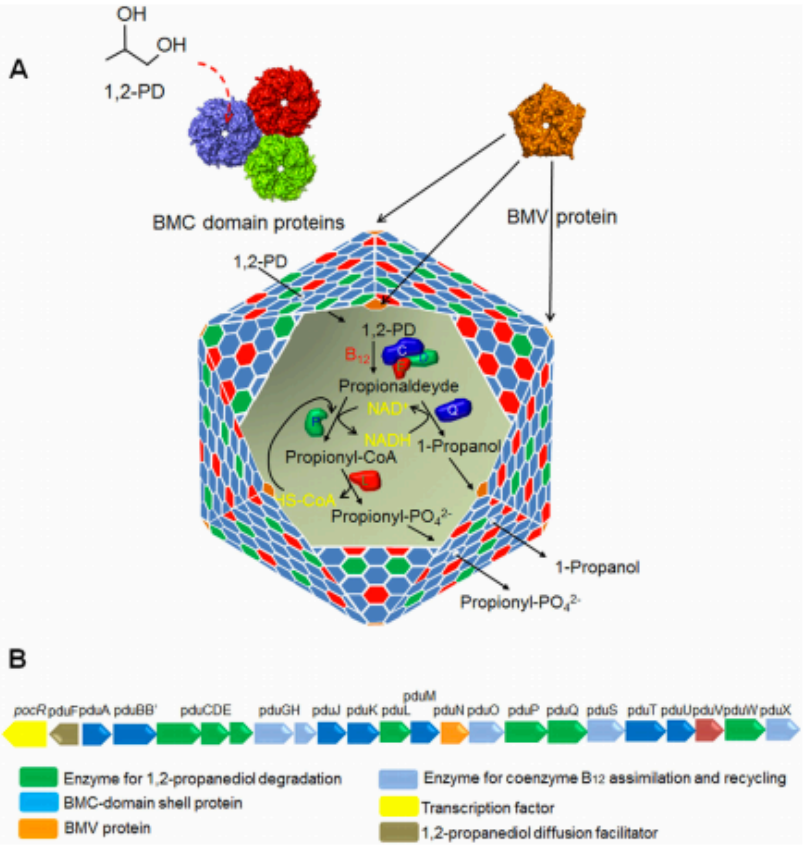


Fig. 1. Model for the Pdu MCP of *Salmonella*.
A. The mosaic structure of the Pdu MCP where different BMC-domain proteins tile together forming facets of the shell. A pentameric non-BMC domain protein (BMV) caps the vertices of the polyhedral structure. The central pore of PduA, a BMC domain protein plays a pivotal role in the selective transport of 1, 2-PD (the substrate) across the shell. A primary reason for the encapsulation of enzymes for 1,2-PD degradation is to sequester a toxic intermediate (propionaldehyde).
B. The organization of *pdu* operon which encodes the known genes of the Pdu MCP including seven BMC-domain shell proteins and 1,2-PD degradative enzymes. The details of the functions of these genes were reviewed recently (Chowdhury *et al.*, 2014; Bobik *et al.*, 2015).

pack tightly side-by-side into extended two-dimensional layers and different types of BMC-domain proteins have edges with complementary shapes. Hence, the protein shells of bacterial MCPs are thought to be composed primarily of mixed sheets of BMC-domain proteins (Fig. 1A) (Heldt *et al.*, 2009; Klein *et al.*, 2009; Sagermann *et al.*, 2009; Crowley *et al.*, 2010). Nearly all MCPs also include at least one bacterial microcompartment vertex (BMV) protein (Tanaka *et al.*, 2008; Wheatley *et al.*, 2013). BMV proteins are pentamers (unrelated to BMC-domain proteins) that cap the vertices of MCPs and help close the polyhedral structure (Fig. 1A) (Tanaka *et al.*, 2008; Wheatley *et al.*, 2013).

MCP function requires that the substrates and cofactors needed by the encapsulated enzymes cross the protein shell and enter the MCP interior (Cheng *et al.*, 2008). However, the MCP must also restrict the egress of certain toxic/volatile intermediates (Price and Badger, 1989; Havemann *et al.*, 2002; Penrod and Roth, 2006; Sampson and Bobik, 2008). Studies indicate that BMC-domain proteins provide the basis for a selectively permeable protein shell that fulfills these requirements (Kerfeld *et al.*, 2005; Crowley *et al.*, 2010). In the case of the Pdu MCP, crystallographic studies showed that the PduA shell protein tiles into tightly packed protein sheets whose most notable openings are central pores at the six-fold axis of the PduA hexamer (Crowley *et al.*, 2010). These pores are lined with numerous hydrogen-bond donors and acceptors suggesting they might preferentially allow the movement of 1,2-PD (the substrate) compared to propionaldehyde (a toxic intermediate) which is less polar (Fig. 1A) (Crowley *et al.*, 2008, 2010). Importantly, Chowdhury and co-workers recently demonstrated selective permeability; structure-guided mutagenesis showed that the PduA pore acts as a major route for entry 1,2-PD into the Pdu MCP while restricting the egress of a propionaldehyde (Crowley *et al.*, 2010; Chowdhury *et al.*, 2015). Interestingly, crystallographic studies have suggested that tandem-BMC domain proteins such as PduB might have an allosterically-gated pores that allow the entry of bulkier enzymatic cofactors (Klein *et al.*, 2009; Sagermann *et al.*, 2009; Tanaka *et al.*, 2010; Cai *et al.*, 2013; Thompson *et al.*, 2015). Several tandem-BMC domain proteins have been crystallized in alternate conformations that have either a large central pore or one that is

occluded suggestive of a gate (Klein *et al.*, 2009; Sagermann *et al.*, 2009; Tanaka *et al.*, 2010; Cai *et al.*, 2013). In addition, the EutL tandem-BMC-domain protein has been shown to bind ethanolamine (the substrate of the Eut MCP) in a manner that indicates allosteric regulation of its pore (Thompson *et al.*, 2015). It is also notable that the PduT shell protein has a [4Fe-4S] cluster where the central pore would typically be found (Crowley *et al.*, 2010; Pang *et al.*, 2011). It has been suggested that PduT might act as an electron conduit or in the transport of Fe-S clusters between the cytoplasm of the cell and the MCP lumen (Crowley *et al.*, 2008; 2010).

The PduJ and PduA proteins are both major components of the shell of the Pdu MCP which is used by *Salmonella* and many other species for the catabolism of 1,2-PD as a carbon and energy source (Havemann and Bobik, 2003). The function of the Pdu MCP is to sequester propionaldehyde to prevent toxicity during 1,2-PD metabolism (Fig. 1A) (Havemann *et al.*, 2002; Sampson and Bobik, 2008). As mentioned above, PduA provides a major route for the selective entry of 1,2-PD into the Pdu MCP (Chowdhury *et al.*, 2015). Sequence alignments show that PduJ and PduA are 80% identical in amino acid sequence and have identical pore-lining residues (Fig. 2A). Hence, it would seem reasonable the PduJ and PduA have similar functions. Surprisingly, however, results presented here show that unlike PduA, the PduJ protein has no role in 1,2-PD transport. We also present the crystal structure of PduJ and show that the pore structures of PduA and PduJ are highly conserved. Lastly, we show that the function of the PduJ protein depends on the chromosomal position of its encoding gene. If a *pduJ* gene is placed in the chromosomal position normally occupied by *pduA*, the PduJ protein can then mediate 1,2-PD transport into the Pdu MCP. Based on these findings, we propose a model in which the specific location of the PduJ protein within MCP shell determines whether it can function in 1,2-PD transport.

Results

A PduJ pore mutant grows normally on 1,2-PD minimal medium

Recent studies showed that the pore structure of the PduA shell protein allows for the selective influx of substrate (1,2-PD) into the Pdu MCP (Chowdhury *et al.*, 2015). In these studies, a PduA S40L pore mutant was shown to be impaired for 1,2-PD transport across the MCP shell (Chowdhury *et al.*, 2015). Because the pore region of PduJ is identical in sequence to PduA (Fig. 2A), we tested whether the PduJ pore might also play a role in 1,2-PD uptake into the Pdu MCP. A chromosomal

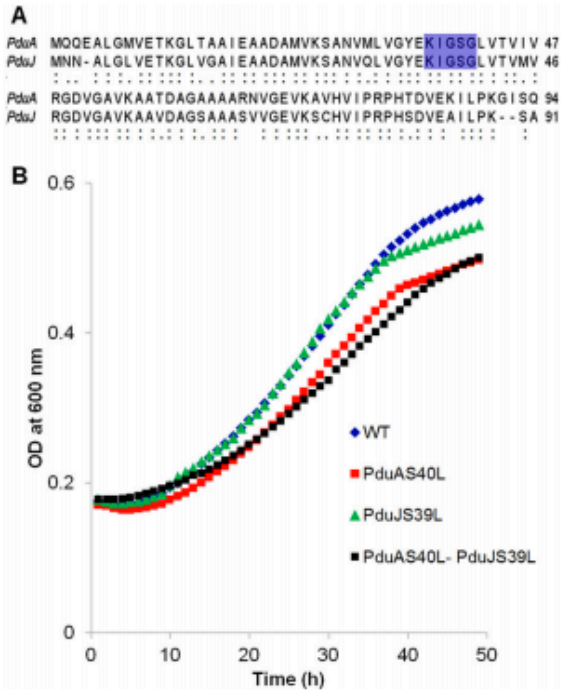


Fig. 2. Differential effects of analogous pore mutations in PduA and PduJ on growth of *Salmonella* on 1,2-PD at limiting B_{12}.
A. Pairwise sequence alignment showing high level of identity between PduA and PduJ. The conserved pore-lining residues are highlighted in blue. In PduA, S40 is a key pore-forming residue that plays a crucial role in 1, 2-PD transport. The analogous residue in PduJ is S39.
B. Growth of PduA S40L and PduJ S39L pore mutants with respect to wild-type (WT). Growth assays were repeated at least three times in triplicate or more and representative curves are shown.

PduJ S39L mutant was constructed (S39 of PduJ corresponds to S40 of PduA). However, in contrast to a PduA S40L mutant which grows slowly in 1,2-PD minimal medium at limiting B_{12} due to impaired 1,2-PD transport (Chowdhury *et al.*, 2015), a PduJ S39L mutant grew similarly to wild-type (Fig. 2B). The doubling times were as follows: wild-type (17.6 $\pm$1.1 h); PduA S40L (22.8 $\pm$ 0.6 h); and PduJ S39L (16.4 $\pm$ 0.8 h). We also tested growth of a PduJ S39L PduA S40L double mutant, and found that it grew similarly to PduA S40L (Fig. 2B) (the doubling time of the PduJ S39L PduA S40L double mutant was 23.1 $\pm$ 0.6 h). The error shown for doubling times is one standard deviation from the mean and in all cases was determined from at least three biological replicates measured in triplicate. The similar growth pattern of PduJ S39L to wild-type tentatively suggests that PduJ does not mediate 1,2-PD transport despite the fact that the pore region of PduJ is identical in sequence to that of PduA. Thus, despite their high sequence identity, PduJ seemed to have a different function than PduA.

MCPs isolated from the PduJ S39L pore mutant exhibited diol dehydratase activity similar to wild-type

Diol dehydratase (DDH) is an MCP lumen enzyme that catalyzes the first step of 1,2-PD degradation (the conversion of 1,2-PD to propionaldehyde); hence, its activity depends on the diffusion of 1,2-PD across the MCP shell. Previous studies showed that MCPs purified from a PduA S40L mutant exhibited lower DDH activity compared to wild-type due to impaired diffusion of 1,2-PD across the MCP shell (Chowdhury *et al.*, 2015). Therefore, as a second test to assess whether the PduJ pore has a role in 1,2-PD transport, we assayed DDH activity of MCPs purified from a PduJ S39L mutant. PduJ S39L MCPs were purified as described (Sinha *et al.*, 2012) with a yield > 90% compared to wild-type. In addition, transmission electron microscopy showed no gross abnormalities in the size and shape of PduJ S39L MCPs (Supporting Information Figs. S1 and S2) and SDS-PAGE followed by densitometry showed no obvious changes in the major proteins bands compared to wild-type MCPs (Supporting Information Fig. S3). However, unlike PduA S40L MCPs which had substantially reduced DDH activity, the DDH activity of PduJ S39L MCPs was indistinguishable from wild-type (Table 1). In addition, MCPs purified from the PduA S40L PduJ S39L double mutant had DDH activity similar to the PduA S40L single mutant further indicating that the PduJ S39L mutant did not restrict 1,2-PD transport (Table 1). As a control, dialysis and sonication were used to disrupt the shells of each type of MCP to allow free diffusion of 1,2-PD to DDH (Chowdhury *et al.*, 2015). After their shells were broken, wild-type, PduJ S39L, PduA S40L and PduA S40L PduJ S39L (double mutant) MCPs all had similar DDH activity (Table 1) showing that DDH was normally recruited and incorporated into the mutant MCPs. Thus, DDH assays supported the growth studies described above and provided further evidence that in contrast to PduA, PduJ has no role in 1,2-PD influx into the Pdu MCP.

Table 1. Effect of pore mutations in PduA and PduJ on the DDH activities of purified Pdu MCPs.

MCPs from Pdu Mutants	Specific Activity (μmol min^{-1} mg^{-1})
Wild type (WT)	28.2 ± 1.2
PduJ S39L	27.6 ± 0.8
PduA S40L	14.8 ± 1.5*
PduA S40L-PduJ S39L	14.4 ± 1.1*
WT-Broken MCP	34.8 ± 1.5*
PduJ S39L-Broken MCP	34.2 ± 0.8*
PduA S40L-Broken MCP	33.8 ± 0.3*
PduA S40L-PduJ S39L-Broken MCP	34.2 ± 0.8*

The DDH activities shown are based on at least three biological replicates measured in triplicate. The error estimate shown is ± one standard deviation.
*P-value <0.0001 compared to wild-type.

The structure of PduJ is nearly identical to that of PduA

The dissimilar transport behaviour of PduA and PduJ, despite identical amino acid sequences in their pore regions, was unexpected. One possibility was that PduJ and PduA have significant structural differences despite their high sequence identity (80%). To examine this, PduJ K25A was crystallized in space group P6 and its structure was solved at 1.5Å resolution. The K25A mutant was used to reduce crystallographic problems that result from edge-to-edge aggregation of BMC hexamers (Sinha *et al.*, 2014). The structure was refined using PHENIX and manual manipulation using COOT, and BUSTER 2.10.0. Protein geometry analysis revealed no Ramachandran outliers, with 98.9% residues in favored regions. The Molprobity clash score after adding hydrogens was 0.78. Data collection statistics are shown in Supporting Information Table S1. In the crystal structure, PduJ K25A forms a hexamer, adopting the canonical BMC hexamer conformation (PDBID 5D6V, Fig. 3A). In the PduJ structure, the C-terminal region is well-ordered and forms a short alpha helix whereas this segment is disordered and not resolved in the PduA structure (Fig. 3B). Structural alignment of the PduJ hexamer with the PduA hexamer (PDBID 3NGK) gives a backbone r.m.s.d. deviation of 0.28Å (Fig. 4A). The pore region of PduJ appears identical to that of the PduA (Fig. 4B). Aligning all atoms of the pore regions (PduJ residues I37-L41 and PduA residues I38-L42) of the monomers gives an r.m.s.d of 0.185Å. Also, the PduJ pore has a similar pore diameter of 5.7Å to that of PduA, 5.6Å (Fig. 4B). Thus, there are no obvious structural differences between PduA and PduJ that could account for the dissimilar transport behaviour.

pduA can complement the growth phenotype of a pduJ deletion mutant but not vice versa

Prior studies showed that a *pduA* deletion mutant grows faster than wild-type at limiting levels of B_{12} (20 nM) (Cheng *et al.*, 2011; Sinha *et al.*, 2014). This phenotype results from a damaged MCP shell which allows increased access of the lumen enzymes to their substrates (Sinha *et al.*, 2014). Hence, growth rate at limiting B_{12} provides a test for the structural integrity of the Pdu MCP (Cheng *et al.*, 2011). Using this test, we examined whether PduJ and PduA are interchangeable for MCP assembly. Control experiments (complementation tests) showed that the fast-growth phenotype of

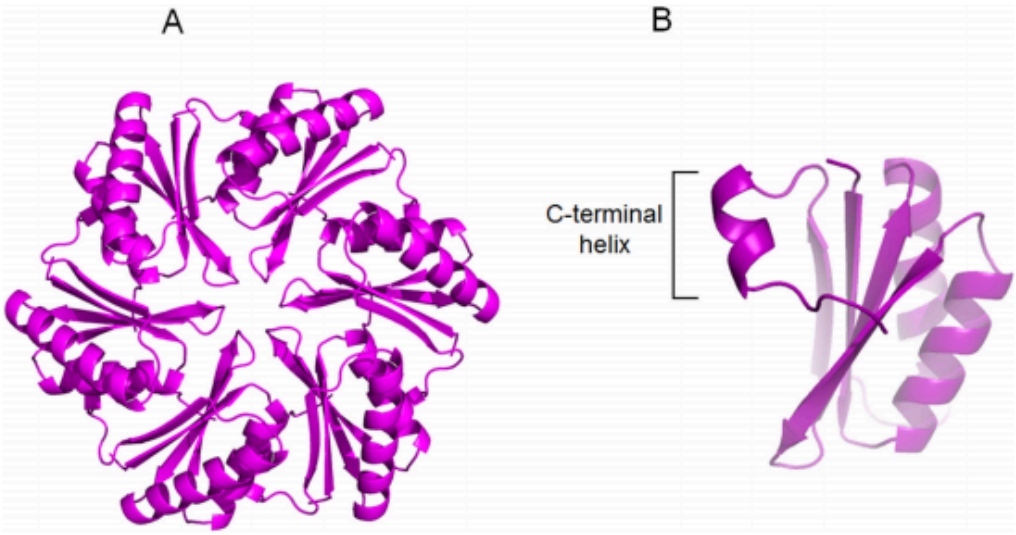


Fig. 3. Crystal structure of PduJ K25A (PDBID 5D6V).
A. Cartoon representation from cytosolic face
B. The C-terminal alpha helix of PduJ is shown which is well-ordered, unlike PduA.

Δ*pduJ* was restored to wild-type levels by expression of *pduJ* from plasmid pLac22 at ~50 μM IPTG (Fig. 5A, Table 2). We also found that production of PduA using 50 μM IPTG largely corrected the fast-growth phenotype of a *pduJ* deletion mutant (Fig. 5B, Table 2). In contrast, a *pduA* deletion was only partially complemented by expression of PduA at optimal IPTG levels (Fig. 5C, Table 2) as reported previously (Cheng *et al.*, 2011), and was not complemented at all by production of PduJ from pLac22 at any IPTG concentration tested (Fig. 5D, Table 2). Thus, with respect to MCP function, PduA can replace PduJ but not vice versa.

We note that the growth phenotypes of certain strains (Δ*pduA*/pLac22-*pduA* and Δ*pduJ*/pLac22-*pduJ*) were more severe (faster growth) without added IPTG (Fig. 5A and C). We speculate that a very low level of expression from pLac22 plasmid might have an antagonistic effect on MCP assembly in these particular cases.

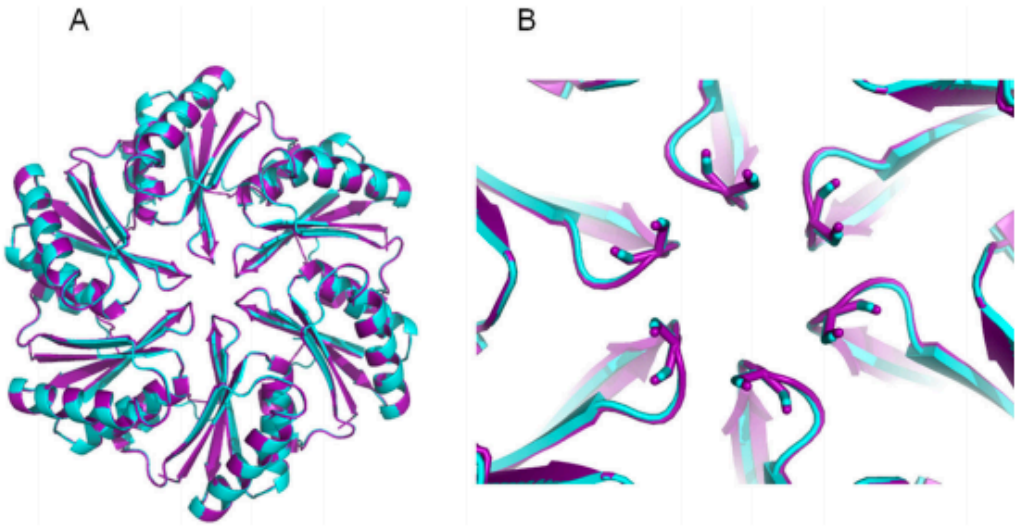


Fig. 4. Structural alignment of PduA and PduJ.
A. PduJ hexamer (purple) and PduA hexamer (cyan).
B. Close up view of the aligned pore regions of PduA and PduJ.

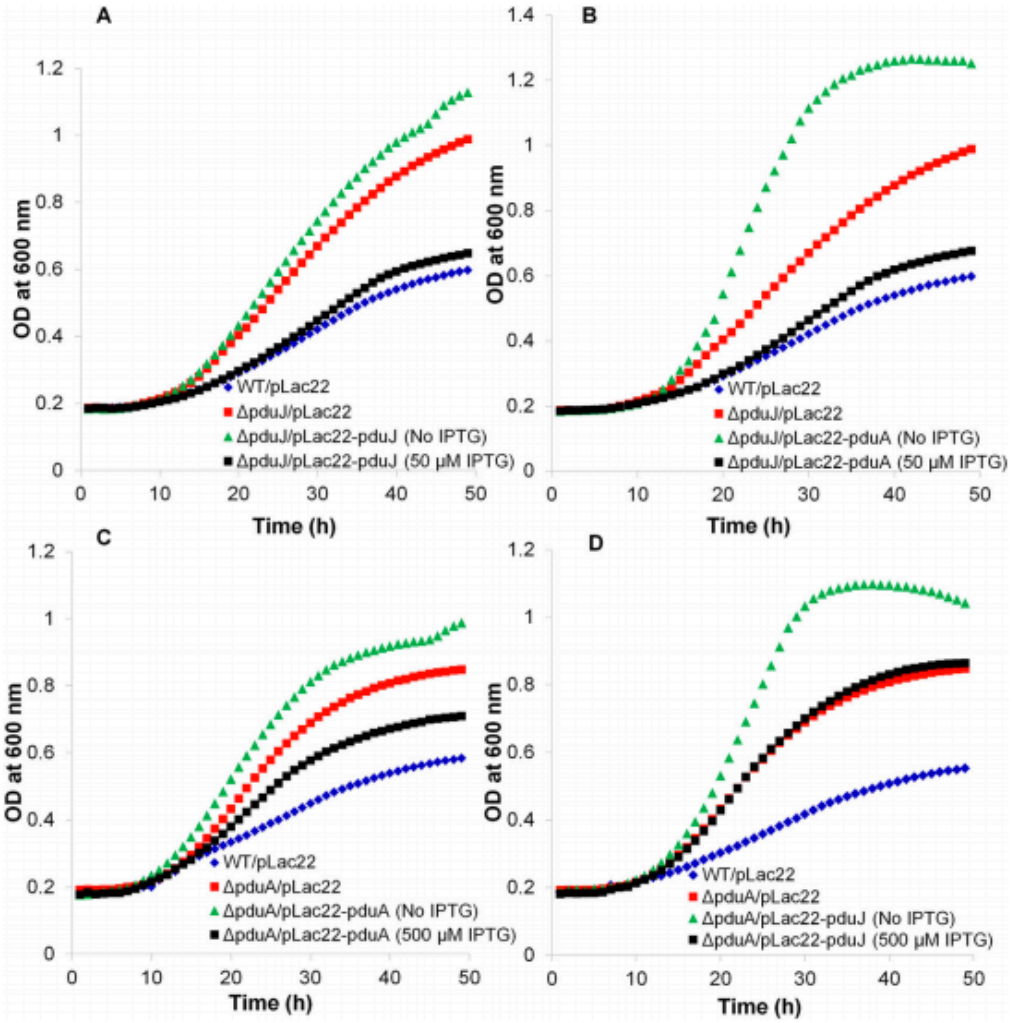


Fig. 5. Complementation and cross-complementation studies with Δ*pduA* and Δ*pduJ* mutants.
A. Complementation of Δ*pduJ* with *pduJ*.
B. Complementation of Δ*pduJ* by *pduA*.
C. Partial complementation of Δ*pduA* by *pduA*.
D. Attempted complementation of Δ*pduA* by *pduJ*. Gene expression from pLac22 was induced by adding IPTG at 0, 10, 20, 30, 50, 100, 200, 500 μM or 1 mM). The results shown in the figure are for the IPTG level at which optimal complementation occurred. Other concentrations of IPTG resulted in similar or worse complementation. All growth assays were done at least three times and in triplicate and a representative outcome is shown.

pduA *can complement the DDH phenotype of a* ΔpduJ *mutant but not vice versa*

MCPs purified from a *pduJ* deletion mutant show higher DDH activity than wild-type MCPs because the MCP shell was disrupted allowing DDH greater access to its substrate 1,2-PD (Table 3) (Sinha *et al.*, 2014). Ectopic production of either PduJ or PduA reduced the DDH activity of purified Δ*pduJ* MCPs substantially closer to wild-type levels (Table 3) consistent with growth studies (Fig. 5). The ectopic production of PduA and PduJ from pLac22 was confirmed by SDS PAGE of purified MCPs (Supporting Information Fig. S3). In contrast, production of PduJ from pLac22 had no effect on the DDH activity of MCPs purified from a *pduA* deletion (Table 3). Thus, in vitro assays with purified MCPs also indicate that PduA can substantially substitute for PduJ, but not vice versa.

The aberrant morphology of ΔpduJ *MCPs is partially corrected by expression of PduA from a plasmid*

A *pduJ* deletion mutant produced elongated MCPs (Fig. 6) as was observed earlier (Cheng *et al.*, 2011; Sinha *et al.*, 2014).

Table 2. Complementation and cross-complementation of *pduA* and *pduJ* mutants.

Pdu mutants	Doubling time (h)
WT/pLac22	19.4 ± 0.8
Δ*pduA*/pLac22	10.5 ± 0.8*
Δ*pduA*/pLac22-*pduA* (induced with 500 μM IPTG)	14.2 ± 0.3*
Δ*pduJ*/pLac22	10.1 ± 0.7*
Δ*pduJ*/pLac22-*pduJ* (induced with 50 μM IPTG)	18.1 ± 0.5
Δ*pduA*/pLac22-*pduJ* (induced with 500 μM IPTG)	9.5 ± 1.2*
Δ*pduJ*/pLac22-*pduA* (induced with 50 μM IPTG)	17.1± 1.1
Δ*pduA*/pLac22-*pduA* (uninduced)	9.5 ± 0.7*
Δ*pduJ*/pLac22-*pduJ* (uninduced)	8.8 ± 1.1*

Doubling times were calculated from at least three biological replicates measured in triplicate. The error estimate shown is ± one standard deviation
*P-value <0.001 compared to wild-type.

The structure of Δ*pduJ* MCPs was partly corrected by production of PduA from a plasmid (Fig. 6). Upon expression of PduA, the elongated MCPs produced by a Δ*pduJ* mutant assumed a rounded shape more similar to wild-type although they had a somewhat more wrinkled or folded appearance (Fig. 6). A Δ*pduA* mutant forms slightly enlarged MCPs (Fig. 6 and Supporting Information Fig. S2) as was previously shown (Cheng *et al.*, 2011). PduJ expressed from a plasmid did not observably change the MCP morphology of Δ*pduA* mutant although any cross complementation would be difficult to discern since Δ*pduA* MCPs are only slightly larger than wild-type MCPs (Fig. 6 and Supporting Information Fig. S2). Nonetheless, overall, three lines of evidence (described above) indicate that PduA can substitute for PduJ to a large extent but not *vice versa*. This emphasizes the question, why can PduA functionally substitute for PduJ but not *vice versa*?

A pduJ *gene can complement a* pduA *deletion if placed at the* pduA *chromosomal locus*

In order to examine the effects of genomic position, the chromosomal copy of the *pduA* gene was replaced with a *pduJ* gene that had altered codon usage (to block recombination with the native copy of *pduJ*) but produced a wild-type PduJ protein. We refer to the resulting mutant as Δ*pduA::pduJ*^{M} (M for modified). Studies showed that the growth rate of Δ*pduA::pduJ*^{M} mutant on 1,2-PD minimal medium with limiting B_{12} was substantially closer to wild-type compared to the *pduA* deletion mutant whose growth rate is markedly increased due to increased permeability of the MCP shell. The doubling times were as follows: wild-type, 17.6 ± 1.1 h; Δ*pduA::pduJ*^{M}, 14.5 ± 0.2 h; and Δ*pduA* was 9.1 ± 0.8 h. (Fig. 7). The error shown for doubling times is one standard deviation from the mean and in all cases was determined from at least three biological replicates measured in triplicate. Hence, these growth tests suggested that PduJ partially replaced PduA when produced from the *pduA* chromosomal position. This contrasted with tests in which PduJ was produced from a plasmid where no observable correction of the Δ*pduA* phenotype was observed at any IPTG level used (0, 10, 20, 30, 50, 100, 200, 500 μM or 1 mM) (Fig. 5D).

We also purified MCPs from a Δ*pduA::pduJ*^{M} mutant and found that they had close to wild-type levels of DDH activity (~29 μmol min^{-1} mg^{-1}) (Table 4). This is in contrast to MCPs purified from the Δ*pduA* mutant which had increased levels of DDH activity (~34 μmol min^{-1} mg^{-1}) (Table 4) due to a damaged shell. Furthermore, control experiments showed that when MCPs of Δ*pduA::pduJ*^{M} mutant were broken by dialysis followed by sonication, the DDH activity the mutant was similar to that of broken wild-type MCPs. This showed that the amount of DDH encapsulated in the PduJ^M MCPs were similar to the parent strain. We also note that the PduJ^M MCPs were purified with ~90% yield compared to wild-type MCPs. Transmission electron microscopy did not reveal any obvious morphological defect in the PduJ^M MCPs (Supporting Information Figs. S1 and S2) and SDS-PAGE followed by densitometry indicated that the major MCP proteins were present at near wild-type levels (minor proteins are not detected by SDS-PAGE) (Supporting Information Fig. S4). The near wild-type levels of DDH activity in purified Δ*pduA::pduJ*^{M} MCPs further supported the idea the *pduJ* can replace *pduA* when expressed from the chromosomal position of *pduA*.

Table 3. Effect of complementation on DDH activity.

MCPs from Pdu mutants	Specific activity (μmol min^{-1} mg^{-1})
WT/pLac22	27.3 ± 0.8
Δ*pduJ*/pLac22	33.7 ± 0.7
Δ*pduJ*/pLac22-*pduJ*	27.8 ± 0.7
Δ*pduA*/pLac22-*pduA*	29.5 ± 0.8*
Δ*pduA*/pLac22	34.8 ± 1.1
Δ*pduA*/pLac22-*pduJ*	34.5 ± 0.9*
Δ*pduJ*/pLac22-*pduA*	28.7 ± 0.7

The DDH activities are based on at least three biological replicates measured in triplicate. The error estimate shown is ± one standard deviation.
*P-value <0.001 compared to wild-type/pLac22.

The pore of PduJ^M *functions in 1,2-PD transport*

Prior studies showed that the PduA shell protein functions in 1,2-PD transport and that a PduA S40L pore mutation impairs movement of 1,2-PD into the MCP lumen (Chowdhury *et al.*, 2015). To determine if a change in chromosomal position allowed the PduJ pore to function in 1,2-PD transport, we mutated serine 39 to leucine in PduJ^M to produce Δ*pduA::pduJ*^{M} S39L.

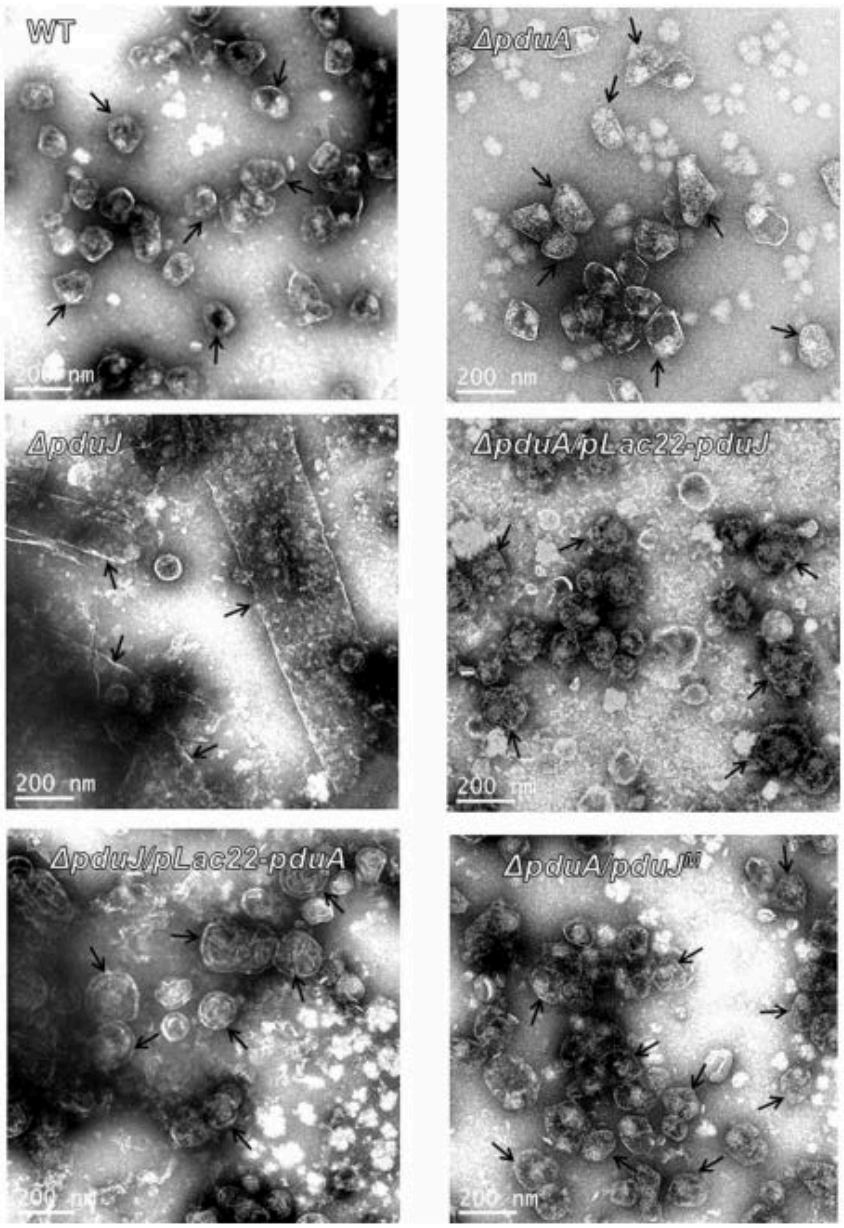


Fig. 6. Transmission electron micrographs of purified MCPs from PduA and PduJ deletion mutants as compared to wild-type. WT is wild-type *Salmonella*. A *pduA* deletion mutant formed somewhat enlarged MCPs (Sinha *et al.*, 2014) while a Δ*pduJ* mutant formed elongated MCPs.

Changing the pore lining serine to leucine (S39L) in $PduJ^M$ resulted in a slower growth rate on 1,2-PD which is indicative of impaired 1,2-PD transport into the Pdu MCP (Supporting Information Fig. S5) as previously described for a PduA S40L mutant (Chowdhury *et al.*, 2015). This is in contrast to S39L mutation in *pduJ* located at its wild-type chromosomal position where no growth defect was observed (Fig. 2B).

As a second test of the pore function in $PduJ^M$, we purified $PduJ^M$ S39L MCPs and measured their DDH activity. The $PduJ^M$ S39L MCPs had lower DDH activity (18.8 ± 1.5 μmol min^{-1} mg^{-1}) than purified wild-type or $PduJ^M$ MCPs (28.2 ± 1.2 and 29.3 ± 0.6 μmol min^{-1} mg^{-1}) (Table 4, line 6). Controls showed that when the $PduJ^M$ S39L MCPs were broken, their DDH activity was similar to that of broken wild-type showing normal DDH

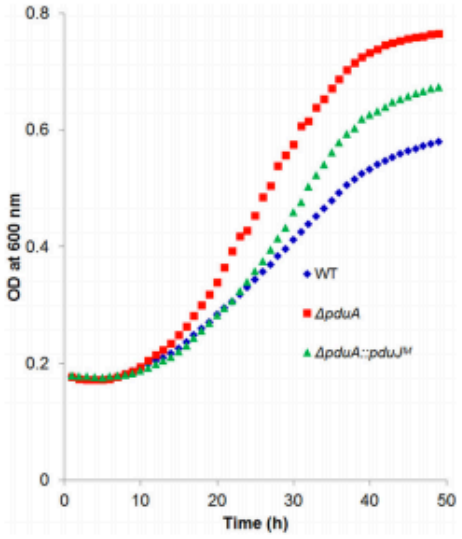


Fig. 7. Growth phenotypes of wild-type, Δ*pduA* and Δ*pduA*::*pduJ*^{M} at limiting B_{12}. WT is wild-type *Salmonella*. Δ*pduA*::*pduJ*^{M} is a mutant were the chromosomal copy of *pduA* has been replaced by a recoded *pduJ* gene.

targeting to the mutant MCPs (Table 4, line 7). Thus, the reduced DDH activity of the PduJ^{M} S39L MCPs is indicative of impaired pore function as previously described for PduA S40L (Chowdhury *et al.*, 2015). These results contrast markedly to those obtained for the PduJ S39L mutation (the *pduJ* gene encoding the S39L mutation is at its wild-type chromosomal location) which had no significant effect on DDH activity (Table 1). Together, these studies provide strong evidence that the function of PduJ changes depending on the location of its encoding gene.

As a further control, we used LC-MS/MS to verify incorporation of both PduJ and PduJ^{M} S39L into purified MCPs isolated from the *pduA*::*pduJ*^{M} S39L mutant. A number of peptides with *m/z* values in good agreement with predictions from the protein sequence confirmed the presence of both PduJ and PduJM S39L (Supporting Information Table S2) in purified MCPs.

Table 4. DDH activity upon change in chromosomal position of PduJ.

MCPs from Pdu mutants	Specific activity (μmol min^{-1} mg^{-1})
Wild type (WT)	28.2 ± 1.2
Δ*pduA*::*pduJ*^{M}	29.3 ± 0.6
Δ*pduA*	34.3 ± 1.2
WT-Broken MCP	33.7 ± 0.8
Δ*pduA*::*pduJ*^{M}-Broken MCP	34.2 ± 1.2
Δ*pduA*::*pduJ*^{M}*S39L*	18.8 ± 1.5*
Δ*pduA*::*pduJ*^{M}*S39L*-Broken MCP	33.4 ± 0.6

Discussion

Bacterial MCPs are distinguished from other known organelles in having a protein shell rather than a lipid membrane as a permeability barrier. MCP shells are generally composed of 5–10 different types of BMC-domain proteins that are functionally diversified to allow the selective transport of metabolites and perhaps to satisfy different architectural requirements (Rae *et al.*, 2013; Yeates *et al.*, 2013; Chowdhury *et al.*, 2014). Regardless of the structural diversification, nearly all BMC domain proteins have a conserved flat hexagonal structure (Rae *et al.*, 2013; Yeates *et al.*, 2013; Chowdhury *et al.*, 2014), and it is thought that shape complementarity along the edges allows different types of BMC-domain proteins to tessellate edge-to-edge to form the mixed protein sheets that comprise the facets of the shell (Fig. 1A) (Rae *et al.*, 2013; Yeates *et al.*, 2013; Chowdhury *et al.*, 2014). This idea is supported by alanine scanning mutagenesis studies that showed a conserved N-R-K triad found along the edges of BMC-domain proteins is essential in sheet formation and MCP assembly (Sinha *et al.*, 2014). The structural similarity and edge complementarity among varied BMC domain proteins raises the question of the rules that determine the incorporation of particular shell proteins into the MCP. Is the shell a random assembly of BMC-domain proteins, or do particular BMC proteins occupy specific positions within the polyhedral shell? Modeling suggested that mixed sheets of the PduA and PduB proteins would allow their edges to align more precisely resulting in greater stability (Pang *et al.*, 2012). Thus, minor variations along the edges of BMC-domain proteins might guide assembly. In addition, prior studies showed that N-terminal extensions are necessary and sufficient for targeting enzymes to the MCP lumen and that these extensions interact with the termini of shell proteins (Fan *et al.*, 2010; 2012). Hence, it was proposed that the cargo enzymes might play a role in organizing the shell. In the case of the carboxysome, a scaffold protein (CcmN) links the cargo proteins to the shell and might serve a role in shell assembly (Cot *et al.*, 2008; Kinney *et al.*, 2012; Cameron *et al.*, 2013). On the other hand, studies have suggested that empty MCP shells can be produced by recombinant strains (although their organization is uncertain) (Parsons *et al.*, 2010). In addition, prior studies have indicated that certain hexameric BMC domain shell proteins are interchangeable (Cai *et al.*, 2015). Hexameric β-carboxysomal CcmK4 was shown to structurally complement a Δ*ccmK2* mutant of *Synechococcus elongatus* PCC 7942 (Cai *et al.*, 2015). Moreover; a chimeric carboxysomal shell was also constructed by replacing β-carboxysomal shell protein CcmK2 with the α-carboxysomal shell protein CsoS1. In the chimeric carboxysomal MCP the α-carboxysomal BMC hexamer fit well into a β-carboxysome shell because of their

structural similarity. In the present study, we found that PduJ was unable to compensate for the structural defects in a Δ*pduA* mutant, when expressed from a plasmid even though PduJ and PduA are very closely related in amino acid sequence (80% identical). Several explanations are possible for this finding: (i) plasmid expression might result in improper protein levels; (ii) the difference in amino acid sequence between PduJ and PduA might dictate interactions with other MCP proteins or result in structural variation or (iii) interactions with proteins encoded by adjacent chromosomal genes might be necessary for proper PduA function. It seems unlikely to us that protein production levels are the cause of the observed behaviour. The vector used in these studies (pLac22) has a low level of expression without IPTG (Warren *et al.*, 2000), and in our prior studies, pLac22 was successfully used in complementation tests with a number of different Pdu proteins including major and minor shell proteins as well as proteins whose dosage is critical to MCP assembly (Cheng *et al.*, 2011). We also found that PduA itself did not fully complement a Δ*pduA* mutant when produced from a plasmid (Fig. 5C) as was shown previously (Cheng *et al.*, 2011). Furthermore, partial complementation (the best result obtained) occurred at an intermediate level of IPTG suggesting that higher or lower levels of PduA expression would not have resulted in improved complementation. Thus, these results suggested that the sequence variation between PduA and PduJ does not fully account for the inability of ectopically produced PduJ to correct a Δ*pduA* mutation. Moreover, we determined the crystal structure of PduJ and found it to be nearly identical to the structure of PduA (Crowley *et al.*, 2010; Sinha *et al.*, 2014). On the other hand, we found that defects in MCP function that result from a Δ*pduA* mutation were largely corrected by replacing *pduA* with *pduJ* chromosomally. The growth phenotype and the MCP morphology of a Δ*pduA*::*pduJ*^{M} mutant were similar to wild-type (Figs. 6 and 7 and Supporting Information Fig. S2), and the DDH activity of MCPs purified from a Δ*pduA*::*pduJ*^{M} mutant was close to normal (Table 4). In addition, we showed that PduJ assumed a new function when expressed from the *pduA* locus. PduJ did not mediate 1,2-PD transport into MCPs when produced from its native chromosomal locus (Fig. 2B, Table 1), but it did so when produced from the *pduA* locus (Supporting Information Fig. S5). We propose that when PduJ is produced from the *pduA* locus, it interacts with new protein partners in a manner that changes its function. The *pdu* operon is transcribed as a polycistronic message (Fig. 1B) (Bobik *et al.*, 1999). The juxtaposition of the *pduA* and *pduB* genes will result in co-localization of their protein products in time and space during and immediately following translation and this might promote protein-protein interactions needed for proper MCP shell assembly. Thus, high sequence similarity allows PduJ to replace PduA and assume its transport function when expressed from the *pduA* locus but not when expressed from its native chromosomal position or from a plasmid. When PduJ is expressed from its native chromosomal position, it is apparently incorporated into the MCPs in a fashion that does not allow it to mediate 1,2-PD transport. The fact that PduJ changes function depending on the chromosomal position of its encoding gene strongly supports this model. At the present time it is unclear whether the observed phenomena reflect context-dependent differences in local protein conformation (e.g., in the pore of PduJ) or more global differences in the way subunits are arranged in the fully assembled MCP. Precedence for gene location influencing protein assembly was recently described for bacterial luciferase (LuxAB) (Shieh *et al.*, 2015). In this case, LuxA and LuxB were shown to assemble close to the site of synthesis, and localization of the *luxA* and *luxB* genes at distant chromosomal sites impaired formation of the LuxAB heterodimer. Importantly, the studies reported here show for the first time that gene location influences the interactions among MCP proteins and is critical to the higher order structure and function of bacterial MCPs. They also provide an example of how gene position can guide assembly and even determine protein function in large multi-protein complexes.

Experimental procedures

Chemicals and reagents

Antibiotics, vitamin B_{12} (CN-B_{12}), coenzyme B_{12} (AdoB_{12}), NAD^+, NADH, $NADP^+$, and NADPH are from Sigma-Aldrich (St. Louis, MO). *KOD* DNA polymerase and restriction enzymes and T4 ligase were respectively from Novagen (Cambridge, MA) and New England Biolabs (Beverly, MA). Isopropyl-β-D-1-thiogalactopyranoside (IPTG) was from Diagnostic Chemicals Limited (Charlotteville, PEI, Canada. Choice *Taq* Blue Mastermix was from Denville Scientific (South Plainfield, NJ). Other chemicals were from Fisher Scientific (Pittsburgh, PA).

Bacterial strains, media and growth conditions

The bacterial strains used are listed in Table 5. All strains are derivatives of *Salmonella enterica* serovar Typhimurium strain LT2, referred to as LT2. The rich medium used was Luria-Bertani/Lennox (lysogeny broth) (LB) medium (Difco, Detroit, MI). For selecting transformants TYE media was used (Chang and Cronan, 1982) where bacto-tryptone, yeast extract and agar were purchased from Difco Laboratories (Difco, Detroit, MI). The minimal medium used was no carbon-E (NCE) medium supplemented with 1 mM $MgSO_4$, 0.3 mM each of valine, isoleucine, leucine, and threonine and 50 μM ferric citrate (Berkowitz *et al.*, 1968; Price-Carter *et al.*, 2001; Cheng *et al.*, 2011).

Table 5. Bacterial strains used in this study.

Strains	Genotype/Variant	Source
WT	*Salmonella enterica* serovar Typhimurium strain LT2	Lab collection
BE293	LT2/pKD46	Lab collection
BE464	LT2, Δ*pduA::frt*	Lab collection
BE184	LT2, Δ*pduJ*	Lab collection
CC33	LT2, Δ*pduJ* Δ*pduA::frt*	This study
CC3	LT2, PduAS40L	This study
CC17	LT2, PduJS39L	This study
CC32	PduAS40L-PduJS39L	This study
BE287	LT2/pLac22	Lab collection
BE1351	LT2, Δ*pduA::frt*/pLac22	Lab collection
TA1414	LT2, Δ*pduJ*/pLac22	Lab collection
BE1350	LT2, Δ*pduA::frt*/pLac22-*pduA*	Lab collection
CC16	LT2, Δ*pduA::frt*/pLac22-*pduJ*	This study
CC10	LT2, Δ*pduJ*/pLac22-*pduA*	This study
TA1412	LT2, Δ*pduJ*/*pLac22-pduJ*	Lab collection
CC76	LT2, Δ*pduA::pduJ*^{M}	This study
CC79	LT2, Δ*pduA::pduJ*^{M}*S39L*	This study

Construction of chromosomal mutations

Chromosomal *pduA* and *pduJ* deletions and the point mutations surrounding the pore region were constructed by recombineering as described (Datsenko and Wanner, 2000; Sinha *et al.*, 2014; Chowdhury *et al.*, 2015). All the mutants were confirmed by sequencing.

P22 Transduction

In order to make double mutants, the *sacB-cat* cassette at *pduA* loci was moved to a Δ*pduJ* mutant by transduction as previously described, using P22 HT105/1 *int-210* (Schmieger, 1971). Transductants were tested for phage contamination and sensitivity by streaking on green plates against H5- *int*. Finally, the *sacB-cat* cassette was replaced by recombination with single stranded oligo having the desired mutation. Deletion mutations were confirmed by PCR followed by DNA sequencing.

Construction of clones for complementation studies

The coding sequences of *pduA*, and *pduJ* were cloned into pLac22 (Warren *et al.*, 2000; Cheng *et al.*, 2011) using *Bgl* II and *Hind* III restriction sites. The DNA sequences of all clones were verified. The recombinant plasmids were transformed into respective deletion mutants using electroporation (Cheng *et al.*, 2011).

Chromosomal replacement of pduA *with* pduJ

To determine if chromosomal position has any effect on complementation, the chromosomal *pduA* gene was replaced with the *pduJ* gene using a recombineering method (Datsenko and Wanner, 2000; Sinha *et al.*, 2014; Chowdhury *et al.*, 2015). A codon-modified *pduJ* gene (*pduJ*^{M}) was designed using a codon optimization tool from Integrated DNA Technologies, (Coralville, IA). A Shine Dalgarno sequence was placed after the stop codon of *pduJ*^{M} to assure proper expression of downstream genes. The *pduJ*^{M} gene with needed flanking sequences was synthesized by Integrated DNA Technolgies as a gBlock with the sequence shown in Supporting Information Fig. S6. This construct was used to transform the cells where a *sacB-cat* cassette was inserted at *pduA* locus in a strain expressing the λ-red recombinase from pKD46 (Datsenko and Wanner, 2000). Positive transformants were confirmed through DNA sequencing.

Growth studies, MCP Purification and Diol dehydratase activity

Growth studies were performed at limiting CN-B_{12} (25 nM) as previously described using a Synergy HT Microplate reader (BioTek, Winooski, VT) (Cheng *et al.*, 2011). Experiments were done in triplicate. Pdu MCPs were purified according to a prior protocol (Sinha *et al.*, 2012; Chowdhury *et al.*, 2015). Their protein content was verified by SDS PAGE. To check MCP integrity, the purified MCPs were negatively stained with uranyl acetate (2%) and visualized using a transmission electron microscope (JEOL 2100, Peabody, MA) as described (Sinha *et al.*, 2012; Chowdhury *et al.*, 2015). Purified MCPs were assayed for diol dehydratase using the coupled NADH-dependent alcohol dehydrogenase reaction as described (Fan and Bobik, 2011; Chowdhury *et al.*, 2015).

Protein identification by LC-MS/MS

To verify the incorporation of PduJ^{M} into microcompartment shell, the MCP particles were purified and fractionated by SDS PAGE. Protein band of interest was excised from a SDS-polyacrylamide gel, cut into small pieces and distained with mixture of 25 mM ammonium bicarbonate and acetonitrile (ACN) (v/v 1:1) according to the protocol described earlier (Shevchenko *et al.*, 2006). The proteins inside the gel were reduced with 1 mM DTT and cysteine residues were alkylated with iodoacetamide. Trypsin (proteomics grade, Promega) was added to the gel pieces and incubated for 16–18 h according to a published procedure (Shevchenko *et al.*, 2006). The trypsin digested peptides were extracted and subjected to liquid chromatography mass spectrometry analysis according to the manufacturer's instructions (Thermo, Q Exactive™ Hybrid Quadrupole-Obrbitrap Mass Spectrometer). Tandem mass spectra were searched against the known protein sequence from *S. enterica* serovar Typhimurium LT2 using both MASCOT and Sequest HT (Thermo Proteome discoverer software from Thermo Fisher) with maximum mixed cleavage sites 2, fixed carbamidomethylation and variable methionine oxidation. The precursor mass tolerance was set to 5 ppm and fragment mass tolerance was 0.05 Da. Maximum delta Cn value was set to 0.05.

PduJ structure determination. For ease of purification and characterization a PduJ K25A mutant was constructed by site-directed mutagenesis method. The mutation greatly increased the solubility of the PduJ protein as was previously observed for the analogous mutation in PduA (K26A). The PduJ K25A and PduA K26A mutations (which are on

the edges of the hexamers) reduce interactions that lead to protein sheet formation at low protein concentrations (Sinha *et al.*, 2014). To construct this mutation, the coding sequence of full-length *pduJ* of *S. enterica* serovar Typhimurium LT2 was cloned into pET22b(+) vector (Novagen). Then, site-directed mutagenesis was performed using a QuikChange kit (Stratagene) to generate the PduJ K25A mutant. The construct was confirmed by DNA sequencing.

PduJ K25A construct was transformed in BL21 CodonPlus(DE3)-RIL *E. coli* cells (Stratagene), grown in LB media with ampicillin (100 μg ml^{-1}) and chloramphenicol (34 μg/ml) at 37°C, shaking at 225 rpm. When cell cultures reached an optical density at wavelength 600 nm (OD_{600}) of 0.4, temperature was dropped to 30°C; and at OD_{600} of 0.6, protein expression was induced with 0.5 mM isopropyl-D-1-thiogalactopyranoside (IPTG) for 4 h. Cells were pelleted at 4,000 r.p.m. for 15 min and resuspended (15 g cell pellet per 35 ml total) in 50 mM Tris-HCl and 200 mM NaCl, pH 9.0 (Buffer A). Resuspended cells were stored at −80°C until purification.

Before purification the cells were thawed and homogenized in the presence of lysozyme, protease inhibitor cocktail, DNase I and 1 mM $MgCl_2$. Cells were lysed by sonication in three cycles, increasing amplitude from 40 to 60% per cycle. Cell lysate was centrifuged at 17,000 × g for 30 min and the supernatant was clarified by 0.45μm filtration. The clarified supernatant was loaded on a HisTrap column (CV = 5 ml, GE Healthcare). The column was washed with 10% Buffer B (50 mM Tris-HCl, 200 mM NaCl, 500 mM Imidazole, pH 8.0) and eluted with a linear gradient (Δ90% Buffer B/45ml of elution). PduJ K25A construct was eluted between 30 and 50% Buffer B. Fractions containing the fairly pure and large quantities of protein migrating at 10 kDa were pooled and dialyzed into crystallization buffer (30 mM Tris. HCl, 50 mM NaCl, 1% v/v glycerol, pH 8.0) overnight at 4°C. PduJ K25A construct was then concentrated to 100–250 mg/ml each, flash frozen with liquid N_2, and stored at − 80°C.

Crystallization trials were set up in 96-well plate by hanging drop method including screens from Emerald Biosystems, Hampton Research, and Qiagen. PduJ K25A (65.7 mg ml^{-1}) crystallized in hexagonal plate crystals in multiple conditions. PduJ K25A crystals that were used for structure determination were optimized from 0.1 MES sodium salt pH 6.5, 2.0 M ammonium sulfate, 5% PEG 400 (w/v).

Before data collection, protein crystals were first cryoprotected with 35% glycerol to 65% well solution. An initial data set was collected in-house and the data were integrated, merged, and scaled using the XDS suite programs for crystallography. A molecular replacement solution was obtained with PHASER using PduA (PDB ID: 3NGK) as a search model. A higher resolution data set was later collected at the Advanced Photon Source, beamline 24-ID-C, at Argonne National Laboratories, IL. Refinement with the new data set began with rigid body refinement using coordinates from the previous molecular replacement.

Acknowledgements

The authors thank the Iowa State University microscopy and nano-imaging facility for help in electron microscopy. The authors also acknowledge the ISU DNA sequencing and synthesis facility and Protein facility for DNA sequence and LC-MS/MS analyses. The authors also thank Michael Collazo and Duilio Cascio for assistance in protein crystallization and data collection, and the synchrotron staff at APS beamline 24-ID-C. This work was supported by NIH grant R01AI081146 (T.O.Y. and T.A.B.). Work in the Yeates lab was supported by the BER program of the DOE Office of Science. S.C. was supported by a CBI NIH training grant (T32-GM008496) and the by the UCLA Graduate Division. We thank M. Capel, K. Rajashankar, N. Sukumar, J. Schuermann, I. Kourinov, and F. Murphy (NECAT Beamline 24-ID-C at Advanced Photon Source, which is supported by National Center for Research Resources Grant 5P41RR015301-10 and National Institute of General Medical Sciences Grant 8 P41 GM103403-10 from the National Institutes of Health). Use of the Advanced Photon Source is supported by the Department of Energy under Contract DE-AC02-06CH11357.

References

Abdul-Rahman, F., Petit, E., and Blanchard, J.L. (2013) The distribution of polyhedral bacterial microcompartments suggests frequent horizontal transfer and operon reassembly. *J Phylogen Evol Biol* **1**: 1–7. doi 10.4172/2329-9002.1000118

Berkowitz, D., Hushon, J.M., Whitfield, H.J., Jr., Roth, J. and Ames, B.N. (1968) Procedure for identifying nonsense mutations. *J Bacteriol* **96**: 215–220.

Bobik, T.A., Havemann, G.D., Busch, R.J., Williams, D.S., and 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 B_{12}-dependent 1, 2-propanediol degradation. *J Bacteriol* **181**: 5967–5975.

Bobik, T.A., Lehman, B.P., and Yeates, T.O. (2015) Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways. *Mol Microbiol* **98**: 193–207.

Cai, F., Sutter, M., Bernstein, S.L., Kinney, J.N., and Kerfeld, C.A. (2015) Engineering bacterial microcompartment shells: chimeric shell proteins and chimeric carboxysome shells. *ACS Synth Biol* **4**: 444–453.

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

Cameron, J.C., Wilson, S.C., Bernstein, S.L., and Kerfeld, C.A. (2013) Biogenesis of a bacterial organelle: the carboxysome assembly pathway. *Cell* **155**: 1131–1140.

Chang, Y.Y., and Cronan, J.E. Jr. (1982) Mapping nonselectable genes of *Escherichia coli* by using transposon Tn10: location of a gene affecting pyruvate oxidase. *J Bacteriol* **151**: 1279–1289.

Cheng, S., Liu, Y., Crowley, C.S., Yeates, T.O., and Bobik, T.A. (2008) Bacterial microcompartments: their properties and paradoxes. *Bioessays* **30**: 1084–1095.

Cheng, S., Sinha, S., Fan, C., Liu, Y., and Bobik, T.A. (2011) Genetic analysis of the protein shell of the

microcompartments involved in coenzyme B_{12}-dependent 1,2-propanediol degradation by *Salmonella*. *J Bacteriol* **193**: 1385–1392.

Chowdhury, C., Chun, S., Pang, A., Sawaya, M.R., Sinha, S., Yeates, T.O., and Bobik, T.A. (2015) Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proc Natl Acad Sci USA* **112**: 2990–2995.

Chowdhury, C., Sinha, S., Chun, S., Yeates, T.O., and Bobik, T.A. (2014) Diverse bacterial microcompartment organelles. *Microbiol Mol Biol Rev* **78**: 438–468.

Cot, S.S., So, A.K., and Espie, G.S. (2008) A multiprotein bicarbonate dehydration complex essential to carboxysome function in cyanobacteria. *J Bacteriol* **190**: 936–945.

Craciun, S., and Balskus, E.P. (2012) Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proc Natl Acad Sci USA* **109**: 21307–21312.

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

Crowley, C.S., Sawaya, M.R., Bobik, T.A., and Yeates, T.O. (2008) *Structure of the PduU shell protein from the Pdu microcompartment of Salmonella*. *Structure* **16**: 1324–1332.

Datsenko, K.A., and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* **97**: 6640–6645.

Erbilgin, O., McDonald, K.L., and Kerfeld, C.A. (2014) Characterization of a planctomycetal organelle: a novel bacterial microcompartment for the aerobic degradation of plant saccharides. *Appl Environ Microbiol* **80**: 2193–2205.

Fan, C., and Bobik, T.A. (2011) The N-terminal region of the medium subunit (PduD) packages adenosylcobalamin-dependent diol dehydratase (PduCDE) into the Pdu microcompartment. *J Bacteriol* **193**: 5623–5628.

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

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

Havemann, G.D., and Bobik, T.A. (2003) Protein content of polyhedral organelles involved in coenzyme B_{12}-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar Typhimurium LT2. *J Bacteriol* **185**: 5086–5095.

Havemann, G.D., Sampson, E.M., and Bobik, T.A. (2002) PduA is a shell protein of polyhedral organelles involved in coenzyme B_{12}-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar typhimurium LT2. *J Bacteriol* **184**: 1253–1261.

Heldt, D., Frank, S., Seyedarabi, A., Ladikis, D., Parsons, J.B., Warren, M.J., and 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.

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

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

Kinney, J.N., Salmeen, A., Cai, F., and Kerfeld, C.A. (2012) Elucidating essential role of conserved carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly. *J Biol Chem* **287**: 17729–17736.

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

Kuehl, J.V., Price, M.N., Ray, J., Wetmore, K.M., Esquivel, Z., Kazakov, A.E., *et al.* (2014) Functional genomics with a comprehensive library of transposon mutants for the sulfate-reducing bacterium *Desulfovibrio alaskensis* G20. *MBio* **5**: e01041–e01014.

Pang, A., Liang, M., Prentice, M.B., and 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.

Pang, A., Warren, M.J., and 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.

Parsons, J.B., Frank, S., Bhella, D., Liang, M., Prentice, M.B., Mulvihill, D.P., and 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.

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

Petit, E., LaTouf, W.G., Coppi, M.V., Warnick, T.A., Currie, D., Romashko, I., Deshpande, S., *et al.* (2013) Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by *Clostridium phytofermentans*. *PLoS One* **8**: e54337.

Price-Carter, M., Tingey, J., Bobik, T.A., and Roth, J.R. (2001) The alternative electron acceptor tetrathionate supports B_{12}-dependent anaerobic growth of *Salmonella enterica* serovar typhimurium on ethanolamine or 1,2-propanediol. *J Bacteriol* **183**: 2463–2475.

Price, G.D., and Badger, M.R. (1989) Isolation and characterization of high CO_2-requiring-mutants of the cyanobacterium *Synechococcus* PCC7942: two phenotypes that accumulate inorganic carbon but are apparently unable to generate CO_2 within the Carboxysome. *Plant Physiol* **91**: 514–525.

Rae, B.D., Long, B.M., Badger, M.R., and Price, G.D. (2013) Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO_2 fixation in cyanobacteria and some proteobacteria. *Microbiol Mol Biol Rev* **77**: 357–379.

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

Sampson, E.M., and Bobik, T.A. (2008) Microcompartments for B_{12}-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. *J Bacteriol* **190**: 2966–2971.

Schmieger, H., (1971) A method for detection of phage mutants with altered transducing ability. *Mol Gen Genet* **110**: 378–381.

Scott, K.P., Martin, J.C., Campbell, G., Mayer, C.D., and Flint, H.J. (2006) Whole-genome transcription profiling reveals genes up-regulated by growth on fucose in the human gut bacterium "*Roseburia inulinivorans*". *J Bacteriol* **188**: 4340–4349.

Seedorf, H., Fricke, W.F., Veith, B., Bruggemann, H., Liesegang, H., Strittmatter, A., *et al.* (2008) The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic features. *Proc Natl Acad Sci USA* **105**: 2128–2133.

Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V., and Mann, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1**: 2856–2860.

Shieh, Y.W., Minguez, P., Bork, P., Auburger, J.J., Guilbride, D.L., Kramer, G., and Bukau, B. (2015) Operon structure and cotranslational subunit association direct protein assembly in bacteria. *Science* **350**: 678–680.

Sinha, S., Cheng, S., Fan, C., and Bobik, T.A. (2012) The PduM protein is a structural component of the microcompartments involved in coenzyme B_{12}-dependent 1,2-propanediol degradation by *Salmonella enterica*. *J Bacteriol* **194**: 1912–1918.

Sinha, S., Cheng, S., Sung, Y.W., McNamara, D.E., Sawaya, M.R., Yeates, T.O., and 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* **426**: 2328–2345.

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

Tanaka, S., Sawaya, M.R., Phillips, M., and Yeates, T.O. (2009) Insights from multiple structures of the shell proteins from the b-carboxysome. *Protein Sci* **18**: 108–120.

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

Thompson, M.C., Cascio, D., Leibly, D.J., and Yeates, T.O. (2015) An allosteric model for control of pore opening by substrate binding in the EutL microcompartment shell protein. *Protein Sci* **24**: 956–975.

Warren, J.W., Walker, J.R., Roth, J.R., and Altman, E. (2000) Construction and characterization of a highly regulable expression vector, pLAC11, and its multipurpose derivatives, pLAC22 and pLAC33. *Plasmid* **44**: 138–151.

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

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

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Supplementary Information

Table S1. Data collection and refinement statistics

	PduJ K25A
Data collection	
Space group	P 6
Cell dimensions	
a, *b*, *c* (Å)	67.9 67.9 26.5
α, *β*, *γ* (°)	90, 90, 120
Wavelength (Å)	0.97950
Resolution (Å)	58.78-1.50 (1.54-1.50) *
R_{sym} or R_{merge}	0.062(0.564)
$I/\sigma I$	18.1(2.79)
Completeness (%)	99.9(99.5)
Redundancy	9.2(5.5)
Refinement	
Resolution (Å)	58.78-1.50
No. of unique reflections	11398
R_{work}/ R_{free}	.191/.219
No. atoms	659
Protein	626
Ions	0
Water	33
B-factors	
Protein	21.6
Ions	0
Water	27.1
R.m.s deviations	
Bond lengths (Å)	.013
Bond angles (°)	1.7

This structure was determined from one crystal.

*Highest resolution shell is shown in parenthesis.

Table S2. The peptides identified by LC-MS/MS

Peptide Sequence	Protein Group	m/z (Da)	Ion Score	Modification	XCorr
AAVDAGSAAASVVGEVK	PduJ/PduJ^{M}	751.39667	127	-	4.62
IGSGLVTVMVR	PduJ	566.33209	103		3.45
GDVGAVKAAVDAGSAAASVVGEVK	PduJ/PduJ^{M}	710.04620	101	-	4.26
IGLGLVTVMVR	PduJ^{M}	579.35773	93	-	2.91
GLVGAIEAADAMVK	PduJ/PduJ^{M}	672.86359	92	-	3.50
MNNALGLVETK	PduJ/PduJ^{M}	603.31293	88	M1 (Oxidation)	3.52
GLVGAIEAADAMVK	PduJ/PduJ^{M}	680.86102	81	M12(Oxidation)	3.47
IGSGLVTVMVR	PduJ	574.32922	80	M9(Oxidation)	3.16
SANVQLVGYEK	PduJ/PduJ^{M}	604.31940	74	-	3.25
IGLGLVTVMVR	PduJ^{M}	587.35553	71	M9(Oxidation)	2.70
SCHVIPRPHSDVEAILPK	PduJ/PduJ^{M}	1028.04248	57	C2(Carbamidomethyl)	4.55
MNNALGLVETK	PduJ/PduJ^{M}	595.31616	57	-	3.09

Foot note: The peptide (IGLGLVTVMVR) comes from PduJ^{M} S39L. The leucine (underlined) corresponds to the leucine at position 39 in PduJ^{M} S39L

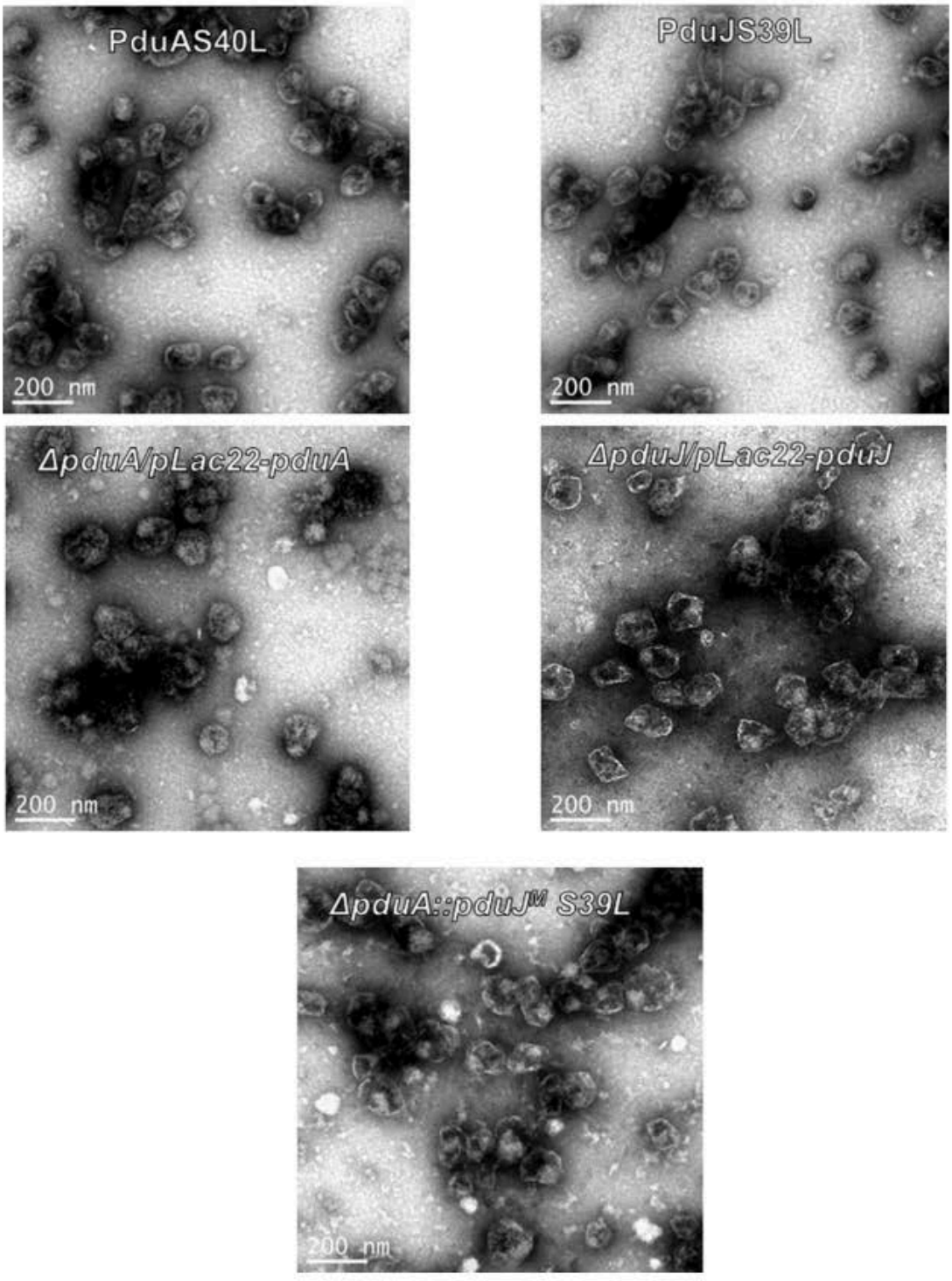


Figure S1. Transmission electron microcopy of purified MCPs. (A) The strain used for MCP purification is indicated in each panel.

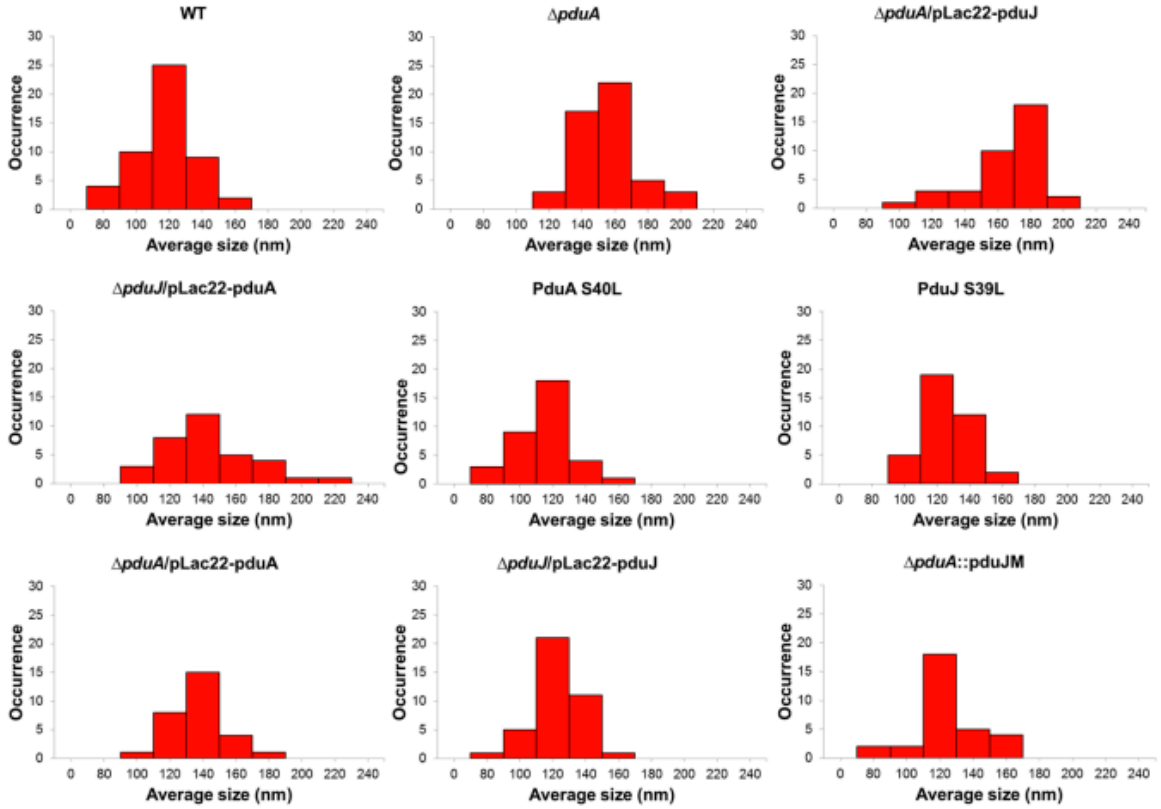


Figure S2. Size distribution of the purified MCPs from different mutants. MCPs were measured using ImageJ software as described (Cheng *et al.*, 2011). Measurements were grouped into 20 nM bins. Bars are centered on the middle value of a bin. The x-axis is the number of MCPs whose size fall within a particular bin. *ΔpduJ* produced elongated and filamentous MCPs and therefore, their sizes were not evaluated.

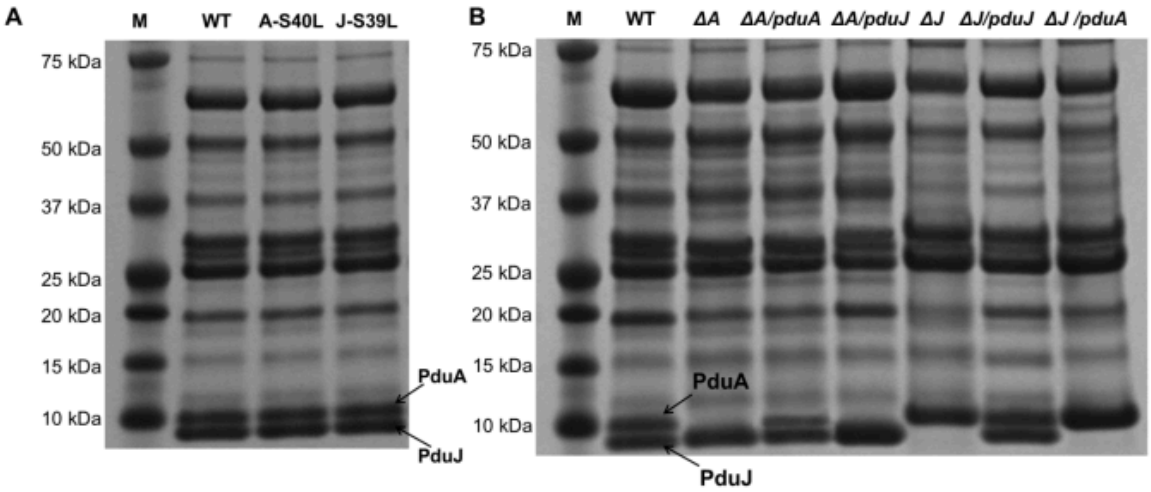
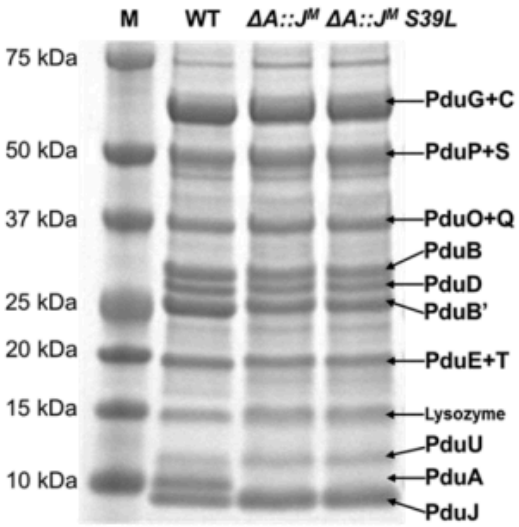


Figure S3. SDS PAGE and of purified MCPs. (A) MCPs purified from PduA S40L and PduJ S39L mutants have a protein profile similar to wild type. **(B)** Deletion of the *pduA* or *pduJ* gene (both encode major shell proteins) resulted in some changes in the relative amounts of proteins present in purified MCPs. This gel also showed that both PduA and PduJ were incorporated into MCPs when produced from plasmid pLac22 (*ΔA* indicates *ΔpduA,* and *ΔA/pduA* indicates *ΔpduA*/pLac22-*pduA*). Approximately 15 µg protein was loaded in each well. NuPAGE 4-12% gels were used according to the manufacturer's instructions (Life Technologies) and were stained with Simplyblue Safe stain (Life Technologies)



Protein bands	Percentage of total proteins (%)		
	WT	*ΔpduA::J^M*	*ΔpduA::J^M-S39L*
PduG+C	13.2 ± 1.2	12.8 ± 0.8	12.5 ± 0.7
PduP+S	11.4 ± 0.5	12.4 ± 0.7	12.2 ± 0.5
PduO+Q	10.7 ± 0.8	10.3 ± 0.9	10.4 ± 0.1
PduB	12.2 ± 0.5	11.6 ± 1.1	11.4 ± 0.8
PduD	6.9 ± 0.5	6.5 ± 0.5	6.2 ± 0.5
PduB'	12.4 ± 0.1	11.1 ± 0.5	11.3 ± 0.7
PduE+T	7.2 ± 0.8	6.5 ± 0.8	6.6 ± 1.1
PduU	2.4 ± 0.2	2.6 ± 0.1	2.5 ± 0.4
PduA	9.2 ± 0.4	-	-
PduJ	10.5 ± 0.2	13.2 ± 0.9	13.4 ± 0.3

Figure S4. SDS-PAGE and densitometry of wild-type (WT) and the key mutants used in this study. The abundance of each MCP protein is shown as a percent of total MCP protein. The values for overlapping bands were added together. For example, PduG+C indicates the percent of total MCP protein contributed by PduG and PduC combined. Image J was used to measure band intensities. The total protein was the sum of the intensities determined for all the bands. A linear response was obtained with loading levels from 5 to 15 µg protein. The values shown above were obtained when approximately 10 µg purified MCPs was used and run on NuPAGE 4-12% gel (Life Technologies) for 1 h at 150 V that were stained with Simplyblue Safe stain (Life Technologies). ImageJ was used according to the online tutorial https://imagej.nih.gov/ij/docs/

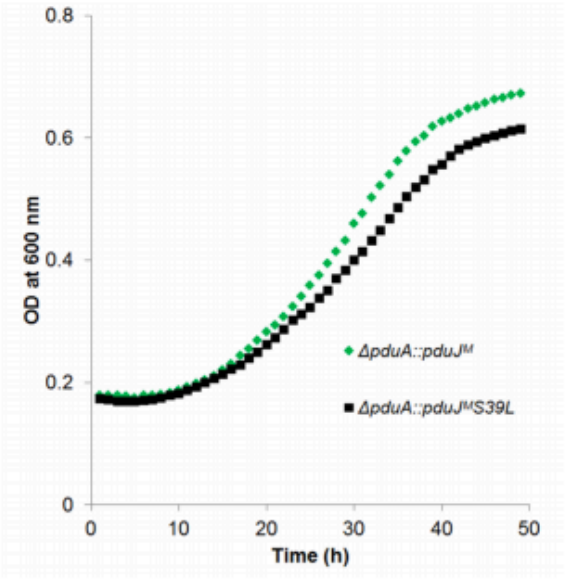


Figure S5. Effect of pore mutation in PduJ^M on growth in 1, 2-PD minimal media. A *ΔpduA::pduJ^{M}S39L* mutant grows slightly slower than the parent strain. Growth assays were repeated in triplicate and representative curves are shown. The *ΔpduA::pduJ^{M}S39L* mutant grew slightly slower that the *ΔpduA::pduJ^M* parent in all cases.

ttctttttcc tcgcatcttc ttatagtccc aactatcgga acactccatg cgaggtcttt ATGAACAATG
CCCTGGGCCT GGTGGAGACC AAAGGCCTTG TGGGTGCGAT CGAAGCAGCC GACGCGATGG TTAAAAGCGC
GAACGTCCAG CTGGTGGGTT ATGAAAAAAT TGGGTCGGGA TTGGTGACGG TTATGGTACG TGGCGATGTC
GGAGCGGTCA AGGCTGCGGT TGATGCCGGC AGCGCGGCAG CGTCTGTCGT CGGCGAGGTT AAATCGTGCC
ACGTCATCCC CCGTCCGCAC TCAGACGTCG AGGCTATTTT ACCAAAATCC GCGTAAagga ggttagccaa
tgagcagcaa tgagctggtg gaacagatca tggcgcaggt g

Figure S6. Nucleotide sequence of *pduJ*^{M} amplicon used in recombineering. The sequence of gBlock is in blue font; the coding sequence of *pduJ*^{M} is shown capital letters. The flanking regions used for recombination are underlined.

Cheng, S., Sinha, S., Fan, C., Liu, Y. and Bobik, T.A., (2011) Genetic analysis of the protein shell of the microcompartments involved in coenzyme B_{12}-dependent 1,2-propanediol degradation by *Salmonella*. *J Bacteriol* **193**: 1385-1392.

Chapter 8

N-term tail targeting of enzymes to Pdu MCP lumen

Introduction

Fan et al. showed that the first 18 aa of the N-terminal tail of PduP and PduD are sufficient for targeting to the Pdu MCP lumen via interaction with short C-terminal regions on PduA and PduJ[39,40,41]. For example, the fragment was successfully used to target GFP, MBP and GST to the MCP lumen. In addition, multiple sequence alignment showed that this N-terminal targeting fragment was found on other MCP proteins including PduE, and EutCG from the Eut MCP. Key residues of the N-terminal tail of PduP were E7, I10, and L14, whereas H81, V84, and L88 are key to the C-terminal tail of PduA. This was also the first evidence indicating that the Pdu MCP shell may act as a scaffold upon which to organize internalized enzymes and that the orientation of the PduA hexamer is concave-side-facing-in due to the orientation of the C-terminus in solved crystal structures of PduA. In light of solving the hexamer assembly structure of PduA and PduJ, we were interested in characterizing their binding modes and binding affinities to the N-terminal peptides of PduCDELP which were identified bioinformatically by Jorda et al (Figure 8.1)[42]. This information would shed insight on Pdu MCP biogenesis, lumenal protein organization and distinguish functional roles between the highly homologous PduA and PduJ (backbone rmsd 1Å).

Jorda et al. predicted a peptide-binding-protein mode using the program Auto-Dock Vina to approximate the binding mode and FlexPepDock to refine the Vina model[42]. Coordinate files were prepared by combining the coordinates of the biological assembly of hexamer protein, PduA (PDB ID: 3NGK) or PduJ (PDB ID: 5D6V), and coordinates of the peptide as short

amphipathic α-helices predicted by Pep-Fold. Using this method, shell proteins PduA and PduJ were predicted to have similar binding modes for the PduP peptide (Figure 8.2). This binding mode suggests that the N-tail peptides bind on the lumenal side of the hexamer protein assembly in grooves between adjacent protomers. This supports the model of BMC biogenesis in the Pdu MCP as in the carboxysome, where the lumenal core may assemble first and recruit the BMC shell proteins via affinity to the lumenal protein N-terminal tail peptides[43].

To reach a more quantitative understanding of Pdu MCP enzyme and shell protein binding, I designed experiments for measuring binding affinities by gel electrophoresis mobility shift assay (EMSA) and fluorescence anisotropy or polarization. Previous attempts using isothermal calorimetry and thermal melting with Sypro Orange did not give interpretable data due to protein aggregation and precipitation upon perturbation, despite varying temperature, spinning speeds, or protein concentration. Here, I present the results of the Pdu Ntail peptide binding studies.

Results

For EMSA, I designed N-tail peptides from PduDEP fused to the N-terminus of superfolder GFP V206K, a mutation that abrogates the native dimeric interaction and results in monomeric protein, with variable linker regions consisting of 0, 2, 5, or 10 glycines (Figure 8.3). Constructs were made using Gibson Assembly cloning into pET22b+ vector (Invitrogen) and transformed into BL21 Rosetta DE3 *E. coli* expression strain for expression. Proteins were purified by Ni^{2+}-affinity purification using C-terminal 6xHis tag on superfolder GFP V206K (pI 4.75). Constructs were then mixed with PduA (pI 7.75) or PduJ (pI 7.75) and run on native gels (running gel at pH 9.0, stacking gel at pH 8.0) in Tris-glycine buffer pH 6.8. Gels were analyzed

by fluorescence scanning using Pharos FX Imager and Coomassie staining. Initial screening to optimize linker length showed that longest linker, 10 glycine residues, produced the best interpretable result indicating binding between PduA and the peptides (Figure 8.4).

Titration of $PduP^{1-18}$-Gly_{10}-sfGFP into PduA (50μM) resulted in upward shift of band with increasing peptide concentration, indicating binding to retard migration of the complex (Figure 8.5A). No shift was observed with PduA only (Figure 8.5B). However, a similar pattern of shift was observed with increasing concentration of $PduP^{1-18}$-Gly_{10}-sfGFP (Figure 8.5C). The shift in $PduP^{1-18}$-Gly_{10}-sfGFP alone may be due to aggregation of the amphipathic peptide, which is functional in one possible mechanism for the native assembly of the BMC core via N-terminal peptide aggregation thereby recruiting the BMC shell[44]. The shift was also observed in $PduP^{1-18}$-Gly_2-sfGFP alone (Figure 8.5D). As results from the binding assay would be convoluted by the peptide's self-association, further study by EMSA was not pursued.

As an independent and alternative approach, I decided to use fluorescence polarization to measure binding of fluorophore-conjugated peptides to shell proteins. Fluorescence polarization,

$$P = \frac{I_{||} - I_{\perp}}{I_{||} + I_{\perp}}$$

can be used to measure binding affinity by reflecting the relative tumbling rates of the fluorophore-conjugated peptide upon binding the shell protein, which causes decrease in tumbling and increase in P or r (Figure 8.6). PduCDELP peptides and peptides with an appended C-terminal sequence KCK for single cysteine fluorophore conjugation were ordered from GenScript (Table 8.1). I designed a protocol for labeling peptides with a small molecule fluorophore such as cyanine, Alexa Fluor®, or rhodamine. Small reaction screening showed

Tetramethylrhodamine-5-maleimide (5-TAMRA; 480 Da, excitation/emission=544nm/570nm) was optimal for fluorophore conjugation. Peptides with KCK were labeled in the reaction: 100mM HEPES pH 7.5, 1mM TCEP pH 7.5, 3.5mM 5-TAMRA (3-fold excess, 1.1% DMSO final), 2.0mg peptide. Reactions were sonicated in a water bath for 3h and allowed to react for 6 days, monitoring reaction progress per day by matrix-assisted laser desorption/ionization (MALDI) time of flight (TOF) mass spectrometry (MS) (Figure 8.7). Only PduP did not label well due to insolubility issues despite varying DMSO and acetonitrile conditions. PduCDEL labeled peptides were purified by reverse-phase HPLC (Higgins Analytical, Inc; Proto 300 C18 column) with a methanol gradient and peak fractions containing the labeled species were lyophilized.

Fluorescence anisotropy was measured on a PTI QuantaMaster Spectrofluorimeter at excitation 544nm and emission 572nm. Initial measurements gave 5-TAMRA-PduC N-tail peptide (5uM) and PduA (20uM) with P= 0.115 +/- 0.004; 5-TAMRA-PduC N-tail peptide alone (5uM) with P= 0.109 +/- 0.006. Titration experiments using the Tecan M1000 plate reader equipped with monochromators for both excitation and emission wavelengths allowed a thorough investigation of binding by measurement of fluorescence polarization. The Tecan M1000 plate reader allowed excitation at 530nm and detected emission at 572nm, bandwidth 5nm. The emission wavelength and concentration of TAMRA-labeled peptide was selected by optimization at 1μM (Figure 8.8).

By fluorescence polarization (FP), PduD Ntail peptide exhibited binding of high affinity to both PduA and PduJ with K_D of 170nM and 19nM, respectively (Figure 8.9). To test the specificity of binding, PduA and labeled PduD peptide were mixed at saturation of binding and

titrated with unlabeled PduD peptide. Contrary to expected decrease in FP, the signal increased in a manner characteristic of binding (Figure 8.10). Indeed, labeled PduD peptide titrated with unlabeled PduD peptide only results in binding with low affinity (K_D = 5μM) (Figure 8.11). Instead, specificity of binding was tested using PduA and PduJ mutants in which mutations were made to the C-terminal helices to disrupt binding to the Ntail peptides: mutation of residues H81, V84, L88, PduA ctm or PduJ ctm; and truncation of the C-termini to the first 80 amino acids of PduA, PduA trunc1-80 (or 79 amino acids in PduJ, PduJ trunc1-79)[45]. The C-terminal mutants of PduA and PduJ exhibited decreased affinity in binding to the PduD Ntail peptide (Figure 8.12). From these results, I conclude that PduD Ntail peptide binds to PduA and PduJ, and that PduD also has affinity for binding itself.

The Pdu enzyme Ntail binding is not likely to be a coincidence. Early literature on the study of the Pdu enzymes describes the amphipathicity of these N-terminal peptides and their role in decreasing protein solubility[46–48]. It is also the basis for one of the popular proposed models of bacterial microcompartment assembly in which a pre-core assembles first by the aggregation induced by amphipathic N-terminal helices on enzymes and the subsequent recruitment of the shell proteins also by interaction with the N-terminal helices[44]. Although only PduD was shown to be necessary and sufficient for PduCDE pulldown in earlier experiments, this insight prompted me to also test the self-binding properties of other Pdu enzyme Ntail peptides. It appears that PduD and PduEL, to a lesser extent, exhibit self-binding (Figure 8.13). However, PduC does not have a self-interaction characteristic of binding. This is in agreement with findings that PduC did not need N-terminal truncation in order to facilitate ease of biochemical study. The self-interactions of the Pdu enzyme N-terminal peptides lends support to the pre-core assembly model in bacterial microcompartment assembly.

Unless otherwise noted, all peptide binding fluorescence polarization measurements were carried out in triplicate at peptide concentration 1μM, excitation wavelength 530nm, emission wavelength 572nm, in 30mM TRIS pH 8.0, 50mM NaCl, on a 96-well black optical bottom plate format on a Tecan M1000 plate reader instrument. All Pdu shell proteins were cloned and purified as edge mutants (PduA K26A and PduJ K25A) for ease of use.

Discussion

From these preliminary results, I conclude that PduD Ntail peptide binds to PduA and PduJ, and that PduD has affinity for binding itself but further work must be done to verify these results and address the challenges in this scenario. First, the self-interaction of the peptides presents a problem to the experimental design of the fluorescence polarization assay approach, which in its simplest applications presumes that the fluorescently labeled species is non-interacting. This also explains why previous attempts by former members of the Yeates Laboratory to study binding by isothermal calorimetry had technical difficulties with a non-uniform baseline, rendering results uninterpretable. Second, the data may be further convoluted by the Pdu shell protein's oligomer formation, whole equilibrium position is expected to be concentration-dependent.

For the fluorescence polarization assay, the total intensity of the fluorescently labeled peptides must be monitored and shown to be constant during titration with binding ligand. Preliminary measurements of total intensity were uninterpretable, perhaps due to malfunction of the photon counter of the Tecan M1000 plate reader instrument (data not shown).

To address the first issue of peptide self-interaction, future work to characterize shell protein and enzyme Ntail peptide binding might focus on using an MBP-fusion based depletion

assay. The idea is that MBP-fused peptides bound to dextrin or amylose beads will present the peptides to the ligand in a more uniform and accessible manner than if the peptides are subject to degrees of aggregation. The MBP-fused peptides on dextrin beads can be added in titrating amounts to a constant concentration of the ligand, shell protein of choice. Centrifuging the samples at high acceleration (~30,000xg) can pellet the beads and measurement of the depletion of the shell protein from solution (not due to shell protein precipitation) can be used to determine the binding affinity between peptide and ligand. The N-terminus of MBP (the gene product of *malE* gene in *E. coli*, PDBID 1ANF) is solvent accessible and ideal for the fusion of N-terminal peptides[49]. Also, MBP increases the solubility of some proteins which would be beneficial in this system[50]. A Gibson assembly plan and sequences of Gibson assembly blocks for MBP-fused peptides, including negative controls, for future studies is included in the Appendix.

For the second issue, the hexamer assembly of the Pdu shell proteins can be verified by SEC-MALS (size exclusion chromatography followed by multi-angle light scattering) experiments. However, thus far, all attempts to measure PduA K26A or PduJ K25A by SEC-MALS have been delayed until repair of the SEC-MALS instrument. In addition, the observance of the hexamer assembly in PduA, PduA mutant, and PduJ crystal structures is strong support for its assembly in solution.

I have also tried surface plasmon resonance to measure shell protein binding peptide immobilized on a chip. Ligand thiol coupling was used to conjugate peptides with KCK for single cysteine conjugation using EDC/NHS activation and PDEA crosslinking on CM5 chip (GE Healthcare). The most efficient coupling reaction was observed for PduE N-tail with KCK peptide. Injections of PduA mutants led to uninterpretable results. The Biacore software could

not fit binding curves to the response curves to model kinetics. To investigate this, I injected PduA K26A (25μM, 240s) to measure k_{ON} and k_{OFF} manually. Injection produced a binding signal of 100 RU (Response Units), but dissociation to the baseline was not achieved. Injection of astringent regeneration buffer (10mM glycine pH 3.0, 150mM NaCl, 30s) did not return the signal to baseline, even after 1800s at the recommended flow rate 30μL/min. This indicates that PduA constructs are binding to the PduE N-tail labeled chip but with slow dissociation, disrupting subsequent binding events. At 25 μM PduA K26A, I estimate $\tau_{1/2,ON}$=40s and $\tau_{1/2,OFF}$=10,475s, thus k_{ON}=0.025s^{-1} and k_{OFF}=95.5x$10^{-6}s^{-1}$ (Figure 8.14). Successive washing of the chip did not remove all the injected protein, prohibiting collection of further data points using the PduE N-tail peptide labeled Biacore chip. Thus, a complete dataset to analyze kinetics of Pdu MCP shell protein binding to Pdu enzyme N-tail peptides was unobtainable.

According to Sutter et al., dynamics of the halophilic myxobacterium *Haliangum ochraceum* BMC hexameric shell protein involves hexamer units assembling into flat sheets[51]. Using high-speed Atomic Force Microscopy and similar experimental conditions at 50mM TRIS pH 7.8, 100mM NaCl, 10mM $MgCl_2$, at ambient temperature and pressure, they observe the translocation of hexamers in large sheets composed of hexamers. They do not observe the translocation of partial hexamers at a protein concentration of approximately 150μM with their method. However, they note that polarity of the distinct faces of the hexamer, concave v. convex, confers affinity for the polar convex side to the mica substrate. A similar phenomenon in PduA may have caused attraction of the PduA hexamer to the CM5 chip surface on the reference channel which is composed of a carboxymethylated dextran matrix.

I note that the self-interactions of the Pdu enzyme N-terminal peptides lends support to the pre-core assembly model in bacterial microcompartment assembly[44]. Further experiments to extend this study could provide additional evidence to support this theory.

Figure 8.1. Clustal W2 alignment of PduCDELP N-terminal peptides.

```
PDUC_SALTY  --MRSKRFEALAKRPVNQDG
PDUD_SALTY  MEINEKLLRQIIEDVLRDMK
PDUE_SALTY  --MNTDAIESMVRDVLSRMN
PDUL_SALTY  --MDKELLQSTVRKVLDEMR
PDUP_SALTY  --MNTSELETLIRTILSEQL
                :  .  :.   .  :
```

Figure 8.2. Peptide docking *in silico* using FlexPepDock predicts similar binding modes for (A) the N-terminal tail of PduP and PduA (courtesy of Jorda et al.) and (B) the N-terminal tail of PduP and PduJ. Overlay in (C).

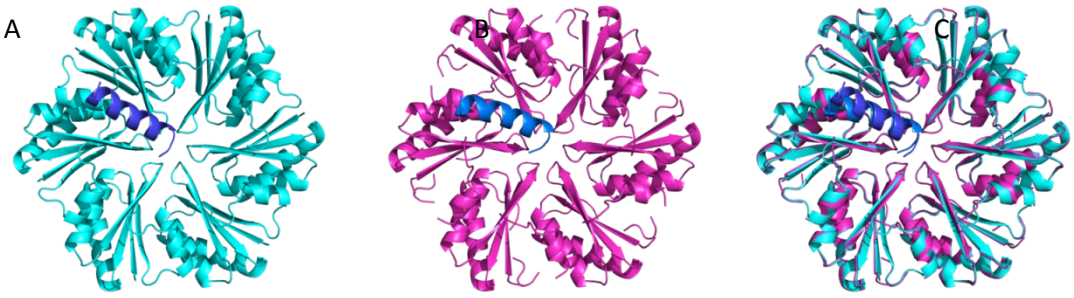


Figure 8.3. Schematic of GFP fusion peptide designs for EMSA (A) and schematic of peptide (yellow) fused GFP (green) binding shell protein (blue)

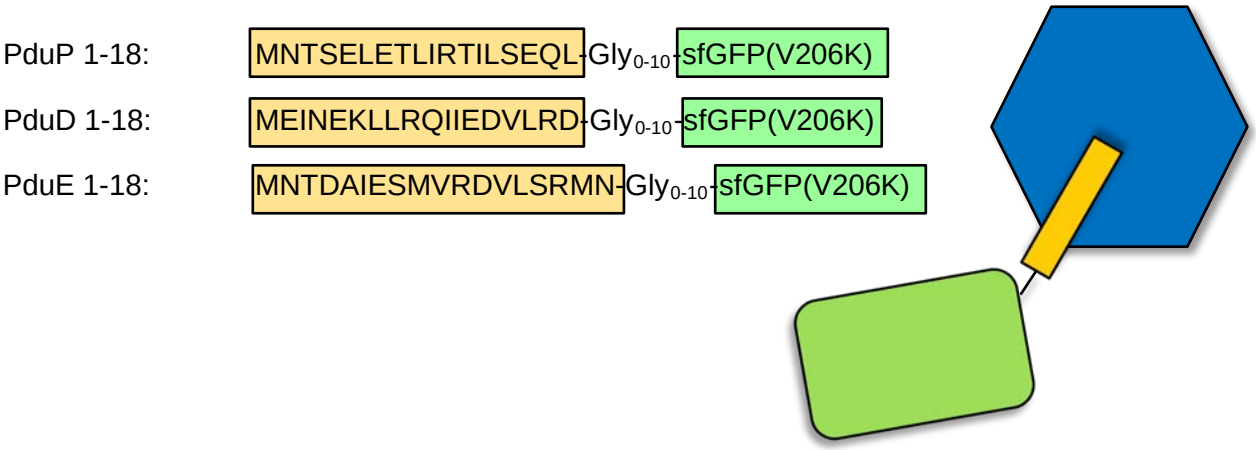


Figure 8.4. Native PAGE (8% acrylamide) of PduA and PduP peptide (20mA, 2h). Lanes: (1) 20μM PduA, (2) 20μM $PduP^{1-18}$-sfGFP, (3) 20μM PduA and 6.7μM $PduP^{1-18}$-sfGFP, (4) 20μM PduA and 20μM $PduP^{1-18}$-sfGFP, (5) 20μM PduA and 60μM $PduP^{1-18}$-sfGFP, (6) 20μM $PduP^{1-18}$-Gly_2-sfGFP, (7) 20μM PduA and 6.7μM $PduP^{1-18}$-Gly_2-sfGFP, (8) 20μM PduA and 20μM $PduP^{1-18}$-Gly_2-sfGFP, (9) 20μM PduA and 60μM $PduP^{1-18}$-Gly_2-sfGFP, (10) 20μM $PduP^{1-18}$-Gly_5-sfGFP, (11) 20μM PduA and 6.7μM $PduP^{1-18}$-Gly_5-sfGFP, (12) 20μM PduA and 20μM $PduP^{1-18}$-Gly_5-sfGFP, (13) 20μM PduA and 60μM $PduP^{1-18}$-Gly_5-sfGFP, (14) 20μM $PduP^{1-18}$-Gly_{10}-sfGFP, (15) 20μM PduA and 6.7μM $PduP^{1-18}$-Gly_{10}-sfGFP, (16) 20μM PduA and 20μM $PduP^{1-18}$-Gly_{10}-sfGFP, (17) 20μM PduA and 60μM $PduP^{1-18}$-Gly_{10}-sfGFP

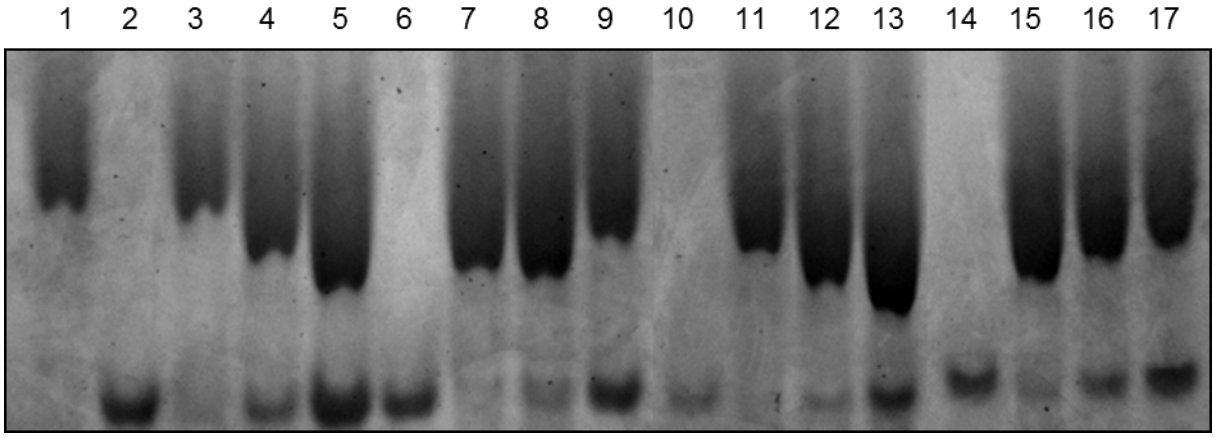


Figure 8.5. Native PAGE (8% acrylamide, 20mA, 1.25h):(A) Titration of PduP$^{1-18}$-Gly$_{10}$-sfGFP into PduA (50μM), (B) PduA only, (C) PduP$^{1-18}$-Gly$_{10}$-sfGFP only, (D) PduP$^{1-18}$-Gly$_{2}$-sfGFP only.

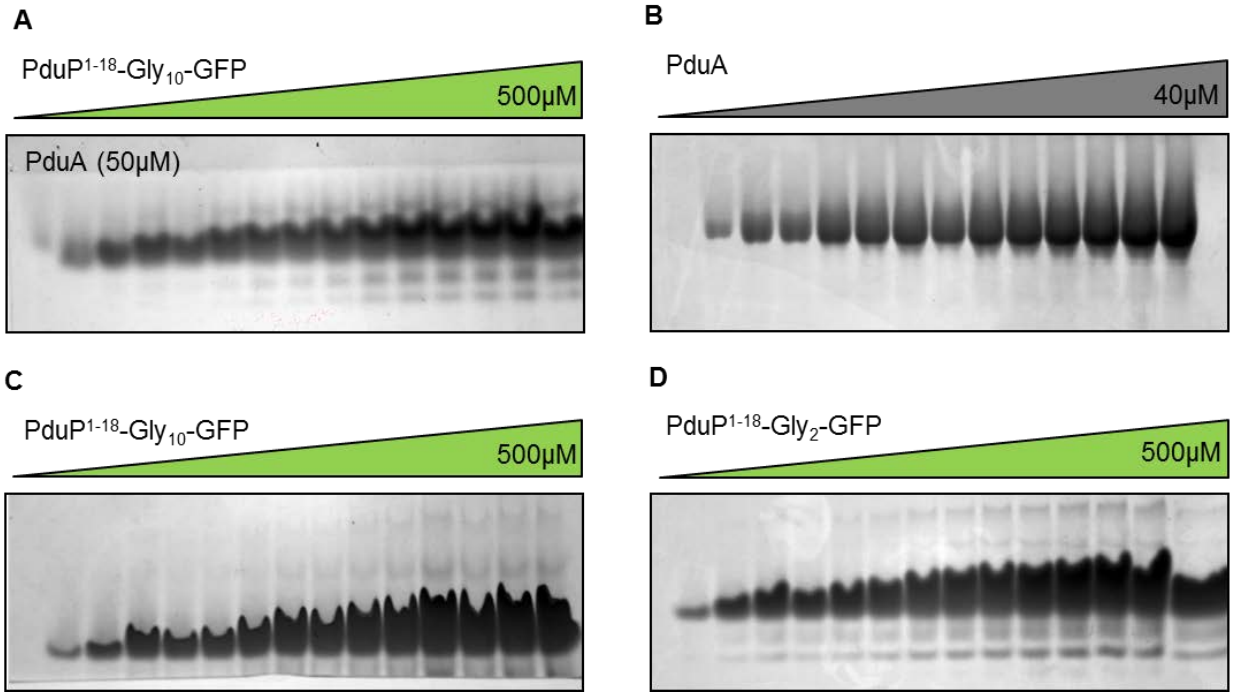


Figure 8.6. Schematic of fluorescence polarization assay where a fluorescently labeled Pdu enzyme N-terminal peptide (fast tumbling) would experience an increase in fluorescence polarization upon binding to a species of larger molecular weight (slow tumbling).

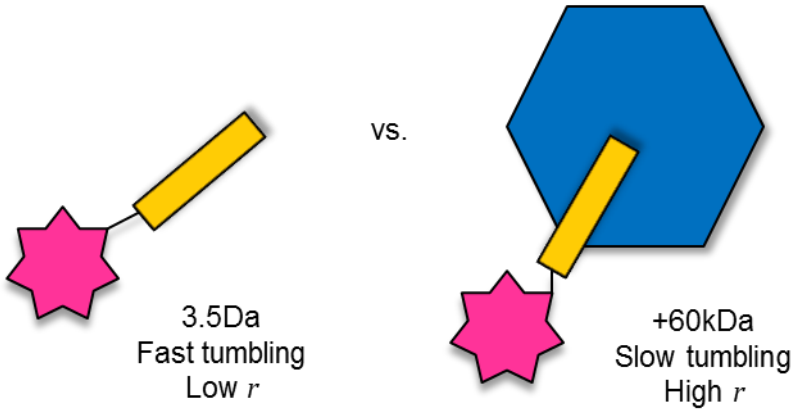
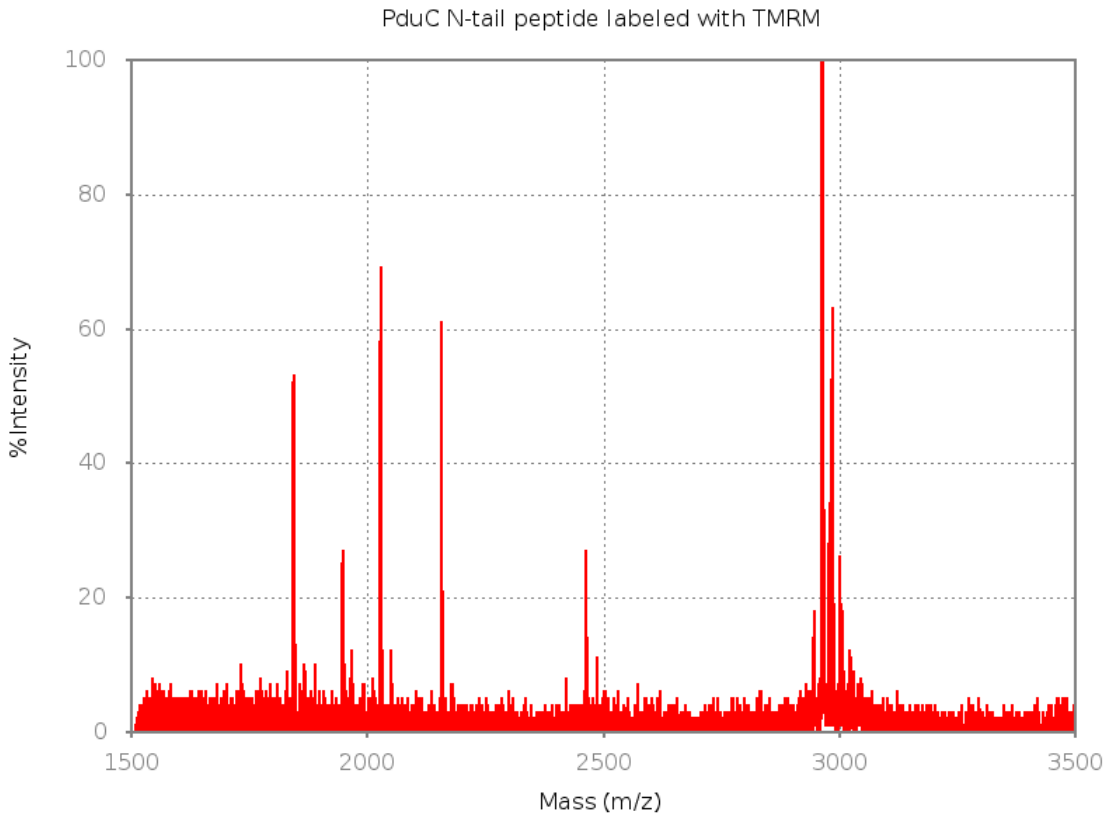
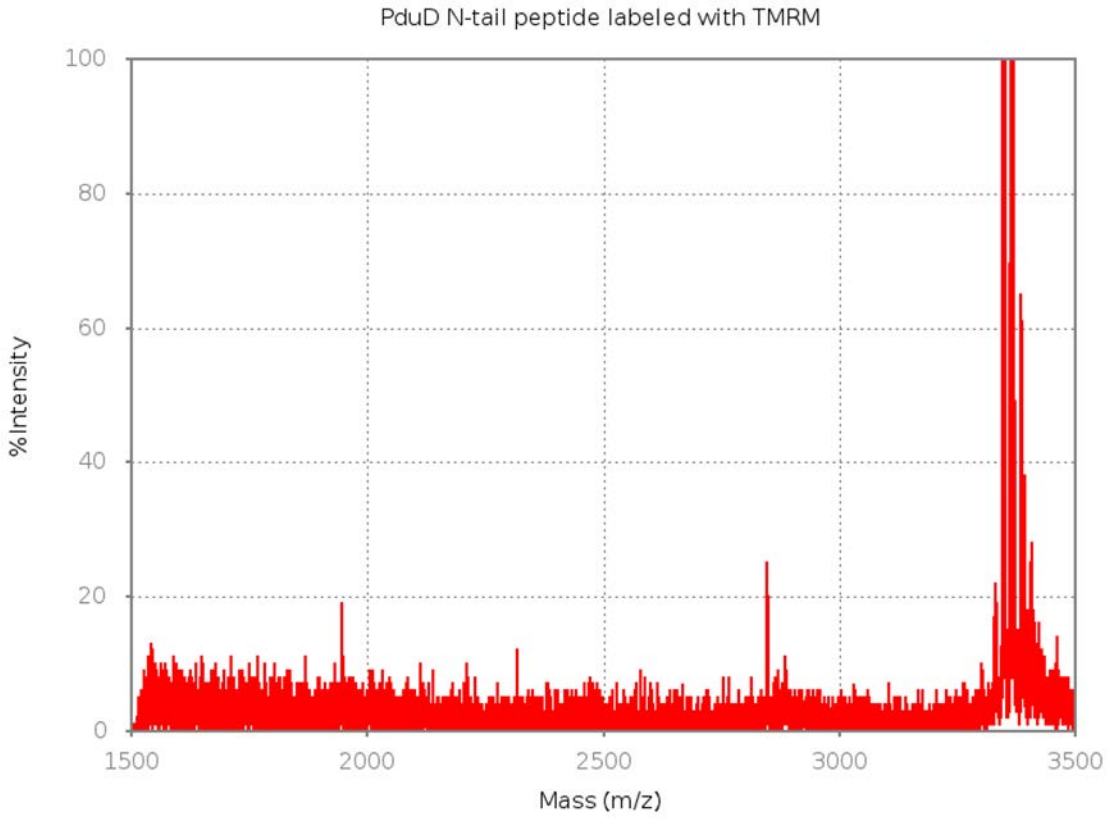


Figure 8.7. MALDI-MS spectra of purified labeled Pdu enzyme N-terminal peptides with tetramethyl-5-rhodamine maleimide (TMRM).

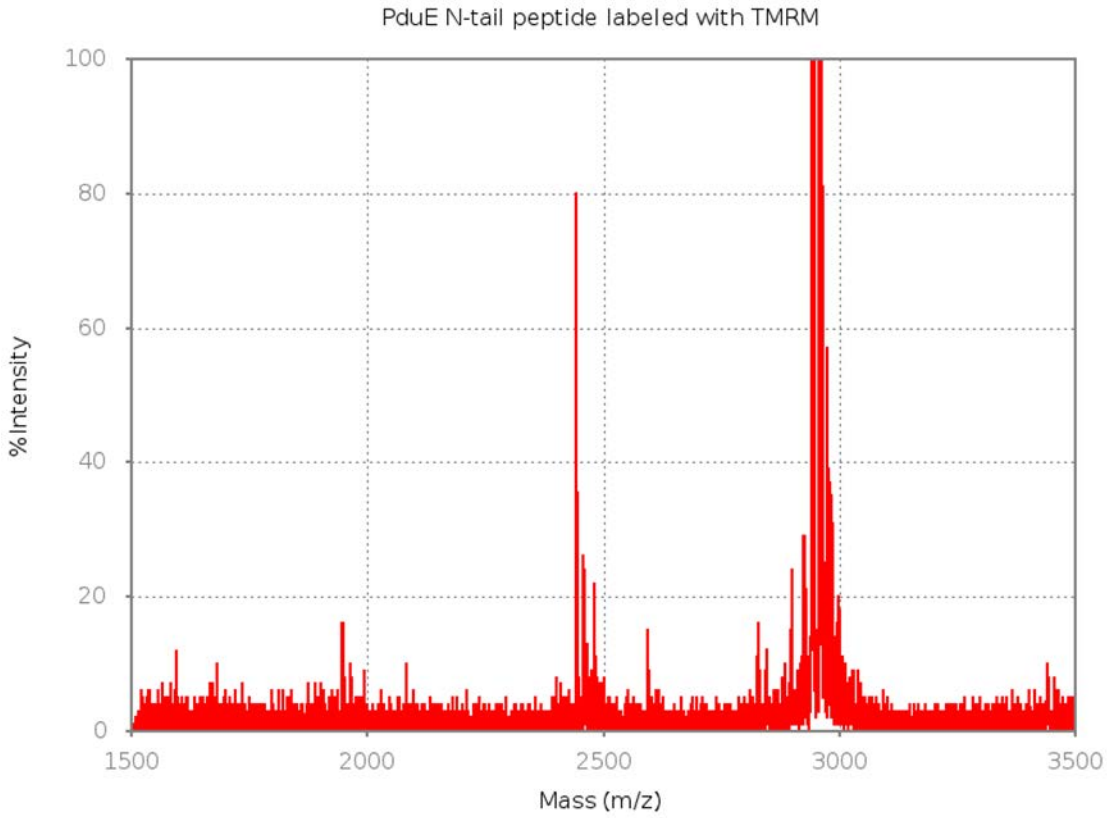
A



B


PduD N-tail peptide labeled with TMRM
% Intensity
100
80
60
40
20
0
1500
2000
2500
3000
3500
Mass (m/z)

C


PduE N-tail peptide labeled with TMRM
% Intensity
100
80
60
40
20
0
1500
2000
2500
3000
3500
Mass (m/z)

D

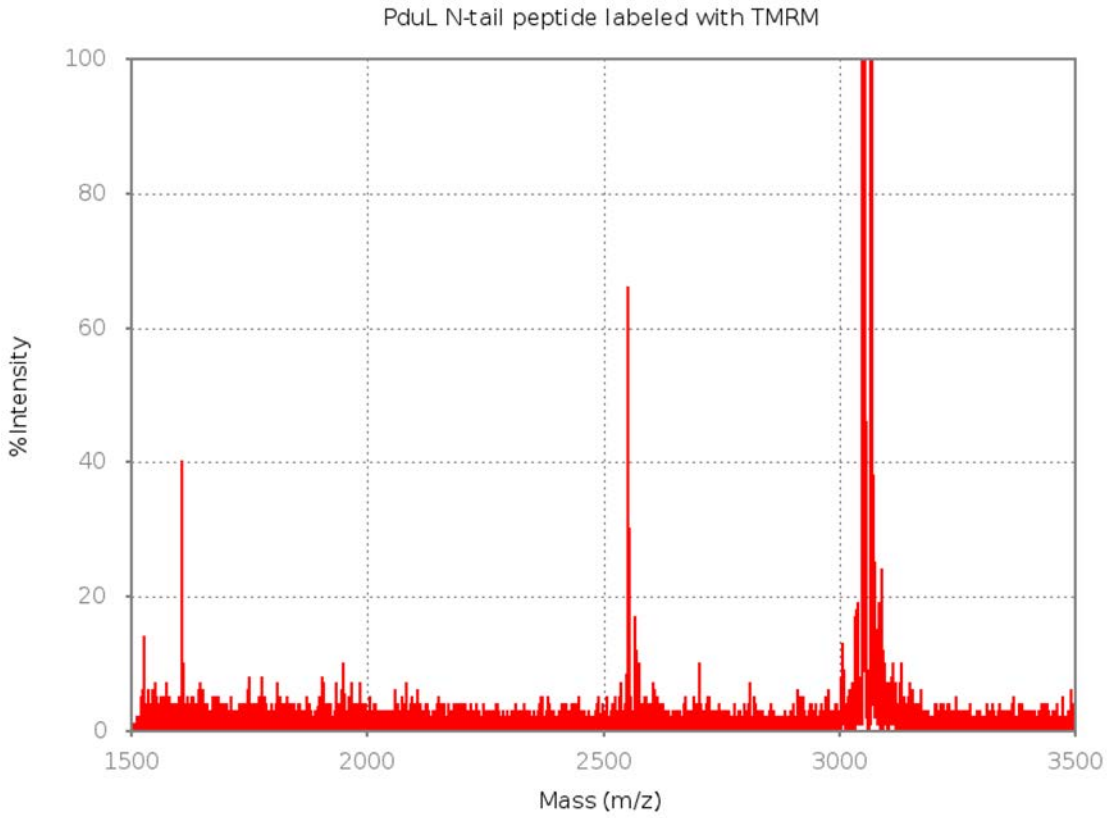

PduL N-tail peptide labeled with TMRM
100
80
60
40
20
0
% Intensity
1500
2000
2500
3000
3500
Mass (m/z)

Table 8.1. Peptides used for fluorescence anisotropy studies.

Peptide	Sequence	Mass (Da)
PduC-Ntail	MRSKRFEALAKRPVNQDG	2103.42
PduC-Ntail-KCK	MRSKRFEALAKRPVNQDGKCK	2462.91
PduD-Ntail	MEINEKLLRQIIEDVLRDMK	2486.96
PduD-Ntail-KCK	MEINEKLLRQIIEDVLRDMKKCK	2846.45
PduENtail	MNTDAIESMVRDVLSRMN	2082.38
PduENtail-KCK	MNTDAIESMVRDVLSRMNKCK	2441.87
PduLNtail	MDKELLQSTVRKVLDEMR	2191.58
PduLNtail-KCK	MDKELLQSTVRKVLDEMRKCK	2551.07
PduPNtail	MNTSELETLIRTILSEQL	2091.39
PduPNtail-KCK	MNTSELETLIRTILSEQLKCK	2450.88

Table 8.2. Fluorophore conjugation of peptides analyzed by MALDI-TOF MS on reaction day 6 corresponding to spectra in Figure 7.

Plate No.	Sample	Calculated (Da)	Experimental (Da, %Int)
02	PduC-KCK + TMRM	2462.91	2463.4 (10)
		2942.91	2963.5 (100)
03	PduD-KCK + TMRM	2846.45	2846.7 (5)
		3326.45	3346.9 (100)

04	PduE-KCK + TMRM	2441.87	2442.6 (12.5)
		2921.87	2941.7 (100)
05	PduL-KCK + TMRM	2551.87	2551.7 (15)
		3031.87	3052.0 (100)

Figure 8.8. Optimization of TAMRA emission

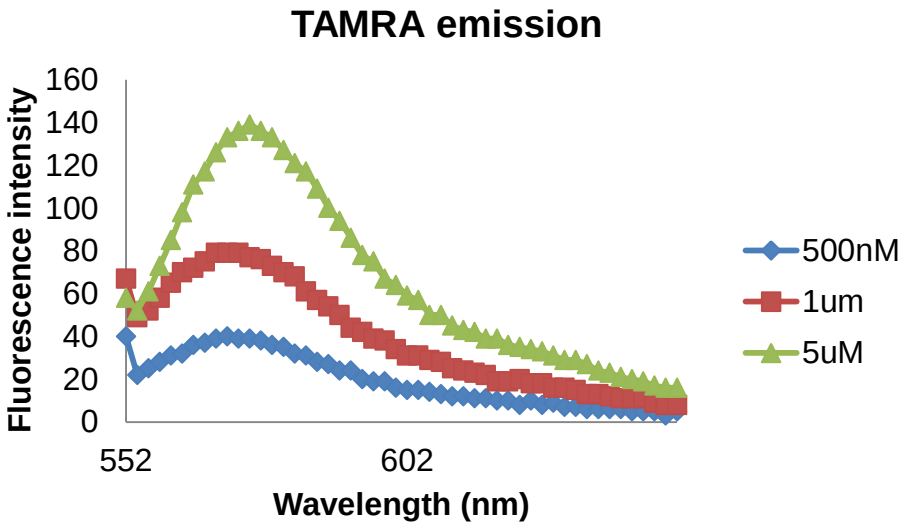
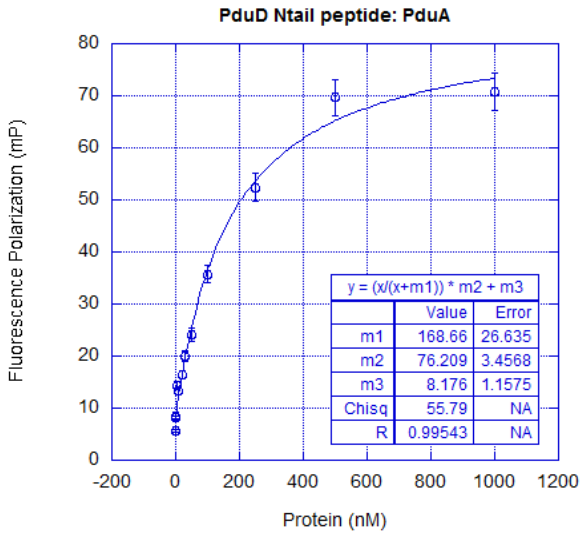
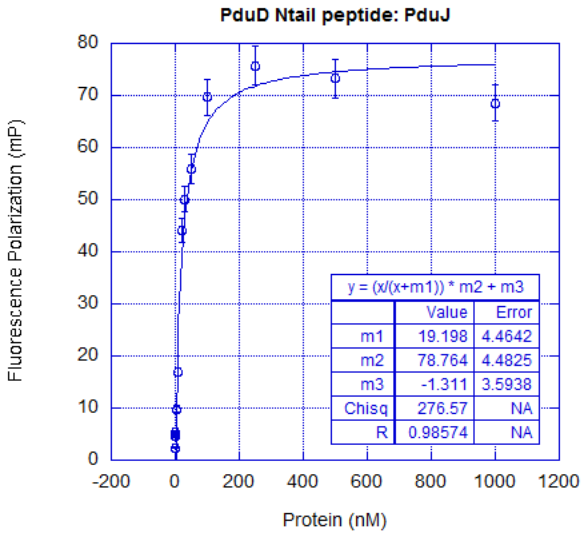


Figure 8.9. Fluorescence polarization measurements show high binding affinities between PduD Ntail peptide and (A) PduA, (B) PduJ. (C) Overlay of the plots in A and B.

A



B



C

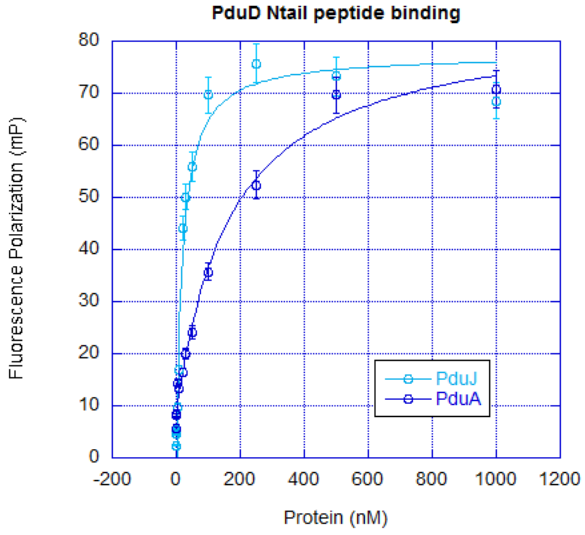


Figure 8.10. Specificity of binding by unlabeled competition assay

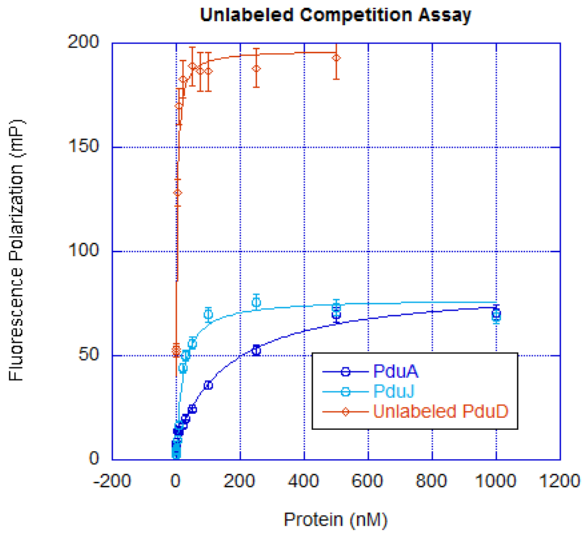


Figure 8.11. PduD Ntail peptide exhibits self-binding properties with low affinity

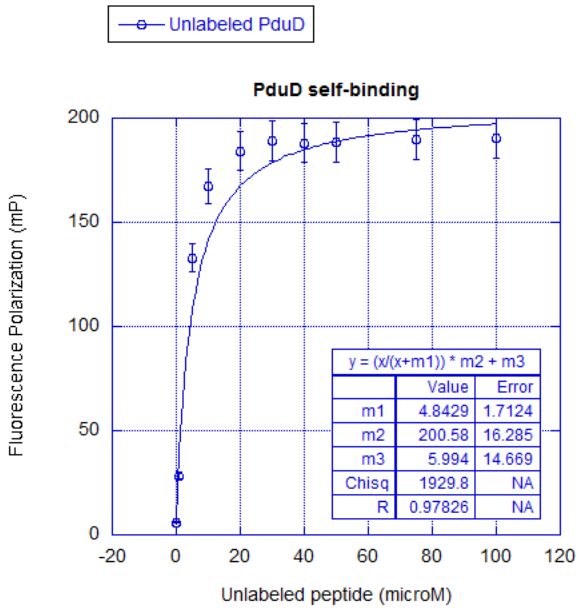
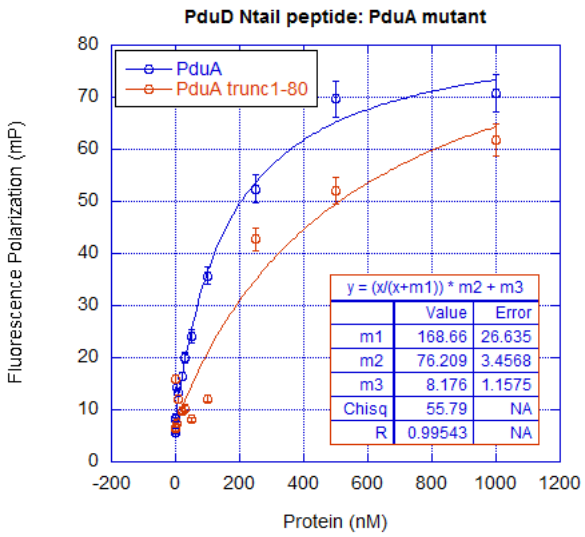
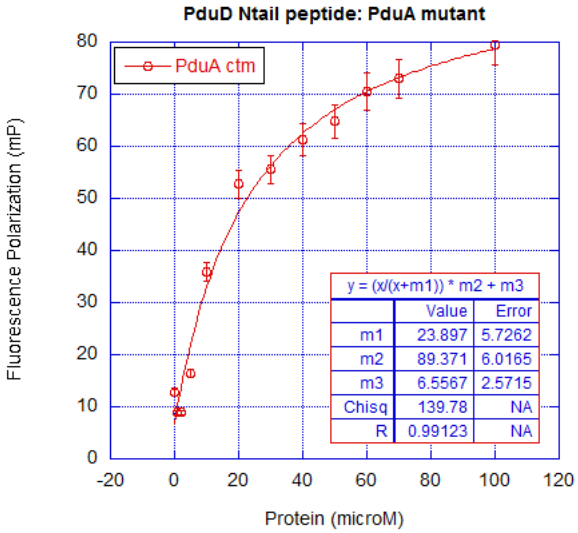


Figure 8.12. PduAJ c-terminal mutants exhibit reduced binding to PduD Ntail peptide. (A) PduD Ntail peptide with PduA trunc1-80. (B) PduD Ntail peptide with PduA ctm. (C) PduD Ntail peptide with PduJ trunc1-80 and PduJ ctm.

A



B



C

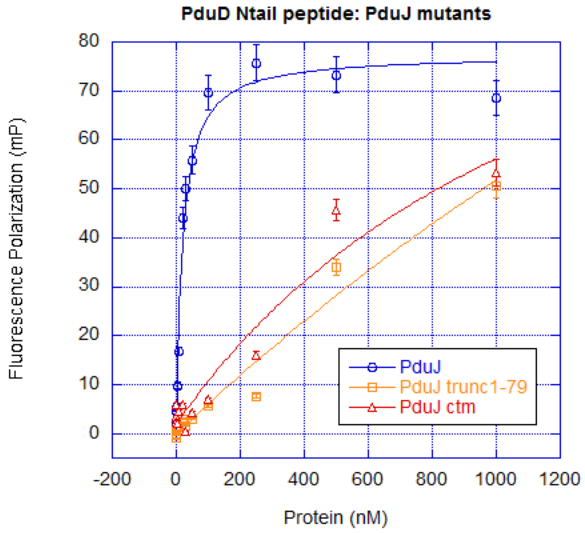
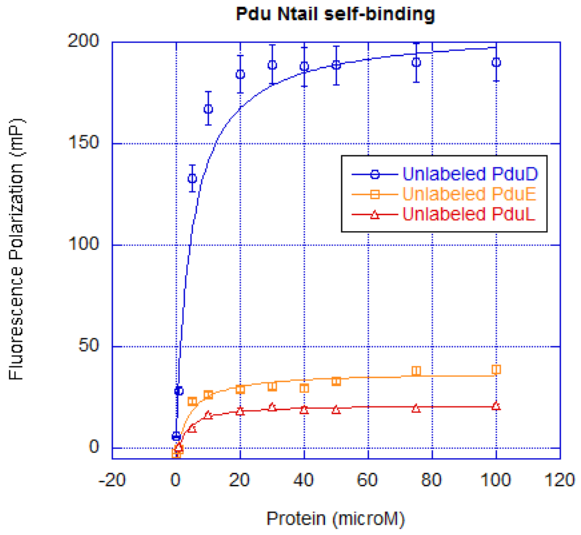

PduD Ntail peptide: PduJ mutants
Fluorescence Polarization (mP)
Protein (nM)
80
70
60
50
40
30
20
10
0
-200
0
200
400
600
800
1000
1200
PduJ
PduJ trunc1-79
PduJ ctm

Figure 8.13. Self-binding properties of the Pdu enzyme Ntail peptides. (A) PduDEL exhibit self-binding properties. (B) PduC does not exhibit binding character

A



B

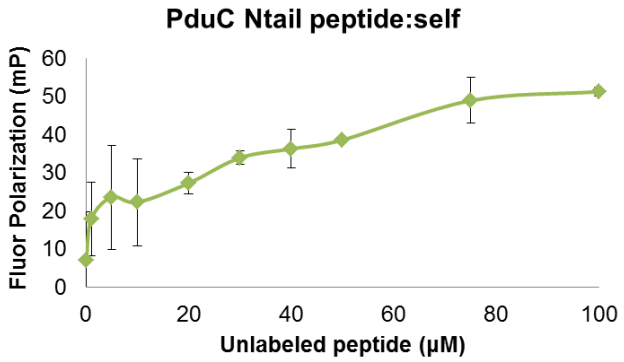
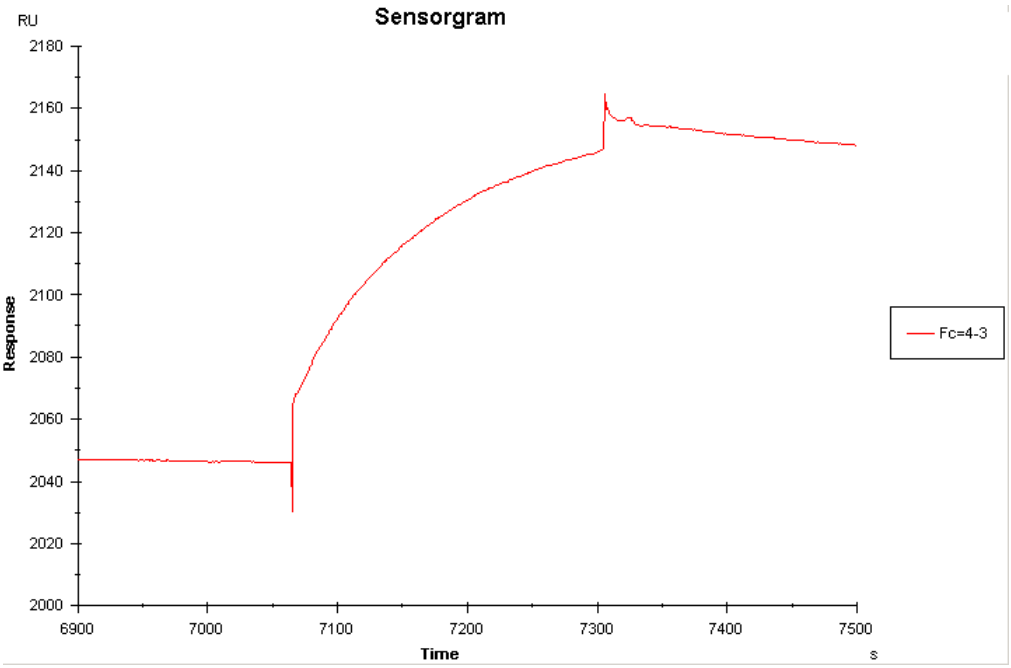
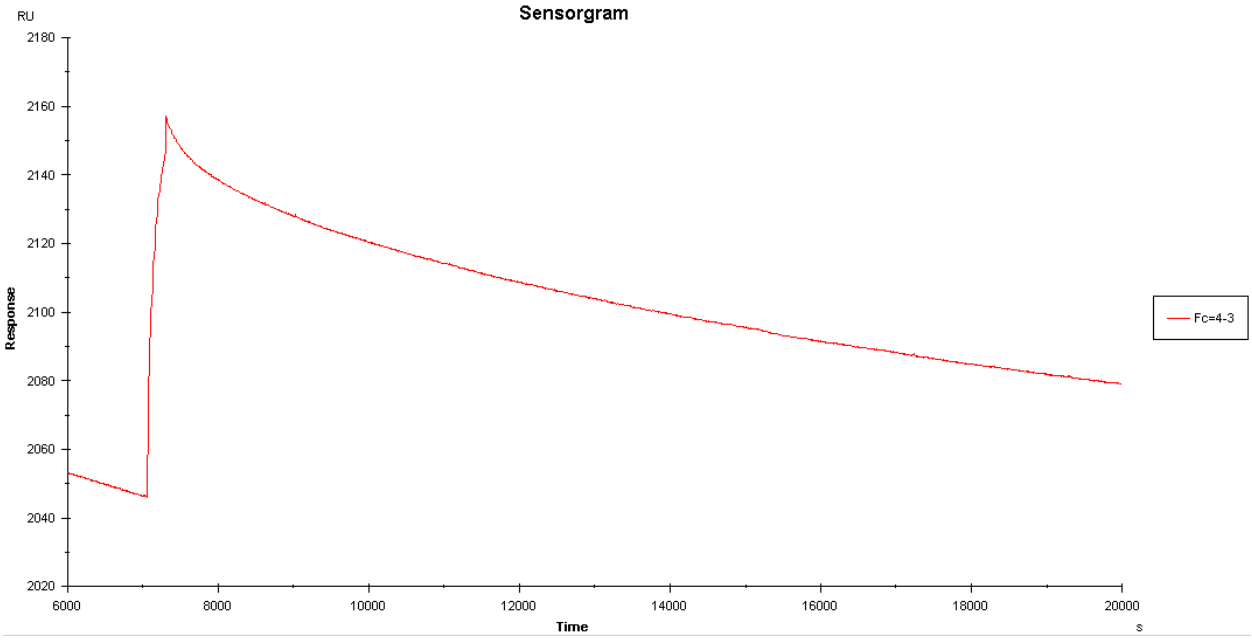


Figure 8.14. Biacore surface plasmon resonance sensogram of PduA K26A injection on PduE Ntail peptide labeled CM5 chip: (A) injection and on-rate, (B) wash and off-rate.

A



B



Appendix

Cloning strategy for MBP-fused Pdu enzyme N-tail peptides

I. Sequences

a. First 18 amino acids in the PduCDELP

PduCNtail- MRSKRFEALAKRPVNQDG

PduDNtail- MEINEKLLRQIIEDVLRDMK

PduENtail- MNTDAIESMVRDVLSRMN

PduLNtail- MDKELLQSTVRKVLDEMR

PduPNtail- MNTSELETLIRTILSEQL

b. MBP DNA sequence from *malE* gene in *E. coli*

ATG AAAATCGAAG AAGGTAAACT GGTAATCTGG ATTAACGGCG ATAAAGGCTA
TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT ACCGGAATTA
AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT
GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT
TGGTGGCTAC GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG
CGTTCCAGGA CAAGCTGTAT CCGTTTACCT GGGATGCCGT ACGTTACAAC
GGCAAGCTGA TTGCTTACCC GATCGCTGTT GAAGCGTTAT CGCTGATTTA
TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA GAGATCCCGG
CGCTGGATAA AGAACTGAAA GCGAAAGGTA AGAGCGCGCT GATGTTCAAC
CTGCAAGAAC CGTACTTCAC CTGGCCGCTG ATTGCTGCTG ACGGGGGTTA

TGCGTTCAAG TATGAAAACG GCAAGTACGA CATTAAAGAC GTGGGCGTGG
ATAACGCTGG CGCGAAAGCG GGTCTGACCT TCCTGGTTGA CCTGATTAAA
AACAAACACA TGAATGCAGA CACCGATTAC TCCATCGCAG AAGCTGCCTT
TAATAAAGGC GAAACAGCGA TGACCATCAA CGGCCCGTGG GCATGGTCCA
ACATCGACAC CAGCAAAGTG AATTATGGTG TAACGGTACT GCCGACCTTC
AAGGGTCAAC CATCCAAACC GTTCGTTGGC GTGCTGAGCG CAGGTATTAA
CGCCGCCAGT CCGAACAAAG AGCTGGCAAA AGAGTTCCTC GAAAACTATC
TGCTGACTGA TGAAGGTCTG GAAGCGGTTA ATAAAGACAA ACCGCTGGGT
GCCGTAGCGC TGAAGTCTTA CGAGGAAGAG TTGGCGAAAG ATCCACGTAT
TGCCGCCACT ATGGAAAACG CCCAGAAAGG TGAAATCATG CCGAACATCC
CGCAGATGTC CGCTTTCTGG TATGCCGTGC GTACTGCGGT GATCAACGCC
GCCAGCGGTC GTCAGACTGT CGATGAAGCC CTGAAAGACG CGCAGACTAA
TTCGAGCTCG

c. MBP amino acid sequence

MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGP
DIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIY
NKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGK
YDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAW
SNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEA
VNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINA
ASGRQTVDEALKDAQTNSSS

II. Gibson assembly plan for total of 17 different constructs

a. Triple digest: 2.5ug vector, w/ controls at .25ug vector

NdeI: 3.75uL, NEB Cutsmart buffer

XhoI: 3.75uL, NEB Cutsmart buffer

BamHI cuts original insert: 3.75uL, NEB Cutsmart buffer

vector = pET24a(+) (Kanamycin resistance)

+Cutsmart buffer (includes BSA)

Total volume: 125uL

PCR purify the triple digested vector (tdVector)

b. Gibson assembly

tdVector

Gblock1 (~500bp)- PduCDELP peptides with or without linker, random peptide, MBP alone

Gblock2 (~700bp)- the end half of MBP

Transform into XL2-Blue, plate on KAN plates

c. Gibson assembly block sequences

i. PduC

Construct: NdeI (Start) + PduC 1-18 (RSKRFEALAKRPVNQDG) + MBP + Stop + XhoI

Shown with pET24a 5' and 3' overlaps in bold; 1215bp total

GTTTAACTTTAAGAAGGAGATATACATATG cgcagcaaacgctttgaagcgctggcgaaacgcccggtg aaccaggatggc AAAATCGAAG AAGGTAAACT GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA GAGATCCCGG CGCTGGATAA AGAACTGAAA GCGAAAGGTA AGAGCGCGCT GATGTTCAAC CTGCAAGAAC CGTACTTCAC CTGGCCGCTG ATTGCTGCTG ACGGGGGTTA TGCGTTCAAG TATGAAAACG GCAAGTACGA CATTAAAGAC GTGGGCGTGG ATAACGCTGG CGCGAAAGCG GGTCTGACCT TCCTGGTTGA CCTGATTAAA AACAAACACA TGAATGCAGA CACCGATTAC TCCATCGCAG AAGCTGCCTT TAATAAAGGC GAAACAGCGA TGACCATCAA CGGCCCGTGG GCATGGTCCA ACATCGACAC CAGCAAAGTG AATTATGGTG TAACGGTACT GCCGACCTTC AAGGGTCAAC CATCCAAACC GTTCGTTGGC GTGCTGAGCG CAGGTATTAA CGCCGCCAGT CCGAACAAAG AGCTGGCAAA AGAGTTCCTC GAAAACTATC TGCTGACTGA TGAAGGTCTG GAAGCGGTTA ATAAAGACAA ACCGCTGGGT GCCGTAGCGC TGAAGTCTTA CGAGGAAGAG TTGGCGAAAG ATCCACGTAT TGCCGCCACT ATGGAAAACG CCCAGAAAGG TGAAATCATG CCGAACATCC CGCAGATGTC CGCTTTCTGG TATGCCGTGC GTACTGCGGT

GATCAACGCC GCCAGCGGTC GTCAGACTGT CGATGAAGCC CTGAAAGACG

CGCAGACTAA TTCGAGCTCG **TAA TAG CTC GAGCAC CAC CAC CAC**

Gblock1 (CMBPfor, 481bp): pET24a vector(30nt) + NdeI (Start) + Peptide + MBP

GTTTAACTTTAAGAAGGAGATATACATATG

cgcagcaaacgctttgaagcgctggcgaaacgcccggtgaaccaggatggc AAAATCGAAG AAGGTAAACT

GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG

GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG

GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC

TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG

GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT

CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC

GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA

ACCCGCCAAA AACCTGGGAA GAGATCCCGG

Construct 2: NdeI (Start) + PduC 1-18 + GGS + MBP + Stop + XhoI

Gblock1 (CggsMBPfor): Same as EMBPfor with ggcggcagc insertion

GTTTAACTTTAAGAAGGAGATATACATATG

cgcagcaaacgctttgaagcgctggcgaaacgcccggtgaaccaggatggc ggcggcagc AAAATCGAAG

AAGGTAAACT GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC

GCTGAAGTCG GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT

TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG

GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC

GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA

CAAGCTGTAT CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA

TTGCTTACCC GATCGCTGTT GAAGCGTTAT CGCTGATTTA

TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA GAGATCCCGG

Construct 3: NdeI (Start) + PduC 1-18 + 2xGGS + MBP + Stop + XhoI

Gblock1 (C2ggsMBPfor): Same as CMBPfor with ggcggcagcggcggcagc insertion

GTTTAACTTTAAGAAGGAGATATACATATG

cgcagcaaacgctttgaagcgctggcgaaacgcccggtgaaccaggatggc ggcggcagcggcggcagc AAAATCGAAG

AAGGTAAACT GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC

GCTGAAGTCG GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT

TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG

GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC

GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA

CAAGCTGTAT CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA

TTGCTTACCC GATCGCTGTT GAAGCGTTAT CGCTGATTTA

TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA GAGATCCCGG

ii. PduD

Construct: NdeI (Start) + PduD 1-18 (EINEKLLRQIIEDVLRDMK) + MBP + Stop + XhoI

Shown with pET24a 5’ and 3’ overlaps in bold; 1215bp total

GTTTAACTTTAAGAAGGAGATATACATATG

gaaattaacgaaaaactgctgcgccagattattgaagatgtgctgcgcgatatgaaa AAAATCGAAG AAGGTAAACT

GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG

GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG

GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC

TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG

GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT

CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC

GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA

ACCCGCCAAA AACCTGGGAA GAGATCCCGG CGCTGGATAA AGAACTGAAA

GCGAAAGGTA AGAGCGCGCT GATGTTCAAC CTGCAAGAAC CGTACTTCAC

CTGGCCGCTG ATTGCTGCTG ACGGGGGTTA TGCGTTCAAG TATGAAAACG

GCAAGTACGA CATTAAAGAC GTGGGCGTGG ATAACGCTGG CGCGAAAGCG

GGTCTGACCT TCCTGGTTGA CCTGATTAAA AACAAACACA TGAATGCAGA

CACCGATTAC TCCATCGCAG AAGCTGCCTT TAATAAAGGC GAAACAGCGA

TGACCATCAA CGGCCCGTGG GCATGGTCCA ACATCGACAC CAGCAAAGTG

AATTATGGTG TAACGGTACT GCCGACCTTC AAGGGTCAAC CATCCAAACC

GTTCGTTGGC GTGCTGAGCG CAGGTATTAA CGCCGCCAGT CCGAACAAAG

AGCTGGCAAA AGAGTTCCTC GAAAACTATC TGCTGACTGA TGAAGGTCTG

GAAGCGGTTA ATAAAGACAA ACCGCTGGGT GCCGTAGCGC TGAAGTCTTA

CGAGGAAGAG TTGGCGAAAG ATCCACGTAT TGCCGCCACT ATGGAAAACG

CCCAGAAAGG TGAAATCATG CCGAACATCC CGCAGATGTC CGCTTTCTGG

TATGCCGTGC GTACTGCGGT GATCAACGCC GCCAGCGGTC GTCAGACTGT

CGATGAAGCC CTGAAAGACG CGCAGACTAA TTCGAGCTCG **TAA TAG CTC**

GAGCAC CAC CAC CAC

Gblock1 (DMBPfor, 481bp): pET24a vector(30nt) + NdeI (Start) + Peptide + MBP

GTTTAACTTTAAGAAGGAGATATACATATG gaaattaacgaaaaactgctgcgccagattattgaagatgt gctgcgcgatatgaaa AAAATCGAAG AAGGTAAACT GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA GAGATCCCGG

Construct 2: NdeI (Start) + PduD 1-18 + GGS + MBP + Stop + XhoI

Gblock1 (DggsMBPfor): Same as EMBPfor with ggcggcagc insertion

GTTTAACTTTAAGAAGGAGATATACATATG gaaattaacgaaaaactgctgcgccagattattgaagatgt gctgcgcgatatgaaa ggcggcagc AAAATCGAAG AAGGTAAACT GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT

GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA

AACCTGGGAA GAGATCCCGG

Construct 3: NdeI (Start) + PduD 1-18 + 2xGGS + MBP + Stop + XhoI

Gblock1 (D2ggsMBPfor): Same as CMBPfor with ggcggcagcggcggcagc insertion

GTTTAACTTTAAGAAGGAGATATACATATG gaaattaacgaaaaactgctgcgccagattattgaagatgt

gctgcgcgatatgaaa ggcggcagcggcggcagc AAAATCGAAG AAGGTAAACT GGTAATCTGG

ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT

CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG

AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC TGACATTATC

TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG GCCTGTTGGC

TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT CCGTTTACCT

GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT

GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA

AACCTGGGAA GAGATCCCGG

iii. PduE

Construct: NdeI (Start) + PduE 1-18 (NTDAIESMVRDVLSRMN) + MBP + Stop + XhoI

Shown with pET24a 5' and 3' overlaps in bold; 1215bp total

GTTTAACTTTAAGAAGGAGATATACATATGAACACCGATGCGATTGAAAGCATGGTGCGC

GATGTGCTGAGCCGCATGAACAAAATCGAAG AAGGTAAACT GGTAATCTGG

ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT

CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG

AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC

ACGACCGCTT TGGTGGCTAC GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA GAGATCCCGG CGCTGGATAA AGAACTGAAA GCGAAAGGTA AGAGCGCGCT GATGTTCAAC CTGCAAGAAC CGTACTTCAC CTGGCCGCTG ATTGCTGCTG ACGGGGGTTA TGCGTTCAAG TATGAAAACG GCAAGTACGA CATTAAAGAC GTGGGCGTGG ATAACGCTGG CGCGAAAGCG GGTCTGACCT TCCTGGTTGA CCTGATTAAA AACAAACACA TGAATGCAGA CACCGATTAC TCCATCGCAG AAGCTGCCTT TAATAAAGGC GAAACAGCGA TGACCATCAA CGGCCCGTGG GCATGGTCCA ACATCGACAC CAGCAAAGTG AATTATGGTG TAACGGTACT GCCGACCTTC AAGGGTCAAC CATCCAAACC GTTCGTTGGC GTGCTGAGCG CAGGTATTAA CGCCGCCAGT CCGAACAAAG AGCTGGCAAA AGAGTTCCTC GAAAACTATC TGCTGACTGA TGAAGGTCTG GAAGCGGTTA ATAAAGACAA ACCGCTGGGT GCCGTAGCGC TGAAGTCTTA CGAGGAAGAG TTGGCGAAAG ATCCACGTAT TGCCGCCACT ATGGAAAACG CCCAGAAAGG TGAAATCATG CCGAACATCC CGCAGATGTC CGCTTTCTGG TATGCCGTGC GTACTGCGGT GATCAACGCC GCCAGCGGTC GTCAGACTGT CGATGAAGCC CTGAAAGACG CGCAGACTAA TTCGAGCTCG **TAA TAG** **CTC GAG**CAC CAC CAC CAC

Gblock1 (EMBPfor, 481bp): pET24a vector(30nt) + NdeI (Start) + Peptide + MBP

GTTTAACTTTAAGAAGGAGATATACATATGAACACCGATGCGATTGAAAGCATGGTGCGC GATGTGCTGAGCCGCATGAACAAAATCGAAG AAGGTAAACT GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG

CGTTCCAGGA CAAGCTGTAT CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA GAGATCCCGG

Construct 2: NdeI (Start) + PduE 1-18 (NTDAIESMVRDVLSRMN)+ GGS + MBP + Stop + XhoI

Gblock1 (EggsMBPfor): Same as EMBPfor with ggcggcagc insertion

GTTTAACTTTAAGAAGGAGATATACATATGAACACCGATGCGATTGAAAGCATGGTGCGC GATGTGCTGAGCCGCATGAAC ggcggcagc AAAATCGAAG AAGGTAAACT GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA GAGATCCCGG

Construct 3: NdeI (Start) + PduE 1-18 (NTDAIESMVRDVLSRMN)+ 2xGGS + MBP + Stop + XhoI

Gblock1 (E2ggsMBPfor): Same as EMBPfor with ggcggcagcggcggcagc insertion

GTTTAACTTTAAGAAGGAGATATACATATGAACACCGATGCGATTGAAAGCATGGTGCGC GATGTGCTGAGCCGCATGAAC ggcggcagcggcggcagc AAAATCGAAG AAGGTAAACT GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA

TTGCTTACCC GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA GAGATCCCGG

iv. PduL

Construct: NdeI (Start) + PduL 1-18 (DKELLQSTVRKVLDEMR) + MBP + Stop + XhoI

Shown with pET24a 5' and 3' overlaps in bold; 1215bp total

GTTTAACTTTAAGAAGGAGATATACATATG

gataaagaactgctgcagagcaccgtgcgcaaagtgctggatgaaatgcgc AAAATCGAAG AAGGTAAACT GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA GAGATCCCGG CGCTGGATAA AGAACTGAAA GCGAAAGGTA AGAGCGCGCT GATGTTCAAC CTGCAAGAAC CGTACTTCAC CTGGCCGCTG ATTGCTGCTG ACGGGGGTTA TGCGTTCAAG TATGAAAACG GCAAGTACGA CATTAAAGAC GTGGGCGTGG ATAACGCTGG CGCGAAAGCG GGTCTGACCT TCCTGGTTGA CCTGATTAAA AACAAACACA TGAATGCAGA CACCGATTAC TCCATCGCAG AAGCTGCCTT TAATAAAGGC GAAACAGCGA TGACCATCAA CGGCCCGTGG GCATGGTCCA ACATCGACAC CAGCAAAGTG AATTATGGTG TAACGGTACT GCCGACCTTC AAGGGTCAAC CATCCAAACC

GTTCGTTGGC GTGCTGAGCG CAGGTATTAA CGCCGCCAGT CCGAACAAAG

AGCTGGCAAA AGAGTTCCTC GAAAACTATC TGCTGACTGA TGAAGGTCTG

GAAGCGGTTA ATAAAGACAA ACCGCTGGGT GCCGTAGCGC TGAAGTCTTA

CGAGGAAGAG TTGGCGAAAG ATCCACGTAT TGCCGCCACT ATGGAAAACG

CCCAGAAAGG TGAAATCATG CCGAACATCC CGCAGATGTC CGCTTTCTGG

TATGCCGTGC GTACTGCGGT GATCAACGCC GCCAGCGGTC GTCAGACTGT

CGATGAAGCC CTGAAAGACG CGCAGACTAA TTCGAGCTCG **TAA TAG** **CTC**

GAGCAC CAC CAC CAC

Gblock1 (LMBPfor, 481bp): pET24a vector(30nt) + NdeI (Start) + Peptide + MBP

GTTTAACTTTAAGAAGGAGATATACATATG gataaagaactgctgcagagcaccgtgcgcaaagtgctg

gatgaaatgcgc AAAATCGAAG AAGGTAAACT GGTAATCTGG ATTAACGGCG

ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT

ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT

CCCACAGGTT GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC

ACGACCGCTT TGGTGGCTAC GCTCAATCTG GCCTGTTGGC TGAAATCACC

CCGGACAAAG CGTTCCAGGA CAAGCTGTAT CCGTTTACCT GGGATGCCGT

ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT GAAGCGTTAT

CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA

GAGATCCCGG

Construct 2: NdeI (Start) + PduL 1-18 + GGS + MBP + Stop + XhoI

Gblock1 (LggsMBPfor): Same as LMBPfor with ggcggcagc insertion

GTTTAACTTTAAGAAGGAGATATACATATG

gataaagaactgctgcagagcaccgtgcgcaaagtgctggatgaaatgcgc ggcggcagc AAAATCGAAG AAGGTAAACT GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA GAGATCCCGG

Construct 3: NdeI (Start) + PduL 1-18 + 2xGGS + MBP + Stop + XhoI

Gblock1 (L2ggsMBPfor): Same as LMBPfor with ggcggcagcggcggcagc insertion

GTTTAACTTTAAGAAGGAGATATACATATG gataaagaactgctgcagagcaccgtgcgcaaagtgctg gatgaaatgcgc ggcggcagcggcggcagc AAAATCGAAG AAGGTAAACT GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA GAGATCCCGG

v. PduP

Construct: NdeI (Start) + PduP 1-18 (NTSELETLIRTILSEQL) + MBP + Stop + XhoI

Shown with pET24a 5’ and 3’ overlaps in bold; 1215bp total

GTTTAACTTTAAGAAGGAGATATACATATG

aacaccagcgaactggaaaccctgattcgcaccattctgagcgaacagctg AAAATCGAAG AAGGTAAACT

GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG

GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG

GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC

TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG

GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT

CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC

GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA

ACCCGCCAAA AACCTGGGAA GAGATCCCGG CGCTGGATAA AGAACTGAAA

GCGAAAGGTA AGAGCGCGCT GATGTTCAAC CTGCAAGAAC CGTACTTCAC

CTGGCCGCTG ATTGCTGCTG ACGGGGGTTA TGCGTTCAAG TATGAAAACG

GCAAGTACGA CATTAAAGAC GTGGGCGTGG ATAACGCTGG CGCGAAAGCG

GGTCTGACCT TCCTGGTTGA CCTGATTAAA AACAAACACA TGAATGCAGA

CACCGATTAC TCCATCGCAG AAGCTGCCTT TAATAAAGGC GAAACAGCGA

TGACCATCAA CGGCCCGTGG GCATGGTCCA ACATCGACAC CAGCAAAGTG

AATTATGGTG TAACGGTACT GCCGACCTTC AAGGGTCAAC CATCCAAACC

GTTCGTTGGC GTGCTGAGCG CAGGTATTAA CGCCGCCAGT CCGAACAAAG

AGCTGGCAAA AGAGTTCCTC GAAAACTATC TGCTGACTGA TGAAGGTCTG

GAAGCGGTTA ATAAAGACAA ACCGCTGGGT GCCGTAGCGC TGAAGTCTTA CGAGGAAGAG TTGGCGAAAG ATCCACGTAT TGCCGCCACT ATGGAAAACG CCCAGAAAGG TGAAATCATG CCGAACATCC CGCAGATGTC CGCTTTCTGG TATGCCGTGC GTACTGCGGT GATCAACGCC GCCAGCGGTC GTCAGACTGT CGATGAAGCC CTGAAAGACG CGCAGACTAA TTCGAGCTCG **TAA TAG CTC GAGCAC CAC CAC CAC**

Gblock1 (PMBPfor, 481bp): pET24a vector(30nt) + NdeI (Start) + Peptide + MBP

GTTTAACTTTAAGAAGGAGATATACATATG aacaccagcgaactggaaaccctgattcgcaccattctga gcgaacagctg AAAATCGAAG AAGGTAAACT GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA GAGATCCCGG

Construct 2: NdeI (Start) + PduP 1-18 + GGS + MBP + Stop + XhoI

Gblock1 (PggsMBPfor): Same as PMBPfor with ggcggcagc insertion

GTTTAACTTTAAGAAGGAGATATACATATG

aacaccagcgaactggaaaccctgattcgcaccattctgagcgaacagctg ggcggcagc AAAATCGAAG

AAGGTAAACT GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA GAGATCCCGG

Construct 3: NdeI (Start) + PduP 1-18 + 2xGGS + MBP + Stop + XhoI

Gblock1 (P2ggsMBPfor): Same as PMBPfor with ggcggcagcggcggcagc insertion

GTTTAACTTTAAGAAGGAGATATACATATG aacaccagcgaactggaaaccctgattcgcaccattctga gcgaacagctg ggcggcagcggcggcagc AAAATCGAAG AAGGTAAACT GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA GAGATCCCGG

vi. Random peptide (negative control)

Construct: NdeI (Start) + random peptide (YDEIKSRAV QAPGNFLT) + MBP + Stop + XhoI

Shown with pET24a 5’ and 3’ overlaps in bold; 1215bp total; http://web.expasy.org/randseq/

GTTTAACTTTAAGAAGGAGATATACATATG

tatgatgaaattaaaagccgcgcggtgcaggcgccgggcaactttctgacc AAAATCGAAG AAGGTAAACT

GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG

GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG

GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC

TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG

GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT

CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC

GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA

ACCCGCCAAA AACCTGGGAA GAGATCCCGG CGCTGGATAA AGAACTGAAA

GCGAAAGGTA AGAGCGCGCT GATGTTCAAC CTGCAAGAAC CGTACTTCAC

CTGGCCGCTG ATTGCTGCTG ACGGGGGTTA TGCGTTCAAG TATGAAAACG

GCAAGTACGA CATTAAAGAC GTGGGCGTGG ATAACGCTGG CGCGAAAGCG

GGTCTGACCT TCCTGGTTGA CCTGATTAAA AACAAACACA TGAATGCAGA

CACCGATTAC TCCATCGCAG AAGCTGCCTT TAATAAAGGC GAAACAGCGA

TGACCATCAA CGGCCCGTGG GCATGGTCCA ACATCGACAC CAGCAAAGTG

AATTATGGTG TAACGGTACT GCCGACCTTC AAGGGTCAAC CATCCAAACC

GTTCGTTGGC GTGCTGAGCG CAGGTATTAA CGCCGCCAGT CCGAACAAAG

AGCTGGCAAA AGAGTTCCTC GAAAACTATC TGCTGACTGA TGAAGGTCTG

GAAGCGGTTA ATAAAGACAA ACCGCTGGGT GCCGTAGCGC TGAAGTCTTA

CGAGGAAGAG TTGGCGAAAG ATCCACGTAT TGCCGCCACT ATGGAAAACG

CCCAGAAAGG TGAAATCATG CCGAACATCC CGCAGATGTC CGCTTTCTGG

TATGCCGTGC GTACTGCGGT GATCAACGCC GCCAGCGGTC GTCAGACTGT

CGATGAAGCC CTGAAAGACG CGCAGACTAA TTCGAGCTCG **TAA TAG CTC**

GAGCAC CAC CAC CAC

Gblock1 (randpMBPfor, 481bp): pET24a vector(30nt) + NdeI (Start) + Peptide + MBP

GTTTAACTTTAAGAAGGAGATATACATATG tatgatgaaattaaaagccgcgcggtgcaggcgccgggc

aactttctgacc AAAATCGAAG AAGGTAAACT GGTAATCTGG ATTAACGGCG

ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT

ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT

CCCACAGGTT GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC

ACGACCGCTT TGGTGGCTAC GCTCAATCTG GCCTGTTGGC TGAAATCACC

CCGGACAAAG CGTTCCAGGA CAAGCTGTAT CCGTTTACCT GGGATGCCGT

ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT GAAGCGTTAT

CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA

GAGATCCCGG

vii. MBP alone (negative control)

Construct: NdeI (Start) + MBP + Stop + XhoI

Shown with pET24a 5’ and 3’ overlaps in bold; 1164bp total

GTTTAACTTTAAGAAGGAGATATACATATG AAAATCGAAG AAGGTAAACT

GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG

GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG

GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC

TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG

GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT

CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC

GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA

ACCCGCCAAA AACCTGGGAA GAGATCCCGG CGCTGGATAA AGAACTGAAA

GCGAAAGGTA AGAGCGCGCT GATGTTCAAC CTGCAAGAAC CGTACTTCAC

CTGGCCGCTG ATTGCTGCTG ACGGGGGTTA TGCGTTCAAG TATGAAAACG

GCAAGTACGA CATTAAAGAC GTGGGCGTGG ATAACGCTGG CGCGAAAGCG

GGTCTGACCT TCCTGGTTGA CCTGATTAAA AACAAACACA TGAATGCAGA

CACCGATTAC TCCATCGCAG AAGCTGCCTT TAATAAAGGC GAAACAGCGA

TGACCATCAA CGGCCCGTGG GCATGGTCCA ACATCGACAC CAGCAAAGTG

AATTATGGTG TAACGGTACT GCCGACCTTC AAGGGTCAAC CATCCAAACC

GTTCGTTGGC GTGCTGAGCG CAGGTATTAA CGCCGCCAGT CCGAACAAAG

AGCTGGCAAA AGAGTTCCTC GAAAACTATC TGCTGACTGA TGAAGGTCTG

GAAGCGGTTA ATAAAGACAA ACCGCTGGGT GCCGTAGCGC TGAAGTCTTA

CGAGGAAGAG TTGGCGAAAG ATCCACGTAT TGCCGCCACT ATGGAAAACG

CCCAGAAAGG TGAAATCATG CCGAACATCC CGCAGATGTC CGCTTTCTGG

TATGCCGTGC GTACTGCGGT GATCAACGCC GCCAGCGGTC GTCAGACTGT

CGATGAAGCC CTGAAAGACG CGCAGACTAA TTCGAGCTCG **TAA TAG CTC**

GAGCAC CAC CAC CAC

Gblock1 (MBPfor, 430bp): pET24a vector(30nt) + NdeI (Start) + MBP

GTTTAACTTTAAGAAGGAGATATACATATG AAAATCGAAG AAGGTAAACT GGTAATCTGG ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA GAGATCCCGG

viii. MBP (also known as Gblock2 for the end half of MBP sequence)

Gblock2 (MBPend, 774bp): MBP + Stop + XhoI + pET24a vector (24nt)

Overlaps with Gblock1 by 40nt, shown underlined.

CTGCTGCCGA ACCCGCCAAA AACCTGGGAA GAGATCCCGG CGCTGGATAA AGAACTGAAA GCGAAAGGTA AGAGCGCGCT GATGTTCAAC CTGCAAGAAC CGTACTTCAC CTGGCCGCTG ATTGCTGCTG ACGGGGGTTA TGCGTTCAAG TATGAAAACG GCAAGTACGA CATTAAAGAC GTGGGCGTGG ATAACGCTGG CGCGAAAGCG GGTCTGACCT TCCTGGTTGA CCTGATTAAA AACAAACACA TGAATGCAGA CACCGATTAC TCCATCGCAG AAGCTGCCTT TAATAAAGGC GAAACAGCGA TGACCATCAA CGGCCCGTGG GCATGGTCCA ACATCGACAC CAGCAAAGTG AATTATGGTG TAACGGTACT GCCGACCTTC AAGGGTCAAC CATCCAAACC GTTCGTTGGC GTGCTGAGCG CAGGTATTAA CGCCGCCAGT CCGAACAAAG AGCTGGCAAA AGAGTTCCTC GAAAACTATC TGCTGACTGA

TGAAGGTCTG GAAGCGGTTA ATAAAGACAA ACCGCTGGGT GCCGTAGCGC

TGAAGTCTTA CGAGGAAGAG TTGGCGAAAG ATCCACGTAT TGCCGCCACT

ATGGAAAACG CCCAGAAAGG TGAAATCATG CCGAACATCC CGCAGATGTC

CGCTTTCTGG TATGCCGTGC GTACTGCGGT GATCAACGCC GCCAGCGGTC

GTCAGACTGT CGATGAAGCC CTGAAAGACG CGCAGACTAA TTCGAGCTCG**TAA**

TAG CTC GAGCAC CAC CAC CAC

Bibliography

1. Toraya, T., Honda, S. & Fukui, S. Fermentation of 1,2-Propanediol and 1,2-Ethanediol by Some Genera of Enterobacteriaceae, Involving Coenzyme B12- Dependent Diol Dehydratase. *J. Bacteriol.* **139,** 39–47 (1979).

2. Rondon, M. R. & Escalante-semerena, J. C. The poc Locus Is Required for 1 , 2-Propanediol-Dependent Transcription of the Cobalamin Biosynthetic (cob) and Propanediol Utilization (pdu) Genes of Salmonella typhimurium. *J. Bacteriol.* **174,** 2267–2272 (1992).

3. Bobik, T. a, Ailion, M. & Roth, J. R. A single regulatory gene integrates control of vitamin B12 synthesis and propanediol degradation. *J. Bacteriol.* **174,** 2253–66 (1992).

4. Obradors, N., Badía, J., Baldomà, L. & Aguilar, J. Anaerobic metabolism of the L-rhamnose fermentation product 1,2-propanediol in Salmonella typhimurium. *J. Bacteriol.* **170,** 2159–62 (1988).

5. Tocheva, E. I. *et al.* Structure and expression of propanediol utilization microcompartments in Acetonema longum. *J. Bacteriol.* **196,** 1651–1658 (2014).

6. Heath, E. C. & Ghalambor, M. A. The Metabolism of l-Fucose: I. The Purification and Properties of l-Fuculose Kinase. *J. Biol. Chem.* **237,** 2423–2426 (1962).

7. Cocks, G. T., Aguilar, J. & Lin, E. C. C. Evolution of L-1 , 2-Propanediol Catabolism in Escherichia coli by Recruitment of Enzymes for L-Fucose and L-Lactate Metabolism. *J. Bacteriol.* **118,** 83–88 (1974).

8. Bobik, T. A., Havemann, G. D., Busch, R. J., Williams, D. S. & Aldrich, H. C. 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 (1999).

9. Jorda, J., Lopez, D., Wheatley, N. M. & Yeates, T. O. Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci.* **22,** 179–95 (2013).

10. Havemann, G. D. & Bobik, T. A. Protein Content of Polyhedral Organelles Involved in Coenzyme B 12 -Dependent Degradation of 1 , 2-Propanediol in Salmonella enterica Serovar Typhimurium LT2 †. *J. Bacteriol.* **185,** 5086–5095 (2003).

11. Sampson, E. M. & Bobik, T. a. Microcompartments for B12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. *J. Bacteriol.* **190,** 2966–71 (2008).

12. Parsons, J. B. *et al.* Biochemical and structural insights into bacterial organelle form and biogenesis. *J. Biol. Chem.* **283,** 14366–75 (2008).

13. Lt, T., Havemann, G. D., Sampson, E. M., Thomas, A. & Bobik, T. A. PduA Is a Shell Protein of Polyhedral Organelles Involved in Coenzyme B 12 in Salmonella enterica Serovar PduA Is a Shell Protein of Polyhedral Organelles Involved in Coenzyme B 12 - Dependent Degradation of 1 , 2-Propanediol in Salmonella enterica Serovar. (2002). doi:10.1128/JB.184.5.1253

14. Crowley, C. S. *et al.* Structural insight into the mechanisms of transport across the Salmonella enterica Pdu microcompartment shell. *J. Biol. Chem.* **285,** 37838–46 (2010).

15. Singleton, C., Howard, T. P. & Smirnoff, N. Synthetic metabolons for metabolic engineering. *J. Exp. Bot.* **65,** 1947–1954 (2014).

16. Kim, E. Y. & Tullman-Ercek, D. Engineering nanoscale protein compartments for synthetic organelles. *Curr. Opin. Biotechnol.* **24,** 627–32 (2013).

17. Parsons, J. B. *et al.* Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. *Mol. Cell* **38,** 305–15 (2010).

18. Sargent, F. *et al.* A synthetic system for expression of components of a bacterial microcompartment. *Microbiology* **159,** 2427–36 (2013).

19. Lassila, J. K., Bernstein, S. L., Kinney, J. N., Axen, S. D. & Kerfeld, C. a. Assembly of Robust Bacterial Microcompartment Shells using Building Blocks from an Organelle of Unknown Function. *J. Mol. Biol.* (2014). doi:10.1016/j.jmb.2014.02.025

20. Cai, F., Sutter, M., Bernstein, S. L., Kinney, J. N. & Kerfeld, C. a. Engineering Bacterial Microcompartment Shells: Chimeric Shell Proteins and Chimeric Carboxysome Shells. *ACS Synth. Biol.* (2014). doi:10.1021/sb500226j

21. Frank, S., Lawrence, A. D., Prentice, M. B. & Warren, M. J. Bacterial microcompartments moving into a synthetic biological world. *J. Biotechnol.* **163,** 273–9 (2013).

22. Kerfeld, C. a *et al.* Protein structures forming the shell of primitive bacterial organelles. *Science* **309,** 936–8 (2005).

23. Chowdhury, C. *et al.* Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proc. Natl. Acad. Sci.* **112,** 2990–2995 (2015).

24. Iannuzzi, M., Laio, A. & Parrinello, M. Efficient Exploration of Reactive Potential Energy Surfaces Using Car-Parrinello Molecular Dynamics. *Phys. Rev. Lett.* **90,** 238302 (2003).

25. Barducci, A., Bussi, G. & Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **100,** 020603 (2008).

26. Bonomi, M. & Parrinello, M. Enhanced Sampling in the Well-Tempered Ensemble. *Phys. Rev. Lett.* **104,** 190601 (2010).

27. Tanaka, S., Sawaya, M. R. & Yeates, T. O. Structure and mechanisms of a protein-based organelle in Escherichia coli. *Science* **327,** 81–4 (2010).

28. Takenoya, M., Nikolakakis, K. & Sagermann, M. 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,** 6056–63 (2010).

29. Schlick, T. Molecular dynamics-based approaches for enhanced sampling of long-time, large-scale conformational changes in biomolecules. *F1000 Biol. Rep.* **1,** 51 (2009).

30. Pierce, L. C. T., Salomon-Ferrer, R., Augusto F de Oliveira, C., McCammon, J. A. & Walker, R. C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **8,** 2997–3002 (2012).

31. Hamelberg, D., Mongan, J. & McCammon, J. A. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **120,** 11919–29 (2004).

32. Durrant, J., de Oliveira, C. & McCammon, J. POVME: An Algorithm for Measuring Binding-Pocket Volumes. *J. Mol. Graph. Model.* **29,** 773–776 (2011).

33. Thompson, M. C., Cascio, D., Leibly, D. J. & Yeates, T. O. An allosteric model for control of pore opening by substrate binding in the EutL microcompartment shell protein. *Protein Sci.* **24,** 956–75 (2015).

34. Schlitter, J., Engels, M., Krüger, P., Jacoby, E. & Wollmer, a. Targeted Molecular Dynamics Simulation of Conformational Change-Application to the T ↔ R Transition in Insulin. *Mol. Simul.* **10,** 291–308 (1993).

35. Klein, M. G. *et al.* Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. *J. Mol. Biol.* **392,** 319–33 (2009).

36. Dryden, K. a, Crowley, C. S., Tanaka, S., Yeates, T. O. & Yeager, M. Two-dimensional crystals of carboxysome shell proteins recapitulate the hexagonal packing of three-dimensional crystals. *Protein Sci.* **18,** 2629–35 (2009).

37. Pang, A., Liang, M., Prentice, M. B. & Pickersgill, R. W. Substrate channels revealed in the trimeric Lactobacillus reuteri bacterial microcompartment shell protein PduB. *Acta Crystallogr. D. Biol. Crystallogr.* **68,** 1642–52 (2012).

38. Leibly, D. J. *et al.* Stabilizing additives added during cell lysis aid in the solubilization of recombinant proteins. *PLoS One* **7,** e52482 (2012).

39. Fan, C. *et al.* Short N-terminal sequences package proteins into bacterial microcompartments. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 7509–14 (2010).

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

41. Fan, C. & Bobik, T. A. The N-Terminal Region of the Medium Subunit (PduD) Packages Adenosylcobalamin-Dependent Diol Dehydratase (PduCDE) into the Pdu Microcompartment The N-Terminal Region of the Medium Subunit (PduD) Packages Adenosylcobalamin-Dependent Diol Dehydratase. *J. Bacteriol.* **193,** 5623 (2011).

42. Jorda, J., Liu, Y., Bobik, T. a & Yeates, T. O. Exploring Bacterial Organelle Interactomes: A Model of the Protein-Protein Interaction Network in the Pdu Microcompartment. *PLoS Comput. Biol.* **11,** e1004067 (2015).

43. Cameron, J. C., Wilson, S. C., Bernstein, S. L. & Kerfeld, C. a. Biogenesis of a bacterial organelle: the carboxysome assembly pathway. *Cell* **155,** 1131–40 (2013).

44. Aussignargues, C., Paasch, B. C., Gonzalez-Esquer, R., Erbilgin, O. & Kerfeld, C. a. Bacterial microcompartment assembly: The key role of encapsulation peptides. *Commun. Integr. Biol.* **8,** e1039755 (2015).

45. Fan, C., Cheng, S., Sinha, S. & Bobik, T. a. Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 14995–5000 (2012).

46. Tobimatsu, T., Kawata, M. & Toraya, T. The N-Terminal Regions of and Subunits Lower the Solubility of Adenosylcobalamin-Dependent Diol Dehydratase. **69,** 455–462 (2005).

47. Akita, K. *et al.* Purification and some properties of wild-type and N-terminal-truncated ethanolamine ammonia-lyase of Escherichia coli. *J. Biochem.* **147,** 83–93 (2010).

48. Shibata, N. *et al.* Crystal structures of ethanolamine ammonia-lyase complexed with coenzyme B12 analogs and substrates. *J. Biol. Chem.* **285,** 26484–93 (2010).

49. Spurlino, J. C., Lu, G.-Y. & Quiocho, F. A. The 2 . 3-A Resolution Structure of the Maltose- or Maltodextrin- binding Protein , A Primary Receptor of Bacterial Active Transport and Chemotaxis *. *J. Biol. Chem.* **266,** 5202–5219 (1991).

50. Kapust, R. B. & Waugh, D. S. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8,** 1668–1674 (1999).

51. Sutter, M. *et al.* Visualization of Bacterial Microcompartment Facet Assembly Using High-Speed Atomic Force Microscopy. *Nano Lett.* (2015). doi:10.1021/acs.nanolett.5b04259