Title
Elucidating the Native Topologies of Bacterial Microcompartments

Author
Ochoa, Jessica Micaela

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Elucidating the Native Topologies of Bacterial Microcompartments

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in Molecular Biology

by

Jessica Micaela Ochoa

2021

ABSTRACT OF THE DISSERTATION

Elucidating the Native Topologies of Bacterial Microcompartments

By

Jessica Micaela Ochoa

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2021

Professor Todd O. Yeates, Chair

Bacterial microcompartments, or MCPs, are supramolecular structures analogous to eukaryotic organelles. These assemblies are understood to organize and compartmentalize the cytosol by sequestering and optimizing a myriad of metabolic reactions. They enable the diffusive transport of substrates, acting as a semi-permeable barrier that prevents the efflux of toxic intermediates. Unlike their membrane-bound eukaryotic counterparts, MCPs are comprised entirely of a proteinaceous outer shell. The so-called BMC domain (for **b**acterial **m**icrocom**p**artment) is the canonical building block for all MCPs. BMCs oligomerize to form hexamers disks which tesselate laterally to form the nearly flat facets of the outer MCP shell. Pentameric proteins form the polyhedral vertices. Despite persistent research in the field, bacterial microcompartments remain poorly understood. This work seeks to enhance our foundational knowledge of MCPs through structurally characterizing BMS shell proteins and highlighting their potentially specialized functions, presenting a novel structural resource that organizes existing structural data and recent work that begins to reveal the native topology and long-range organization of MCP shells.

The dissertation of Jessica Micaela Ochoa is approved.

Ellen May Sletten

Feng Guo

James U. Bowie

Joseph Ambrose Loo

Todd O. Yeates, Committee Chair

University of California, Los Angeles

2021

For my loving family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I want to first thank my family. Without them this journey would not have been possible or meaningful. My grandparents, Lolo and Elena, my parents, Robert and Letty and my sister, Marissa. I want to thank the complete Ochoa and Mendoza families especially my Tías Elvira, Lucha and Cota. I want to thank personal support team and my dear friends at UCLA; Jazlyn Mooney, Andrew Lopez, Christopher Robles, Christian Henry who have been on this road with me since day one. I want to thank my friends who encouraged me and helped me maintain my sanity when I needed a break from the academy: Julia Lawniczak, Renee Dell'Acqua, Sylvia Vetrone, Lupe Alvarez and Jessica and Jason Odle.

Next, I would like to acknowledge my advisor Todd Yeates, whose support, patience, encouragement, and faith in me has helped me develop into the scientist I am today, thank you. I would also like to thank the individuals who have shaped and encouraged my scientific journey; Vann Priest, Auliki Flagan, Jeffrey Hammer, Lindsey Durbin, Devin Iimoto, Ralph Isovitsch and especially Sylvia Vetrone. I also thank the other members of my committee Ellen Sletten, Jim Bowie, Feng Guo and Joe Loo, who have given me their insightful feedback and words of support.

Lastly, I wish to acknowledge my lab mates, collaborators and colleagues who have helped guide me through my research and this PhD. In particular, from the Yeates Lab past and present: Justin Miller, Kevin Cannon, Kyle Meador, Roger Graells Castells, Yuxi Liu, Matthew Agdanowski, Eric Lee, David Leibly, Sunny Chang and Cindy Chau; The brilliant undergraduates I had the privilege to work with especially Vy Nguyen, Mengxiao Nie, Andrea Acosta, Oscar Mijares, Nancy Rivera-Leon and Xavier Escoto; The UCLA DOE: Duillio Cascio, Michael Sawaya, Dan Anderson, Mark Arbing, Michael Collazo, Tom Holton and Genesis Falcon My second-floor support team; Einav Tayeb-Fligellman, Jeannette Bowler and Sarah Greiner. The Loo lab, especially: Janine Fu, Rachel Loo and Kate Liu; The Bobik lab from ISU especially; Thomas

Bobik, Chiranjit Chowdhurry and Sujit Mohanty; The Rodriguez lab, especially Jose Rodriguez and Marcus Gallagher-Jones.

VITA

EDUCATION

2012	Bachelor of Arts, Biology Whittier College | Whittier, CA
2012	Bachelor of Arts, Chemistry Whittier College | Whittier, CA

PROFESSIONAL

2014 – 2015	Lean Project Manager Eurofins Eaton Analytical | Monrovia, CA
2012 – 2014	Scientist I Eurofins Easton Analytical | Monrovia, CA

PUBLICATIONS

Jessica M. Ochoa and Todd O. Yeates. Recent Structural Insights into Bacterial Microcompartment Shells. Current Opinion in Microbiology. April (2021). *Accepted*

Ochoa JM, Bair K, Holton T, Bobik TA, Yeates TO (2021) MCPdb: The bacterial microcompartment database. PLOS ONE 16:e0248269.

Ochoa, J.M., Nguyen, V.N., Nie, M., Sawaya, M.R., Bobik, T.A., and Yeates, T.O. (2020). Symmetry breaking and structural polymorphism in a bacterial microcompartment shell protein for choline utilization. Protein Science *29*, 2201–2212.

Cannon, K. A*., **Ochoa, J.***, Yeates, T. O. High-symmetry protein assemblies: patterns and emerging applications. Current Opinion in Structural Biology. (2019); 77-84. doi: 10.1016/j.sbi.2019.03.008. **KAC and JMO contributed equally.*

Marcos Herrera-Vaquero, Danielle Bouquio, Martin Kallab, Karl Biggs, Gayatri Nair, **Jessica Ochoa**, Antonio Heras-Garvin, Christian Heid, Inesa Hadrovic, Werner Poewe, Gregor K. Wenning, Frank-Gerrit Klärner, Thomas Schrader, Gal Bitan, Nadia Stefanova. The molecular tweezer CLR01 reduces aggregated, pathologic, and seeding-competent α-synuclein in experimental multiple system atrophy (2019) Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease,1865(11):165513.

Callaway MK, **Ochoa JM**, Perez EE, Ulrich PE, Alocilja EC, and Vetrone SA, et al. (2013) Investigation of the Toxicity of Amine-coated, Carboxyl-coated and Polyaniline-coated FeO Magnetic Nanoparticles in *Caenorhabditis elegans*. J Biosens Bioelectron 4: 145.

Setterington, E.B., Cloutier, B.C., **Ochoa, J.M.**, Cloutier, A.K., Jain, P., Alocilja, E.C. 2011. Rapid, sensitive, and specific immunomagnetic separation of foodborne pathogens. International Journal of Food Safety, Nutrition and Public Health, 4(1):83-100.

Chapter 1: Introduction

Bacterial microcompartments, or MCPs, are organelle-like structures found in nearly 20% of bacteria across several phyla that act as semipermeable barriers capable of sequestering substrates and enzymes and carrying out sensitive metabolic reactions [1–3]. However, unlike their eukaryotic membrane-bound counterparts, MCPs are comprised entirely of proteins. These are supramolecular structures are approximately 100 - 400 nm in diameter and help to organize and compartmentalize the cytosol by providing an enclosed environment surrounded by a selectively permeable proteinaceous shell [2,4–6]. These extraordinary structures optimize numerous multi-step metabolic reactions by creating microenvironment that bring various enzymes, substrates and cofactors into close proximity. The co-localization of these various metabolic components helps to optimize metabolic flux while preventing the escape of toxic and/or volatile intermediates into the rest of the cytosol [2,7–9]. MCPs are defined by their signature enzyme. The founding member, the carboxysome, enhance carbon dioxide fixation by utilizing sequentially acting enzymes, carbonic anhydrase and ribulose- 1,5-bisphosphate carboxylase/oxygenase(RuBisCO), convert bicarbonate and ribulose-bisphosphate to its subsequent products [2,10,11]. By contrast, metabolomes utilize various key enzymes in order to consume a variety of substrates including 1,2-propanediol and ethanolamine for the propanediol utilization (Pdu) MCP and ethanolamine utilization (Eut) MCP, respectively [10,12,13]. Recent work has also revealed that other microcompartment types are capable of utilizing glycyl-radical chemistry (including the glycyl-radical propanediol MCP and the choline utilization MCP) while others have metabolic functions that remain only partially understood including the RMM/AAUM (Rhodococcus and Mycolicibacterium Microcompartment, subsequently renamed AAUM for its apparent role in amino acetone utilization) MCP and the Etu (ethanol utilization) MCP [14–20].

While MCPs are functionally diverse, they are unified by their basic architecture; thousands of copies of the BMC (bacterial microcompartment) domain oligomerize to form the

canonical hexameric disks that tesselate to form the nearly flat facets of the outer MCP shell [5]. Pentameric BMV (bacterial microcompartment vertex) proteins, which are evolutionary distinct from BMCs, cap and form the vertices of the nearly polyhedral MCP shell [5,21]. For any given MCP type, there are numerous paralogs, ranging in number from two to seven, that comprise the full MCP shell. Different BMC shell types enable modularity and afford specialized roles including binding internal enzymes and substrates and also providing routes for diffusion and molecular transport. This modularity arises from variations in the canonical BMC domain. BMC-H shell proteins (H is for hexamer), the most abundant type, is comprised of a single BMC domain that forms a cyclic homohexamer [22–25]. Permuted BMCs retain their canonical tertiary structure and overall hexameric architecture but give rise to alternate topological forms because of a circular permutation in the BMC sequence [26,27]. BMC-Ts (T is for tandem) contain two tandem repeats of the BMC domain and form trimers or pseudohexameric disks [28,29].

In recent years, biochemical, genetic and novel engineering studies have helped to shed light on microcompartments and their shell proteins. Despite growing research and advancements, there are several unresolved questions. The roles and potentially specialized functions of permuted BMCs remain poorly understood, with previous structural characterization creating ambiguity in overall quaternary structure. Definitive mechanisms of molecular transport across the protein shell remains elusive prompting questions surrounding the roles of surface electrostatic potentials of BMC shell proteins and their associated charged substrates. Lastly, outstanding questions remain surrounding the native topology and long-range organization of MCP shells. This work seeks to enhance our fundamental understanding of bacterial microcompartments by addressing these questions.

In **Chapter 2**, I present a review of high-symmetry protein assemblies. Here I highlight large supramolecular protein structures that have been characterized, their roles in guiding novel engineering feats and their recent applications for exterior display and interior encapsulation. I

survey the PDB for known natural and designed cubic protein assemblies (tetrahedral, octahedral and icosahedral), providing biological insight and presenting recent applications. This work has been published as:

Cannon, K. A.*, Ochoa, J. M.*, & Yeates, T. O. (2019). High-symmetry protein assemblies: Patterns and emerging applications. *Current Opinion in Structural Biology*, *55*, 77–84. (* indicates equal contribution).

In **Chapter 3**, I present a brief review that focuses on bacterial microcompartment shell proteins. I present the key structural features of the BMC domain (Pfam 00936) as well as their unique variations. An abundance of structural information on the BMC domain can be found in the PDB, with more than 110 microcompartment shell proteins structurally characterized and deposited since 2005. I provide an overview on the five BMC variations; BMC-H, Permuted BMC, BMC-$T^{(+)}$, Permuted BMC-$T^{(+)}$ and Permuted BMC-$T^{(-)}$. I present variations in observed quaternary structures from flat hexamers, twisted or bent hexamers, six-fold screws with varying pitch and stacked disks with open and closed central pores and current hypotheses on their specialized roles in diffusion and transport. Lastly, I present a growing list of miniaturized recombinant microcompartment shells, discussing their overall topology and roles in articulating broader assembly principles of bacterial microcompartments. A version of this work has been accepted for publication in *Current Opinions in Microbiology.*

In **Chapter 4**, I structurally characterize the numerous polymorphs of CutR, a permuted bacterial microcompartment shell protein from *Streptococcus intermedius*. The permuted BMC domain has long since been a shell protein of interest, with previous work demonstrating this homolog was capable for forming traditional flat hexamers as well as twisted/bent versions. In this work solved five structures: two traditional flat hexamers, two six-fold screws of varying pitch and one novel dimer. The observed dimer possessed a domain swap however I believe this quaternary form to be a crystallographic artifact and not biologically relevant. I also found that

CutR was capable of forming two distinct six-fold screws of varying pitch (41.9 Å and 33.8 Å). These screws retain key intermolecular interactions between monomers and give rise to the overall hexameric shape when looking down the six-fold axis of symmetry. I found that the presence of an N-terminal His-6 purification tag does not preclude formation of a traditional flat hexamer. I also compare these four CutR hexamers to previously published permuted BMC hexamers. I describe their variation and deformation using a novel graphical approach to compare adjacent monomer, demonstrating the greatest degree of deformation is observed in the previously characterized EutS homolog. This work adds to the growing library of permuted BMC structures and provides additional evidence that permuted BMCs are capable of forming non-canonical quaternary suggesting the potential for specialized roles in dynamics, flexibility and overall MCP structure. This work has been published as:

Ochoa, J. M., Nguyen, V. N., Nie, M., Sawaya, M. R., Bobik, T. A., & Yeates, T. O. (2020). Symmetry breaking and structural polymorphism in a bacterial microcompartment shell protein for choline utilization. *Protein Science*, *29*(11), 2201–2212.

In **Chapter 5**, I present a novel online tool, MCDPdb: The Bacterial Microcompartment Database (https://mcpdb.mbi.ucla.edu/). To date, there has been no centralized resource that quickly and efficiently provides members with a rapid way of understanding function, structure and diversity of the more than 150 microcompartment-associated structures in the PDB. I briefly present the canonical topological variations of BMC and BMV (**b**acterial **m**icrocompartment **v**ertex) proteins and describe the broad categories of MCP and encapsulin-related structures in the PDB including; BMC-H, Permuted BMC, BMC-T, large assemblies, enzymes, encpasulins and other. I collected relevant data from the Protein Data Bank (PDB) UniProt and linked data using the PDB ID as a primary key. I also provide expertly curated annotations for each of the 163 entries by combining data from the PDB and UniProt to describe the *MCP Type* from which the structure derives, the *MCP Classification*, the *Protein Type and Topology* and *Observed*

Assembly Form. Further, I generated a set of figure-ready images, PyMOL sessions and an in-browser 3D-viewer, providing users with a seamless and effortless way to learn about an engage with the numerous structures in this database. This work provides novices and experts alike with a novel and efficient way of learning about and engaging with structural data about MCPs. More importantly, it provides expertly-curated information about their relevant biological assemblies. This work has been published as:

Ochoa, J. M., Bair, K., Holton, T., Bobik, T. A., & Yeates, T. O. (2021). MCPdb: The bacterial microcompartment database. *PLOS ONE*, *16*(3), e0248269.

In **Chapter 6**, I structurally characterize six hexameric shell proteins derived from four BMC proteins from two types of choline utilization bacterial microcompartments. Despite the vast amounts of structural data available for BMC shell proteins, there is a limited amount of information on hexameric shell proteins derived from choline utilization (Cut) MCPs. Thus, I sought to structurally characterize one homolog from the type I Cut MCP, CutN (from *Streptococcus intermedius*) and three paralogs from the type II Cut MCP; CmcA, CmcB and CmcC (from *Escherichia coli 536)*. Because their choline substrate is charged, I also sought to describe the differences in surface electrostatics. To this end, I structurally characterized point mutants and used the Adaptive Poisson-Boltzman Solver plugin in PyMOL to calculate the surface electrostatic potentials of all six structures. I solved the structures to 1.7 – 3.0 Å resolution and found that the monomers of all four homologs have an RMSD of less than 0.42 Å. Furthermore, I found that all homologs possess a positively charged central pore on their flat faces, while there is variability in overall charge distribution between the dimpled faces of the four homologs. Thus, this work provides foundational data that can be used in future work in order to assess and characterize the recruitment of the choline substrate to Cut MCPs. A version of this work has been submitted for publication as:

Ochoa JM, Mijares O, Acosta AA, Escoto X, Leon-Rivera N, Marshall JD, Sawaya MR and Yeates TO. Structural characterization of hexameric shell proteins from two types of choline utilization bacterial microcompartments. *Acta Crystallographica Section F*.

In **Chapter 7**, I describe recent work in which we seek to characterize the overall topology of endogenous 1,2-propanediol (Pdu) microcompartments. In this ongoing work I describe optimized facile methods for the robust expression and purification of Pdu MCPs. Consistent, structurally-sound, high-yield preps have enabled novel areas for probing the Pdu microcompartment. I describe current efforts to characterize the native orientation of BMC shell proteins in the context of native MCPs by utilizing endogenous metabolic machinery to overproduce propionaldehyde, a highly-reactive reaction intermediate, and tandem mass spectrometry to investigate differences in lysine modifications between internal lumen-facing lysines versus external cytosol-oriented lysines. I also present current work in optimization and data processing as I seek to define the overall geometries of the Pdu MCP and work towards using existing structural data and cryo-electron tomography to define and characterize individual BMC shell proteins within the context of a full MCP shell.

It is important to note that here and throughout this dissertation, MCP refers to the full microcompartment while BMC refers specifically to the shell proteins, which possess the BMC domain and whose quaternary structures are hexameric (or pseudohexmaric in the case of tandem-BMC domain shell proteins).

References

1. Jorda J, Lopez D, Wheatley NM, Yeates TO: Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. Protein Science 2013, 22:179–195.
2. Yeates TO, Kerfeld CA, Heinhorst S, Cannon GC, Shively JM: Protein-based organelles in bacteria: carboxysomes and related microcompartments. Nature Reviews Microbiology 2008, 6:681–691.
3. Kerfeld CA, Aussignargues C, Zarzycki J, Cai F, Sutter M: Bacterial microcompartments. Nature Reviews Microbiology 2018, doi:10.1038/nrmicro.2018.10.
4. Cheng S, Liu Y, Crowley CS, Yeates TO, Bobik TA: Bacterial microcompartments: their properties and paradoxes. Bioessays 2008, 30:1084–1095.
5. Yeates TO, Thompson MC, Bobik TA: The protein shells of bacterial microcompartment organelles. Current Opinion in Structural Biology 2011, 21:223–231.
6. Chowdhury C, Sinha S, Chun S, Yeates TO, Bobik TA: Diverse Bacterial Microcompartment Organelles. Microbiol Mol Biol Rev 2014, 78:438–468.
7. Bobik TA, Havemann GD, Busch RJ, Williams DS, Aldrich HC: The Propanediol Utilization (pdu) Operon ofSalmonella enterica Serovar Typhimurium LT2 Includes Genes Necessary for Formation of Polyhedral Organelles Involved in Coenzyme B12-Dependent 1,2-Propanediol Degradation. J Bacteriol 1999, 181:5967–5975.
8. Shively JM, Ball F, Brown DH, Saunders RE: Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of Thiobacillus neapolitanus. Science 1973, 182:584–586.
9. Stewart AM, Stewart KL, Yeates TO, Bobik TA: Advances in the World of Bacterial Microcompartments. Trends in Biochemical Sciences 2021, doi:10.1016/j.tibs.2020.12.002.

10. Kerfeld CA, Heinhorst S, Cannon GC: Bacterial microcompartments. Annu Rev Microbiol 2010, 64:391–408.
11. Cannon GC, Bradburne CE, Aldrich HC, Baker SH, Heinhorst S, Shively JM: Microcompartments in Prokaryotes: Carboxysomes and Related Polyhedra. Appl Environ Microbiol 2001, 67:5351–5361.
12. Bobik TA, Lehman BP, Yeates TO: Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways. Molecular Microbiology 2015, 98:193–207.
13. Stewart KL, Stewart AM, Bobik TA: Prokaryotic Organelles: Bacterial Microcompartments in E. coli and Salmonella. EcoSal Plus 2020, 9.
14. Herring TI, Harris TN, Chowdhury C, Mohanty SK, Bobik TA: A Bacterial Microcompartment Is Used for Choline Fermentation by Escherichia coli 536. J Bacteriol 2018, 200:e00764-17.
15. Kalnins G, Cesle E-E, Jansons J, Liepins J, Filimonenko A, Tars K: Encapsulation mechanisms and structural studies of GRM2 bacterial microcompartment particles. Nature Communications 2020, 11.
16. Axen SD, Erbilgin O, Kerfeld CA: A Taxonomy of Bacterial Microcompartment Loci Constructed by a Novel Scoring Method. PLOS Computational Biology 2014, 10:e1003898.
17. Ferlez B, Sutter M, Kerfeld CA: Glycyl Radical Enzyme-Associated Microcompartments: Redox-Replete Bacterial Organelles. mBio 2019, 10.
18. Zarzycki J, Sutter M, Cortina NS, Erb TJ, Kerfeld CA: In Vitro Characterization and Concerted Function of Three Core Enzymes of a Glycyl Radical Enzyme - Associated Bacterial Microcompartment. Sci Rep 2017, 7.

19. Mallette E, Kimber MS: Structural and kinetic characterization of (S)-1-amino-2-propanol kinase from the aminoacetone utilization microcompartment of Mycobacterium smegmatis. J Biol Chem 2018, doi:10.1074/jbc.RA118.005485.
20. Ravcheev DA, Moussu L, Smajic S, Thiele I: Comparative Genomic Analysis Reveals Novel Microcompartment-Associated Metabolic Pathways in the Human Gut Microbiome. Front Genet 2019, 10.
21. Wheatley NM, Gidaniyan SD, Liu Y, Cascio D, Yeates TO: Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. Protein Sci 2013, 22:660–665.
22. Havemann GD, Sampson EM, Bobik TA: PduA Is a Shell Protein of Polyhedral Organelles Involved in Coenzyme B12-Dependent Degradation of 1,2-Propanediol in Salmonella enterica Serovar Typhimurium LT2. J Bacteriol 2002, 184:1253–1261.
23. Chowdhury C, Chun S, Sawaya MR, Yeates TO, Bobik TA: The function of the PduJ microcompartment shell protein is determined by the genomic position of its encoding gene. Molecular Microbiology 2016, 101:770–783.
24. Tanaka S, Kerfeld CA, Sawaya MR, Cai F, Heinhorst S, Cannon GC, Yeates TO: Atomic-Level Models of the Bacterial Carboxysome Shell. Science 2008, 319:1083–1086.
25. Kerfeld CA, Sawaya MR, Tanaka S, Nguyen CV, Phillips M, Beeby M, Yeates TO: Protein Structures Forming the Shell of Primitive Bacterial Organelles. Science 2005, 309:936–938.
26. Crowley CS, Sawaya MR, Bobik TA, Yeates TO: Structure of the PduU Shell Protein from the Pdu Microcompartment of Salmonella. Structure 2008, 16:1324–1332.
27. Tanaka S, Sawaya MR, Yeates TO: Structure and Mechanisms of a Protein-Based Organelle in Escherichia coli. Science 2010, 327:81–84.

28. Pang A, Warren MJ, Pickersgill RW: Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe–4S cluster-binding site. Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr 2011, 67:91–96.
29. Sutter M, Greber B, Aussignargues C, Kerfeld CA: Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. Science 2017, 356:1293–1297.

Chapter 2: High-symmetry protein assemblies: patterns and emerging applications

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

High-symmetry protein assemblies: patterns and emerging applications

Kevin A Cannon[1,2,4], Jessica M Ochoa[1,3,4] and Todd O Yeates[1,2,3]

The accelerated elucidation of three-dimensional structures of protein complexes, both natural and designed, is providing new examples of large supramolecular assemblies with intriguing shapes. Those with high symmetry – based on the geometries of the Platonic solids – are particularly notable as their innately closed forms create interior spaces with varying degrees of enclosure. We survey known protein assemblies of this type and discuss their geometric features. The results bear on issues of protein function and evolution, while also guiding novel bioengineering applications. Recent successes using high-symmetry protein assemblies for applications in interior encapsulation and exterior display are highlighted.

Addresses
[1] UCLA Department of Chemistry and Biochemistry, United States
[2] UCLA-DOE Institute for Genomics and Proteomics, United States
[3] UCLA Molecular Biology Institute, United States

Corresponding author: Yeates, Todd O (yeates@mbi.ucla.edu)
[4] KAC and JMO contributed equally.

Current Opinion in Structural Biology 2019, **55**:77–84

This review comes from a themed issue on **Macromolecular assemblies**

Edited by **Andrzej Joachimiak** and **Ilya Vakser**

https://doi.org/10.1016/j.sbi.2019.03.008

Introduction

Nearly half of all known natural proteins form homo-oligomeric complexes through the assembly of multiple copies of the same (or homologous) subunits [1–3]. Such structures have evolved for diverse purposes, ranging from cooperative binding behavior (as exemplified by hemoglobin) to architectural function (as exemplified by microtubules and other large cellular structures). Yet specific assembly advantages are known for only a fraction of the vast number of homo-oligomeric protein and enzyme structures seen in nature. The reasons for protein oligomerization have been discussed as far back as Monod [4], and carefully surveyed by Goodsell [5]. The wide-ranging advantages that have been offered as explanations – enhanced stability, functional regulation, and mitigation of cellular crowding, to name a few – suggest the possibility of multifunctional advantages. Another perspective is that the diversity of explanations for homomeric assemblies belies an incomplete understanding of the phenomenon, suggesting that continued investigations of trends and outlying cases might be informative.

A nearly universal observation is that the subunits in homo-oligomeric assemblies are arranged in symmetric ways. Compared to the more open question of why so many proteins form homo-oligomers in the first place, the explanation of symmetry is clearer [5]. Symmetric arrangements require the fewest distinct kinds of interfaces between equivalent subunits. This makes symmetric arrangements more likely to occur through natural evolutionary events, as articulated by Crick and Watson in their prescient 1956 prediction that viral capsids would assemble according to cubic symmetries [6]. The idea of minimum contact types has also guided developments in the area of designing highly symmetric protein assemblies.

Within the natural hierarchy of symmetry types in three dimensions, from cyclic to dihedral to cubic, the latter category offers special features for investigation and exploitation. By their nature, cubic symmetries take the forms of the Platonic solids. As a consequence, such structures are closed, like a cage or shell, with defined interiors. As limiting cases or outliers on the symmetry spectrum, cubic protein assemblies offer intriguing case studies. As with other symmetric homo-oligomeric assemblies, the functional purpose for cubic symmetry is clear in some cases (i.e. viral capsids) and less clear in others. In addition to their potential biological implications, cubic assemblies offer unique advantages in bioengineering applications. Recent studies have begun to explore a range of novel uses, including interior encapsulation and exterior multivalent display. Below we survey the known cubic protein assemblies, both natural and designed, as interesting cases for biological insight and as starting points for diverse applications in medicine, biomaterials, and synthetic biology.

High-symmetry assemblies in the PDB

Known cubic structures

The Protein Data Bank (PDB) provides a rich source of information about oligomeric protein structures. However, discerning whether observed subunit arrangements in crystals represent biologically relevant assembly states

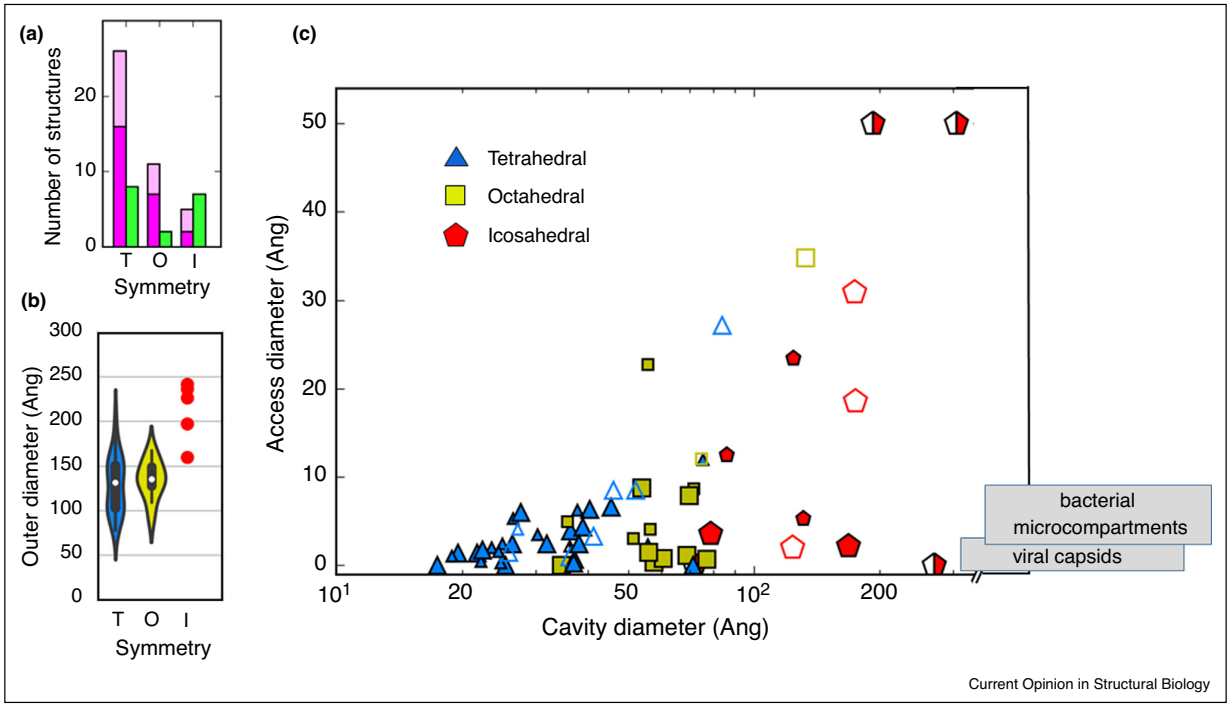
remains a long-standing and challenging problem [7–10]. A recent computational analysis by Dey *et al.* (available online at www.qsbio.org as the QSbio database) provides arguably the most robust assignments to date, with an estimated error rate of 15% for predicting true biological assembly forms [11]. The occurrence of cubic assemblies among natural proteins is rare enough that manual curation of the literature is feasible. Thus, we were able to augment computational inferences in the QSbio database, which are based largely on geometric analysis of interfaces and their preservation across structures of homologous proteins. Literature analyses showed the confidence assignments from QSbio to be generally conservative; several prospective cubic assemblies bearing 'low' or 'very low' confidence ratings could be validated by solution data. Excluding viral capsids, and counting only structurally unique representatives, the set of cubic protein assemblies – tetrahedral, octahedral, and icosahedral – includes 46 unique natural protein assemblies, twelve designed protein cages, and five other assemblies we designate as 'semi-synthetic' (see Supplemental Materials).

The known cubic assemblies span a wide range of sizes with distinct geometric properties, which may be relevant for function and critically important for prospective engineering applications. For each structure, Figure 1 shows the size of the interior cavity as well as the size of the largest opening or 'window' from the exterior. The former quantity bears on encapsulation capacity while the latter relates to porosity and access to the interior by other molecules. Suitable ranges for these parameters will vary depending on the application. Applications related to molecular containment and delivery will benefit from large cavity size and small window size. This has been articulated in recent studies (discussed below), where the goal was to encapsulate large nucleic acid molecules without permitting access to nucleases of varying sizes [12•,13••].

Connections to function and evolution

Relatively rare cases where high-symmetry assemblies have arisen in nature prompt questions about function. In many instances, a closed shape is clearly necessary for a protein's native function. Viruses require a capsid to

Figure 1



Geometric properties of known cubic and icosahedral protein assemblies. **(a)** Abundance of symmetry types T (tetrahedral), O (octahedral), and I (icosahedral) among natural structures (magenta) and designed structures (green) in the PDB. Natural assemblies whose coinfidence scores according to automated annotation were 'low' to 'very low' are colored light magenta [11]. **(b)** Distribution of sizes (outer diameter). **(c)** Sizes of interior cavities and window openings in known structures. Symmetry types are indicated by shape and color. Solid symbols indicate natural structures. Empty symbols indicate designed assemblies. Half-filled symbols indicate re-engineered forms of natural cage or shell structures. Smaller symbols indicate assemblies that had lower automatically assigned confidence levels but were judged reliable following manual curation. See Table S1 for additional details.

protect their encapsulated genetic material; ferritins form a shell to store iron; chaperonins encapsulate misfolded proteins to promote re-folding; and so on. However, numerous proteins form closed, highly symmetric structures without any obvious functional explanation [14,15]. It has been suggested that protein cage formation could have evolved in some enzymes to protect them against environmental stressors [16–18], such as in the case of ornithine carbamoyltransferase in thermophilic bacteria (PDB: 1A1S). There, the catalytically competent homotrimeric unit seen in other species assembles further into a 12-mer tetrahedron (i.e. a tetramer of trimers), with the increased subunit interactions presumed to contribute to stability [19]. A somewhat similar argument has been presented for the 2-hydroxypentadienoic acid hydratase enzyme from *Escherichia coli* (PDB: 2WQT). Its formation of a tightly packed 60-mer icosahedral cage from twelve copies of a pentameric subunit seems to be restricted to extreme conditions (e.g. low pH and high phosphate buffer concentrations). Most proteins of the same family carry out their function as simple C5 pentamers, and it remains unclear what environmental conditions in the cell, if any, might trigger cage formation in a biological setting [15].

Some cubic assemblies exploit geometric advantages that arise from spatial clustering of enzymatic active sites. This can improve the flux through pathways that involve multiple sequentially acting enzymes. The pyruvate dehydrogenase complex (PDC) is a well-studied case of a multi-enzyme system built around a highly symmetric core. Early work by Izard *et al.* demonstrated that the porous cage-like structure of the complex can take the form of either a 24-mer cube (PDB: 1DPB) or a 60-mer dodecahedron (PDB: 1B5S) depending on the species of origin [20]. Long, flexible poly-peptide tails on the exterior of the assembly recruit the other enzymes of the complex to perform their functions. The pyruvate substrate is shuttled via a ping-pong mechanism from one exterior enzyme to the octahedral or icosahedral core's active site and then back out to the third enzyme in the sequence to complete its conversion to acetyl-CoA [21,22]. A recent 3.1 Å structure obtained by cryo-electron microscopy (cryo-EM) demonstrates the first high-resolution structure of a mammalian PDC dodecahedral core and provides new insights into its properties in solution [23].

The integrity of cubic assemblies requires multiple distinct types of interactions between subunits, which naturally suggests evolutionary routes from simpler assembly forms. Previous studies have detailed likely pathways for the evolution of higher-order symmetric structures, with lower symmetry forms of the protein assemblies presumed to represent evolutionary intermediates [24]. Understanding how simpler symmetries can combine together to generate higher symmetry cubic structures has also provided a foundation for recent developments in designing novel protein cages and other geometrically ordered materials.

Designed assemblies

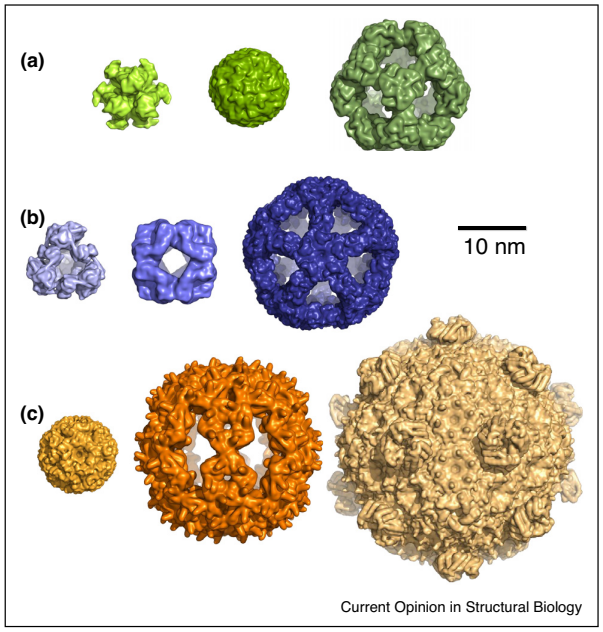
Highly symmetric natural protein assemblies have motivated design efforts in the creation of novel structures. Symmetry-based design principles and the earliest designed protein cage were described by Padilla *et al.* [25]. There, a tetrahedral cage was formed by the genetic fusion of simple oligomeric protein components (i.e. dimers and trimers) by continuous alpha helical linkers. Tetrahedral and octahedral designs were later demonstrated using this approach [26–28]. Larger collections of symmetric cages, generally less porous and more rigid than those created by the fusion approach, were created by King, Bale, Baker *et al.*, based on computational amino acid sequence design of new interfaces between oligomeric components [29–31,32••]. As described below, some of the cages generated by interface design are finding utility in novel applications.

Designing novel protein assemblies remains challenging, as engineered proteins often present difficulties in protein folding, expression, and proper assembly. To date, twelve designed cages have been validated in atomic detail by X-ray crystallography and have been deposited in the PDB. These include seven tetrahedral structures, two octahedral structures, and three icosahedral structures. The geometries of these successful designs are described in Figure 1. Seven additional icosahedral assemblies based on interface design have been shown to form symmetric particles by electron microscopy [31,32••]. Two additional cages – one tetrahedral and one octahedral – have been designed by more flexible fusions between a trimeric protein and various coiled coil segments. The designs were confirmed at the level of low resolution EM [33–35]. Designs in these latter groups are presumed to represent successful designs but have not been validated in atomic detail.

Different design strategies have led to geometric protein structures of somewhat less regular forms, including some that are smaller [36,37] and some that are larger [38] than the highly symmetric ones discussed here. Moreover, introducing metal-binding sites into simpler oligomeric building blocks has also fortuitously led to cage-like structures in other design studies [39].

Augmenting the designed novel structures above, a few recent studies have re-engineered proteins already known to form cages or shells, in order to create intriguing alternative assembly forms. We refer to these structures as 'semi-synthetic' (Figure 2c). In these studies, exploring sequence mutations or variations in subunit composition have led to novel structures whose detailed forms, which

Figure 2



Representative structures of cubic assemblies from various categories highlighting the diversity in symmetry, shape, size, and porosity. **(a)** Naturally evolved structures (left to right): PDB: 1A1S, ornithine carbamoyltransferase; PDB: 3BVE — *Helicobacter pylori* ferritin; PDB: 1B5S — dihydrolipoyl transacetylase. **(b)** Designed assemblies (left to right): PDB: 4QES — Designed tetrahedral cage, 3 + 2 symmetry type; PDB: 3VCD — Designed octahedral cage, threefold plus twofold interface symmetry type; PDB: 5IM4 — two-component icosahedral cage, 5 + 2 symmetry type. **(c)** Semi-synthetic structures obtained as variations on natural cages or shells (left to right): PDB: 5V74 — Icosahedral shell formed by a subset of bacterial microcompartment proteins from *Haliangium ochraceumi*; PDB: 5MQ3 — Expanded icosahedral cage based on lumazine synthetase; and PDB: 5HPN — Icosahedral cage from circularly permuted bacterial microcompartment protein PduA.

could not be divined in advance, were illuminated by X-ray crystallography and cryo-EM. Jorda *et al.* fortuitously obtained a $T = 1$ 60-mer icosahedral cage from a circular permutation of a bacterial microcompartment (BMC) shell protein [40•]. Sutter *et al.* demonstrated a larger $T = 9$ icosahedral shell built from a subset of BMC shell proteins present in a native bacterial microcompartment [41•]. In another remarkable set of structures, Sasaki *et al.* have shown the scalability of lumazine synthase assemblies, demonstrating that by adding negatively charged residues to the interior of the icosahedral cage, they could generate expanded cage forms of either 180 or 360 subunits [42•].

The growing suite of natural, designed, and semi-synthetic protein assemblies is providing new opportunities for diverse applications.

Recent applications and future directions

Protein cages or capsids have been explored for numerous applications. Viral capsids have been widely exploited in prior work, and a few non-viral proteins, including ferritin [43] and the thermophilic heat shock protein Hsp have been similarly investigated [44–47]. Those systems have been widely reviewed [48,49], including recently by Zhang *et al.* [50]. Among large but non-cubic assemblies, eukaryotic vaults [51,52] and chaperonins have also been explored [53,54]. Here we focus on some of the most recent studies and highlight a few newly emerging protein systems.

Interior encapsulation

The closed nature of cubic assemblies leads naturally to prospective applications in containment and delivery. Ferritin, a ubiquitous and well-characterized natural octahedral protein assembly (PDB: 3BVE, Figure 2), has unique physical and chemical properties that have made it a work-horse in previous applications in drug delivery, vaccine development, bioassays, and molecular imaging (reviewed by López-Sagaseta *et al.* and He and Marles-Wright) [55,56]. The rigid ferritin cage encloses an 8-nm interior cavity, is capable of reversibly disassembling under acidic conditions, and is biocompatible. Prior studies have shown ferritin's capacity to encapsulate and deliver anti-tumor drugs [57–60]. Recent studies have demonstrated ferritin's ability to encapsulate doxorubicin and cross the blood–brain barrier [61], and new work by Fan *et al.* showed that ferritin could selectively target glioma cells and release its drug payload to kill tumor cells *in vivo* [62•].

Additional cases of non-viral protein cages provide new prospective applications, with each system offering specific advantages in terms of geometry, amenability to mutagenesis and heterologous expression, and chemical and physical stability (as explored recently by Heinze *et al.* [63]). Taking a designed icosahedral cage as a starting point, Butterfield *et al.* showed that mutated versions of this cage (modified by introduction of an interior positive charge or RNA binding motifs) could encapsulate RNA molecules encoding the capsid shell proteins, mimicking the way a virus contains its genome within its own capsid [13••]. Subsequent rounds of optimization were employed, with roughly 9% of the cages successfully encapsulating mRNA. The interior RNA binding was non-sequence specific, but packaging was strongly correlated with expression levels such that 74% of encapsulated RNA was found to encode the capsid. Capsids were stable in blood for up to six hours.

Recent studies have similarly exploited the lumazine synthase system as a framework for evolving a nucleocapsid [64]. Mutants bearing a designed RNA-binding peptide tag fused to the lumenal side of the capsid achieved comparable levels of mRNA encapsulation,

with ~10% of capsids containing the full-length cage genome [65••]. In another case, Azuma *et al.* were able to engineer a distinct icosahedral lumazine synthase variant in order to encapsulate smaller mRNA sequences (up to 300 nucleotides) with tunable size-selectivity by using different lengths of poly-arginine tags on the lumenal surface of the cage [66]. Controlling the length of encapsulated nucleic acids provides an element of partial selectivity. In another study, a smaller (octahedral) designed cage was mutated to encapsulate short nucleic acid segments (approximately 21 nucleotides), again by the addition of an interior positive charge [12•]. Because of the large 35 Å windows into this cage, small guest nucleic acids could be added after protein purification in order to incorporate specific RNA molecules. However, these openings were large enough for nucleases, such as RNase A, to also gain access. The cages were taken up by mammalian cells via endocytosis, and cargo RNA was released inside the cytoplasm upon competitive binding of native tRNA to the cage. Loading the cages with siRNA led to successful gene knockdown of GFP and showed low levels of cytotoxicity. These novel cage systems are opening new avenues in drug delivery applications, with concomitant challenges related to nucleic acid sequence specificity, immunogenicity, and susceptibility to cellular nucleases and proteases.

Exterior display

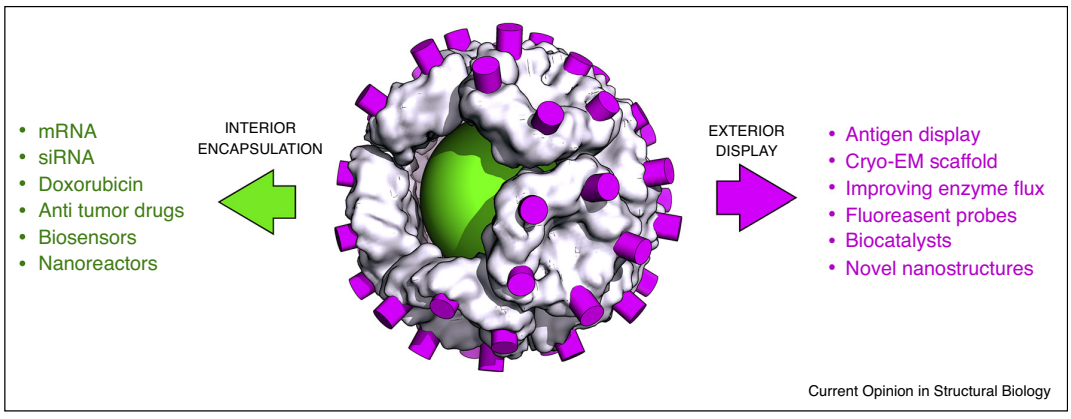
The high copy number of cubic assemblies offers prospects for polyvalent external attachments for varied purposes (Figure 3). Viral capsids, ferritin, and other diverse assemblies have been popular choices in pioneering efforts to functionalize the outer surfaces of protein cages, including for vaccine design and therapeutic and biomaterials applications [43,67–72]. New work by Dostalova *et al.* showed that decorating the surface of Dox-loaded apoferritin cages with mouse antibodies that target a prostate-specific membrane antigen leads to more reliably targeted payload release and lower off-target toxicity [73].

Protein cages of the designed variety have also started to find applications for exterior display. Votteler *et al.* adapted designed icosahedra to form extracellular vesicles *in vivo* via genetic fusion of a membrane-binding peptide motif on the cage's exterior. These cages were able to achieve cellular escape by recruiting endosomal sorting complexes required for transport (ESCRT) machinery, thus demonstrating a viable technology for the transfer of molecular cargo from one cell to another [74••].

Inspired by complexes that bring the active sites of sequentially-acting enzymes into spatial proximity to improve pathway flux [75], a recent study has demonstrated the use of a designed cage as a scaffold for the multi-copy display of enzymes that break down plant-derived cellulosic material to glucose. Attaching an endocellulase and an exocellulase to the exterior of a designed cage increased enzymatic activity by more than five-fold compared to free enzymes in solution (McConnell, Cannon *et al.*, unpublished). The power of high valency was also emphasized in a different kind of application in a recent study by Phippen *et al.*, in which genetically fusing antifreeze protein motifs to the exposed C-terminus of a designed cage resulted in a greater than 50-fold increase in the freezing point depression of water compared to free protein [76•].

A recent work has also demonstrated the powerful utility of cubic protein cages as rigid scaffolding for imaging small proteins by cryo-electron microscopy. Despite a recent technological boom in the cryo-EM field, it has been impossible to gain atomic-level structural detail

Figure 3



Current and future applications for symmetric cubic assemblies. Shells or cages with large interiors allow for encapsulation of other proteins, nucleic acids, metals, fluorophores, and drug molecules. The multivalent nature of cubic assemblies enables exterior high-copy display of antigens, cell-targeting molecule, bioactive motifs, and small proteins for cryo-EM imaging.

from proteins that are smaller than ~50 kDa, due to a low signal-to-noise ratio in this size regime. Designed Ankyrin Repeat Proteins (DARPins) can be obtained by *in vitro* selection methods to bind wide-ranging target proteins with high affinity. Exploiting their modularity, Liu *et al.* used DARPins as an adaptor component to overcome the cryo-EM size barrier by rigidly fusing DARPins to a tetrahedral cage. By design, the 13-nm particle symmetrically displays twelve copies of the target cargo protein bound to DARPins on the cage's exterior. An 18 kDa DARPin itself, not bound to a target, could be visualized by single particle cryo-EM at 3.5 Å–5 Å resolution [77••], and the structure of the first bound cargo protein, super-folder GFP (27 kDa), has now been imaged at a resolution of 3.8 Å by single particle cryo-EM (Liu [78]).

Future directions

Future studies will benefit from further examples of highly symmetric protein assemblies and their applications for novel purposes. Broader choices for protein frameworks will enable more-tailored design features to be realized. Relevant properties include: overall shape, interior accessibility, charge distribution, and positioning of chain termini for applications involving genetic fusion to other components. New strategies for selecting cages with favorable properties, such as robust assembly, from large libraries of mutants will be important for improving the pipeline for new designs. A recent foray by Orner *et al.* showed how successful cage formation could be connected to a fluorescent readout [79]. In newly published work, Marcandalli et al. used a designed protein cage to present a stabilized viral protein from respiratory syncytial virus (RSV) with favorable immunogenic properties [80]. Systematic analysis of successes and failures will enable more complex designs including active systems where specific phenomena (e.g. release) are triggered by particular cellular events or interactions, and where multiple signals can be integrated into systems that exhibit complex digital logic for modern synthetic biology applications.

Conflict of Interest statement

Nothing declared.

Acknowledgements

The authors thank Joshua Laniado, Justin Miller, and other members of the Yeates lab for ideas and helpful discussions. This work was supported by DOE BER Office of Science award # DE-FC02-02ER63421 and National Science Foundation grant CHE-1629214.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.sbi.2019.03.008.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Ahnert SE, Marsh JA, Hernández H, Robinson CV, Teichmann SA: **Principles of assembly reveal a periodic table of protein complexes**. *Science* 2015, **350**.

2. André I, Strauss CEM, Kaplan DB, Bradley P, Baker D: **Emergence of symmetry in homooligomeric biological assemblies**. *Proc Natl Acad Sci U S A* 2008, **105**:16148-16152.

3. Marsh JA, Teichmann SA: **Structure, dynamics, assembly, and evolution of protein complexes**. *Annu Rev Biochem* 2015, **84**:551-575.

4. Monod J: **On symmetry and function in biological systems**. *Nobel Symposium. Symmetry and Function of Biological System at the Macromolecular Level, 11th, Stockholm*. Wiley; 1968:15-27.

5. Goodsell DS, Olson AJ: **Structural symmetry and protein function**. *Annu Rev Biophys Biomol Struct* 2000, **29**:105-153.

6. Crick FHC, Watson JD: **Structure of small viruses**. *Nature* 1956, **177**:473-475.

7. Janin J, Rodier F: **Protein–protein interaction at crystal contacts**. *Proteins* 1995, **23**:580-587.

8. Henrick K, Thornton JM: **PQS: a protein quaternary structure file server**. *Trends Biochem Sci* 1998, **23**:358-361.

9. Ponstingl H, Kabir T, Thornton JM: **Automatic inference of protein quaternary structure from crystals**. *J Appl Crystallogr* 2003, **36**:1116-1122.

10. Krissinel E, Henrick K: **Inference of macromolecular assemblies from crystalline state**. *J Mol Biol* 2007, **372**:774-797.

11. Dey S, Ritchie DW, Levy ED: **PDB-wide identification of biological assemblies from conserved quaternary structure geometry**. *Nat Methods* 2018, **15**:67-72.

12. • Edwardson TGW, Mori T, Hilvert D: **Rational Engineering of a designed protein cage for siRNA delivery**. *J Am Chem Soc* 2018, **140**:10439-10442.
The authors successfully load a designed octahedral protein cage with siRNA that can enter cells via endocytosis and release their cargo in the cytosol. Successful gene knockdown in mammalian cells expressing GFP led to a 70% decrease in GFP fluorescence with lower cytotoxicity compared to other delivery systems.

13. •• Butterfield GL, Lajoie MJ, Gustafson HH, Sellers DL, Nattermann U, Ellis D, Bale JB, Ke S, Lenz GH, Yehdego A *et al.*: **Evolution of a designed protein assembly encapsulating its own RNA genome**. *Nature* 2017, **552**:415-420.
The authors report the use of a designed icosahedral protein cage to encapsulate its own RNA genome via several rounds of directed evolution, with about 1 in 11 cages containing the correct RNA sequence. This synthetic nucleocapsid was demonstrated to be stable in blood and circulate *in vivo* in mice for several hours.

14. McNamara DE, Cascio D, Jorda J, Bustos C, Wang T-C, Rasche ME, Yeates TO, Bobik TA: **Structure of dihydromethanopterin reductase, a cubic protein cage for redox transfer**. *J Biol Chem* 2014, **289**:8852-8864.

15. Montgomery MG, Coker AR, Taylor IA, Wood SP: **Assembly of a 20-nm protein cage by *Escherichia coli* 2-hydroxypentadienoic acid hydratase**. *J Mol Biol* 2010, **396**:1379-1391.

16. Ha N-C, Oh S-T, Sung JY, Cha KA, Lee MH, Oh B-H: **Supramolecular assembly and acid resistance of *Helicobacter pylori* urease**. *Nat Struct Mol Biol* 2001, **8**:505-509.

17. D'Abusco AS, Casadio R, Tasco G, Giangiacomo L, Giartosio A, Calamia V, Di Marco S, Chiaraluce R, Consalvi V, Scandurra R *et al.*: **Oligomerization of *Sulfolobus solfataricus* signature amidase is promoted by acidic pH and high temperature**. *Archaea* 2005, **1**:411-423.

18. Carter EL, Tronrud DE, Taber SR, Karplus PA, Hausinger RP: **Iron-containing urease in a pathogenic bacterium**. *Proc Natl Acad Sci U S A* 2011, **108**:13095-13099.

19. Villeret V, Clantin B, Tricot C, Legrain C, Roovers M, Stalon V, Glansdorff N, Beeumen JV: **The crystal structure of *Pyrococcus furiosus* ornithine carbamoyltransferase reveals a key role for oligomerization in enzyme stability at extremely high temperatures**. *Proc Natl Acad Sci U S A* 1998, **95**:2801-2806.

20. Izard T, Aevarsson A, Allen M, Westphal A, Perham R, de Kok A, Hol WGJ: **Principles of quasi-equivalence and Euclidean geometry govern the assembly of cubic and dodecahedral cores of pyruvate dehydrogenase complexes**. *Proc Natl Acad Sci U S A* 1999, **96**:1240-1245.

21. Tsai CS, Burgett MW, Reed LJ: **α-Keto acid dehydrogenase complexes XX. A kinetic study of the pyruvate dehydrogenase complex from bovine kidney**. *J Biol Chem* 1973, **248**:8348-8352.

22. Zhou ZH, McCarthy DB, O'Connor CM, Reed LJ, Stoops JK: **The remarkable structural and functional organization of the eukaryotic pyruvate dehydrogenase complexes**. *Proc Natl Acad Sci U S A* 2001, **98**:14802-14807.

23. Jiang J, Baiesc FL, Hiromasa Y, Yu X, Hui WH, Dai X, Roche TE, Zhou ZH: **Atomic structure of the E2 inner core of human pyruvate dehydrogenase complex**. *Biochemistry* 2018, **57**:2325-2334.

24. Levy ED, Erba EB, Robinson CV, Teichmann SA: **Assembly reflects evolution of protein complexes**. *Nature* 2008, **453**:1262-1265.

25. Padilla JE, Colovos C, Yeates TO: **Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments**. *Proc Natl Acad Sci U S A* 2001, **98**:2217-2221.

26. Lai YT, Cascio D, Yeates TO: **Structure of a 16-nm cage designed by using protein oligomers**. *Science* 2012, **336**:1129.

27. Lai YT, Tsai KL, Sawaya MR, Asturias FJ, Yeates TO: **Structure and flexibility of nanoscale protein cages designed by symmetric self-assembly**. *J Am Chem Soc* 2013, **135**:7738-7743.

28. Lai YT, Reading E, Hura GL, Tsai KL, Laganowsky A, Asturias FJ, Tainer JA, Robinson CV, Yeates TO: **Structure of a designed protein cage that self-assembles into a highly porous cube**. *Nat Chem* 2014, **6**:1065-1071.

29. King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, Andre I, Gonen T, Yeates TO, Baker D: **Computational design of self-assembling protein nanomaterials with atomic level accuracy**. *Science* 2012, **336**:1171-1174.

30. King NP, Bale JB, Sheffler W, McNamara DE, Gonen S, Gonen T, Yeates TO, Baker D: **Accurate design of co-assembling multi-component protein nanomaterials**. *Nature* 2014, **510**:103-108.

31. Hsia Y, Bale JB, Gonen S, Shi D, Sheffler W, Fong KK, Nattermann U, Xu C, Huang P-S, Ravichandran R *et al.*: **Design of a hyperstable 60-subunit protein icosahedron**. *Nature* 2016, **535**:136-139.

32. •• Bale JB, Gonen S, Liu Y, Sheffler W, Ellis D, Thomas C, Cascio D, Yeates TO, Gonen T, King NP *et al.*: **Accurate design of megadalton-scale two-component icosahedral protein complexes**. *Science* 2016, **353**:389-395.

This study reports the computational design and characterization of large, 120-subunit icosahedral protein cages. Three crystal structures are reported, and six more designs are verified by electron microscopy. These represent the largest atomically precise designed protein assemblies to date.

33. Cristie-David AS, Koldewey P, Meinen BA, Bardwell JCA, Marsh ENG: **Elaborating a coiled-coil-assembled octahedral protein cage with additional protein domains**. *Protein Sci* 2018, **27**:1893-1900.

34. Sciore A, Su M, Koldewey P, Eschweiler JD, Diffley KA, Linhares BM, Ruotolo BT, Bardwell JCA, Skiniotis G, Marsh ENG: **Flexible, symmetry-directed approach to assembling protein cages**. *Proc Natl Acad Sci U S A* 2016, **113**:8681-8686.

35. Badieyan S, Sciore A, Eschweiler JD, Koldewey P, Cristie-David AS, Ruotolo BT, Bardwell JCA, Su M, Marsh ENG: **Symmetry-directed self-assembly of a tetrahedral protein cage mediated by de novo-designed coiled coils**. *ChemBioChem* 2017, **18**:1888-1892.

36. Lapenta F, Aupič J, Strmšek , Jerala R: **Coiled coil protein origami: from modular design principles towards biotechnological applications**. *Chem Soc Rev* 2018, **47**:3530-3542.

37. Gradišar H, Božič S, Doles T, Vengust D, Hafner-Bratkovič I, Mertelj A, Webb B, ali A, Klavžar S, Jerala R: **Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments**. *Nat Chem Biol* 2013, **9**:362-366.

38. Fletcher JM, Harniman RL, Barnes FR, Boyle AL, Collins A, Mantell J, Sharp TH, Antognozzi M, Booth PJ, Linden N *et al.*: **Self-assembling cages from coiled-coil peptide modules**. *Science* 2013, **340**:595-599.

39. Ni TW, Tezcan FA: **Structural characterization of a microperoxidase inside a metal- directed protein cage**. *Angew Chem Int Ed* 2010, **49**:7014-7018.

40. • Jorda J, Leibly DJ, Thompson MC, Yeates TO: **Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein**. *Chem Commun* 2016, **52**:5041-5044.

The authors report the serendipitous formation of a minimal $T = 1$ icosahedron from a circular permutation of a shell protein from the Pdu bacterial microcompartment, opening up new possibilities for the use of microcompartment proteins in synthetic biology and nanotechnological applications.

41. • Sutter M, Greber B, Aussignargues C, Kerfeld CA: **Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell**. *Science* 2017, **356**:1293-1297.

This study reports the EM and crystal structure of a $T = 9$ icosahedron made up of a subset of natural microcompartment shell proteins. Notably, the concave side of the shell proteins faces to the outside of the cage in this structure.

42. • Sasaki E, Böhringer D, Van De Waterbeemd M, Leibundgut M, Zschoche R, Heck AJR, Ban N, Hilvert D: **Structure and assembly of scalable porous protein cages**. *Nat Commun* 2017, **8**:1-10.

The authors demonstrate the scalability of lumazine synthase cage assemblies. By mutating amino acids on the lumenal surface of the protein, remarkably large assemblies of 180 and 360 subunits were observed, the latter case representing a highly unusual subunit arrangement.

43. Wang Z, Gao H, Zhang Y, Liu G, Niu G, Chen X: **Functional ferritin nanoparticles for biomedical applications**. *Front Chem Sci Eng* 2017, **11**:633-646.

44. Murata M, Narahara S, Umezaki K, Toita R, Tabata S, Piao JS, Abe K, Kang JH, Ohuchida K, Cui L *et al.*: **Liver cell specific targeting by the preS1 domain of hepatitis B virus surface antigen displayed on protein nanocages**. *Int J Nanomedicine* 2012, **7**:4353-4362.

45. Toita R, Murata M, Abe K, Narahara S, Piao JS, Kang JH, Hashizume M: **A nanocarrier based on a genetically engineered protein cage to deliver doxorubicin to human hepatocellular carcinoma cells**. *Chem Commun (Camb)* 2013, **49**:7442-7444.

46. Uchida M, Kosuge H, Terashima M, Willits DA, Liepold LO, Young MJ, McConnell MV, Douglas T: **Protein cage nanoparticles bearing the LyP-1 peptide for enhanced imaging of macrophage-rich vascular lesions**. *ACS Nano* 2011, **5**:2493-2502.

47. Abedin MJ, Liepold L, Suci P, Young M, Douglas T: **Synthesis of a cross-linked branched polymer network in the interior of a protein cage**. *J Am Chem Soc* 2009, **131**:4346-4354.

48. Flenniken ML, Uchida M, Liepold LO, Kang S, Young MJ, Douglas T: **A library of protein cage architectures as nanomaterials**. *Viruses Nanotechnol* 2009, **327**:71-93 http://dx.doi.org/10.1007/978-3-540-69379-6_4.

49. Czapar AE, Steinmetz NF: **Plant viruses and bacteriophages for drug delivery in medicine and biotechnology**. *Curr Opin Chem Biol* 2017, **38**:108-116.

50. Zhang Y, Ardejani MS, Orner BP: **Design and applications of protein-cage-based nanomaterials**. *Chem Asian J* 2016, **11**:2814-2828.

51. Yu K, Yau YH, Sinha A, Tan T, Kickhoefer VA, Rome LH, Lee H, Shochat SG, Lim S: **Modulation of the vault protein-protein interaction for tuning of molecular release**. *Sci Rep* 2017, **7**:14816.

52. Rome LH, Kickhoefer VA: **Development of the vault particle as a platform technology**. *ACS Nano* 2013, **7**:889-902.

53. Yuan Y, Du C, Sun C, Zhu J, Wu S, Zhang Y, Ji T, Lei J, Yang Y, Gao N *et al.*: **Chaperonin-GroEL as a smart hydrophobic drug delivery and tumor targeting molecular machine for tumor therapy**. *Nano Lett* 2018, **18**:921-928.

54. Hoersch D, Roh S-H, Chiu W, Kortemme T: **Reprogramming an ATP-driven protein machine into a light-gated nanocage**. *Nat Nanotechnol* 2013, **8**:928-932.

55. López-Sagaseta J, Malito E, Rappuoli R, Bottomley MJ: **Self-assembling protein nanoparticles in the design of vaccines**. *Comput Struct Biotec J* 2016, **14**:58-68.

56. He D, Marles-Wright J: **Ferritin family proteins and their use in bionanotechnology**. *New Biotechnol* 2015, **32**:651-657.

57. Falvo E, Tremante E, Arcovito A, Papi M, Elad N, Boffi A, Morea V, Conti G, Toffoli G, Fracasso G *et al.*: **Improved doxorubicin encapsulation and pharmacokinetics of ferritin-fusion protein nanocarriers bearing proline, serine, and alanine elements**. *Biomacromolecules* 2016, **17**:514-522.

58. Zhen Z, Tang W, Chen H, Lin X, Todd T, Wang G, Cowger T, Chen X, Xie J: **RGD-modified apoferritin nanoparticles for efficient drug delivery to tumors**. *ACS Nano* 2013, **7**:4830-4837.

59. Zhen Z, Tang W, Guo C, Chen H, Lin X, Liu G, Fei B, Chen X, Xu B, Xie J: **Ferritin nanocages to encapsulate and deliver photosensitizers for efficient photodynamic therapy against cancer**. *ACS Nano* 2013, **7**.

60. Uchida M, Flenniken ML, Allen M, Willits DA, Crowley BE, Brumfield S, Willis AF, Jackiw L, Jutila M, Young MJ *et al.*: **Targeting of cancer cells with ferrimagnetic ferritin cage nanoparticles**. *J Am Chem Soc* 2006, **128**:16626-16633.

61. Chen Z, Zhai M, Xie X, Zhang Y, Ma S, Li Z, Yu F, Zhao B, Zhang M, Yang Y *et al.*: **Apoferritin nanocage for brain targeted doxorubicin delivery**. *Mol Pharm* 2017, **14**:3087-3097.

62. • Fan K, Jia X, Zhou M, Wang K, Conde J, He J, Tian J, Yan X: **Ferritin nanocarrier traverses the blood brain barrier and kills glioma**. *ACS Nano* 2018, **12**:4105-4115.

The authors use a ferritin cage loaded with doxorubicin to deliver the anti-tumor drug across the blood brain barrier and kill glioma cells in mice. They report that the ferritin cages accumulate in the lysosomal compartment, selectively killing tumor cells and not the surrounding healthy brain tissue.

63. Heinze K, Sasaki E, King NP, Baker D, Hilvert D, Wuite GJL, Roos WH: **Protein nanocontainers from nonviral origin: testing the mechanics of artificial and natural protein cages by AFM**. *J Phys Chem B* 2016, **120**:5945-5952.

64. Wörsdörfer B, Woycechowsky KJ, Hilvert D: **Directed evolution of a protein container**. *Science* 2011, **331**:589-592.

65. •• Terasaka N, Azuma Y, Hilvert D: **Laboratory evolution of virus-like nucleocapsids from nonviral protein cages**. *Proc Natl Acad Sci U S A* 2018, **115**:5432-5437.

The authors redesigned a lumazine synthase icosahedral cage to contain an RNA-binding peptide sequence on its lumenal surface, and after multiple rounds of directed evolution achieved a cage with the ability to encapsulate its own mRNA genome with 10% success rate.

66. Azuma Y, Edwardson TGW, Terasaka N, Hilvert D: **Modular protein cages for size-selective RNA packaging in vivo**. *J Am Chem Soc* 2018, **140**:566-569.

67. Karch CP, Burkhard P: **Vaccine technologies: from whole organisms to rationally designed protein assemblies**. *Biochem Pharmacol* 2016, **120**:1-14.

68. Kanekiyo M, Bu W, Joyce MG, Meng G, Whittle JRR, Baxa U, Yamamoto T, Narpala S, Todd J-P, Rao SS *et al.*: **Rational design of an Epstein-Barr virus vaccine targeting the receptor-binding site**. *Cell* 2015, **162**:1090-1100.

69. Kanekiyo M, Wei C-J, Yassine HM, McTamney PM, Boyington JC, Whittle JRR, Rao SS, Kong W-P, Wang L, Nabel GJ: **Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies**. *Nature* 2013, **499**:102-106.

70. Negahdaripour M, Golkar N, Hajighahramani N, Kianpour S, Nezafat N, Ghasemi Y: **Harnessing self-assembled peptide nanoparticles in epitope vaccine design**. *Biotechnol Adv* 2017, **35**:575-596.

71. Sliepen K, Ozorowski G, Burger JA, van Montfort T, Stunnenberg M, LaBranche C, Montefiori DC, Moore JP, Ward AB, Sanders RW: **Presenting native-like HIV-1 envelope trimers on ferritin nanoparticles improves their immunogenicity**. *Retrovirology* 2015, **12**:82.

72. Zhao Q, Li S, Yu H, Xia N, Modis Y: **Virus-like particle-based human vaccines: quality assessment based on structural and functional properties**. *Trends Biotechnol* 2013, **31**:654-663.

73. Dostalova S, Polanska H, Svobodova M, Balvan J, Krystofova O, Haddad Y, Krizkova S, Masarik M, Eckschlager T, Stiborova M *et al.*: **Prostate-specific membrane antigen-targeted site-directed antibody-conjugated apoferritin nanovehicle favorably influences in vivo side effects of doxorubicin**. *Sci Rep* 2018, **8**:1-13.

74. •• Votteler J, Ogohara C, Yi S, Hsia Y, Nattermann U, Belnap DM, King NP, Sundquist WI: **Designed proteins induce the formation of nanocage-containing extracellular vesicles**. *Nature* 2016, **540**:292-295.

By displaying a membrane-binding peptide motif on the exterior of a designed icosahedral cage, the authors report the formation of cage-containing extracellular vesicles. These vesicle-forming cages could be used to transport molecular cargo between cells.

75. Moon TS, Dueber JE, Shiue E, Prather KL: **Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli***. *Metab Eng* 2010, **12**:298-305.

76. • Phippen SW, Stevens CA, Vance TDR, King NP, Baker D, Davies PL: **Multivalent display of antifreeze proteins by fusion to self-assembling protein cages enhances ice-binding activities**. *Biochemistry* 2016, **55**:6811-6820.

Using a designed tetrahedral protein cage as a scaffold, the authors genetically fused antifreeze proteins to the exterior for multivalent display. These nanoparticles were able to achieve a greater than 50-fold increase in the freezing point depression of water compared to free enzyme, with a similar increase in ice-recrystallization inhibition.

77. •• Liu Y, Gonen S, Gonen T, Yeates TO: **Near-atomic cryo-EM imaging of a small protein displayed on a designed scaffolding system**. *Proc Natl Acad Sci U S A* 2018, **115**:3362-3367.

This study reports the first use of a designed protein cage as a scaffold for the structural determination of sub-50 kDa proteins by displaying them on the surface of a tetrahedron via the fusion of a DARPin motif that can be selected to tightly bind various target proteins. The authors report structural detail of the 18 kDa DARPin at 3.5–5 Å resolution.

78. Liu Huynh D, Yeates TO: **A 3.8 Å resolution Cryo-EM structure of a small protein bound to a modular imaging scaffold**. *Nat Commun* 2019. (in press).

79. Cornell TA, Fu J, Newland SH, Orner BP: **Detection of specific protein-protein interactions in nanocages by engineering bipartite FlAsH binding sites**. *J Am Chem Soc* 2013, **135**:16618-16624.

80. Marcandalli J *et al.*: **Induction of potent neutralizing antibody responses by a designed protein nanoparticle vaccine for respiratory syncytial virus**. *Cell* 2019, **176**:1420-1431.

Chapter 3: Recent Structural Insights in Bacterial Microcompartment Shells

Recent Structural Insights into Bacterial Microcompartment Shells

Jessica M. Ochoa[1] and Todd O. Yeates[1,2,3]

[1]UCLA Molecular Biology Institute

[2]UCLA DOE Institute for Genomics and Proteomics

[3]UCLA Department of Chemistry and Biochemistry

Abstract

Bacterial microcompartments are organelle-like structures that enhance a variety of metabolic functions in diverse bacteria. Composed entirely of proteins, thousands of homologous hexameric shell proteins tesselate to form facets while pentameric proteins form the vertices of a polyhedral shell that encapsulates various enzymes, substrates and cofactors. Recent structural data have highlighted nuanced variations in the sequence and topology of microcompartment shell proteins, emphasizing how variation and specialization enable the construction of complex molecular machines. Recent studies engineering synthetic miniaturized microcompartment shells provide additional frameworks for dissecting principals of microcompartment structure and assembly. This review updates our current understanding of bacterial microcompartment shell proteins, providing new insights and highlighting outstanding questions.

Introduction

Bacterial Microcompartments (MCPs or alternatively BMCs) are a class of supramolecular structures found in approximately 20% of bacteria. Ranging in size from roughly 100 - 400 nm in diameter, MCPs encapsulate and optimize a myriad of metabolic pathways by concentrating together enzymes and substrates to accelerate catalysis and to prevent the escape of toxic or volatile intermediates [1– 3]. Unlike membrane-bound eukaryotic organelles, MCPs are composed entirely of proteins. So-called BMC proteins (based on Pfam PF00936) form hexameric building blocks shaped like hexagonal disks (Figure 2.1). These tessellate side by side to form flat extended facets [4–8] with lateral associations driven by highly conserved perimeter residues

[7,9–11]. In addition, and distinct from BMC proteins, pentameric BMV proteins (Pfam PF03319) form the vertices of the polyhedral shell, which is in some cases roughly icosahedral (Figure 1) [12–14]. The heterogenous shells of MCPs are formed from two to seven BMC paralogs that are often expressed within a single operon [1–3]. Narrow pores located at the center of BMC proteins, ranging from 5 - 10 Å in diameter, provide channels for the diffusion of specific substrates and cofactors. The external protein shell remains the hallmark of all MCPs.

Despite their structural similarity, MCPs are functionally diverse, carrying out various multi-step metabolic reactions in different bacteria. MCPs can be broadly classified into two major categories: carboxysomes and metabolosomes. The founding MCP type, carboxysomes, utilize bicarbonate (HCO3) and ribulose- 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) in order to enhance CO2 fixation [15–17]. By contrast, metabolosomes metabolize a variety of substrates. Two types of B12-dependent metabolosomes have been extensively studied and shown to degrade 1,2-propanediol (Pdu MCP) and ethanolamine (EuT MCP) [1,3,16]. Another recently discovered class of MCPs, glycyl radical enzyme metabolosomes (GRMs), perform glycyl radical chemistry using distinct signature enzymes that define an assortment of subclasses, GRMs 1-5 [18–22]. Some GRMs metabolize choline (Cut MCPs, Type I and Type II), some consume 1,2-propanediol in a B12-independent manner (Grp MCP) and others process fucose and rhamnose [18,23–26]. Additionally, recent discoveries have begun to characterize MCPs that use S-1-amino-2-propanol-kinase to process aminoacetone (AAUMs or formerly RMMs), MCPs that degrade ethanol (Etu MCPs) and a new class of MCPs that are predicted to degrade xanthine [20,21,27–29].

This review focuses on current structural data, highlighting and updating our understanding of the roles of bacterial microcompartment shell proteins. We draw attention to the unusual variations in sequence and topology in BMC shell proteins and emphasize structural polymorphisms in certain subsets, which are likely to relate to functional specialization. Finally,

we investigate the implications of recently characterized mini MCP shells, addressing strengths in identifying MCP assembly principals and overall shell topology.

Sequence and topological variations and their implications

To date, the structures of some 110 MCP shell proteins have been deposited in the protein data bank (PDB) [30]. A specialized database focusing on MCPs has recently been established to facilitate the analysis and study of their structures [31]. Here we update the tertiary structure variations discovered so far.

Though structurally similar, BMC shell proteins exhibit topological differences of various types (Figure 2a). The canonical BMC protein is comprised of a roughly 100 amino acid domain (Pf00936), and is referred to as BMC-H for its hexameric assembly (Figure 2.2b). The hexagonal disks formed by BMC proteins have distinctly shaped top and bottom faces, with one relatively flat face and the other bearing a central depression that creates a concave surface. Looking down the flat face and following the sequence from the N to C-terminus, the secondary structure elements of BMC-H proteins are arranged in a roughly clockwise fashion (Figure 2.2b). A unique subset of hexameric BMC domain-containing shell proteins, Permuted BMCs, have been discovered to have cyclically permuted sequences and structures. While they possess a similar overall tertiary structure, the circular permutation results in differently poised N and C-termini relative to their BMC-H counterparts [32–34]. In cases that have been structurally characterized (including Permuted BMCs from Eut, Pdu and Cut MCP types), a novel extension at the N-terminus forms a right-handed 6-stranded beta-barrel (with one strand from each subunit) protruding from the otherwise flat face (Figure 2.2b). Another unique variation of BMC shell proteins has arisen from gene-duplication events, thereby producing tandem domain structures. So- called BMC-T proteins, comprising two BMC domains, oligomerize to form trimeric pseudohexamers (Figure 2b) whose overall architectures closely resemble a canonical hexameric BMC disk [8,35–37].

Three-dimensional all-against-all comparisons between the known BMC structures reveal further types of variation, particularly among the BMC-T proteins (Figure 2.2). These relate to surprising differences in the way the sequential tandem domains are arranged, and whether the individual domains are permuted. Remarkably, different BMC-T proteins present sequentially connected BMC domains arranged in either a clockwise or counterclockwise fashion in the context of the trimeric (pseudohexameric) disk. These varied forms can be accommodated with a more finely articulated naming convention; BMC-T(+), Permuted BMC-T(+) and Permuted BMC-T(-). Here the superscript conveys the clockwise (+) or counterclockwise (-) ordering of domains when viewed from the flat face (Figure 2b). Interestingly, examples of the BMC-T(+) type include cases where the central pore presents three symmetry-related cysteine residues (one from each protein chain) for coordinating an 4Fe-4S cluster [38–40]. A general theme is that the evolution of trimeric BMC assemblies appears to have allowed for greater versatility at the pore because of lower symmetry. Indeed, examples of both Permuted BMC-T(+) and Permuted BMC- T(-) proteins have been shown to form trimers in which the central pore can apparently convert between open and closed forms, with important implications for the transport of larger substrates or cofactors [33,36,37,41–45].

The central pores of BMC shell proteins provide routes for the diffusion of molecules across the MCP shell. Mutagenesis experiments suggest that the narrow pores in BMC-H hexamers are the primary routes of substrate influx [46–48]. Different MCP types operate on and thus transport different substrates, suggesting that sequence and structural variations in the pores of BMC proteins are likely important for diverse metabolic functions. Electrostatic properties of BMC pores have been analyzed, with particular implications for their roles in MCPs with charged substrates (e.g. the carboxysome [bicarbonate], Eut [ethanolamine], and Aaum [aminoacetone]) [4,5,14,42,42]. Several recent molecular dynamics (MD) and flux modeling studies have begun to examine the atomic details and mathematical aspects of pore transport. Important questions

concern the degree to which pores in BMC-H proteins are selective for their cognate metabolic substrate. Optimal metabolic function would presumably occur with a combination of facile substrate influx and restricted metabolic intermediate efflux. MD studies on the PduA (BMC-H) protein suggested a modest level of selection in this regard, with a preference for its propanediol substrate that is 3 to 10 times greater than its propionaldehyde intermediate [49]. Similar MD studies on carboxysome shell proteins have also reported a range of selectivity for its substrate, with values in one case as high as 1000 times greater than its corresponding intermediate [47]. Interestingly, transport and metabolic flux modeling on both the carboxysome and the Pdu MCP have emphasized that high selectivity might not be critical for function if internal consumption of the intermediate is sufficiently rapid [50,51]. Nonetheless, the pores of BMC-H proteins present useful targets for modulating MCP function, including by mutagenesis to occlude pores or to insert non-native 4Fe-4S clusters [40,46]. Additional structural studies on more remote BMC homologs could shed further light on transport mechanisms. Based on sequence alignment, several BMC shell proteins from a recently-proposed xanthine MCP [29] appear to have three to four residue insertions near the loop region, opening the possibility for identifying novel pore features and functionalities.

BMC-T proteins present additional puzzles for transport. As noted above, in some cases they have been found to harbor larger open pores. The potential presence of larger pores in the shell presents a dilemma, as retaining key metabolic intermediates is essential for proper MCP function. Different ideas have been put forward on the subject. In some cases it appears that the large BMC-T pores are regulated and could be occluded by allosteric binding events, e.g. by substrates when the MCP is active [39,41,42,52]. As described below, alternative mechanisms of opening and closing have been proposed in other cases [33,36,37,41–44,52].

Structural polymorphisms

MCP shell proteins display a surprising degree of flexibility (Figure 2.3). This is particularly true of Permuted BMC proteins. An early study found that the EutS shell protein crystallized in two forms: 1) a canonical flat disk and 2) a hexamer with a twisting or bending deformation down the two-fold axis of symmetry (Fig. 2.3a) [33]. Recent structural characterization of another Permuted BMC homolog, CutR from a Choline Utilization Type II MCP, showed other forms of flexibility including the formation of flat disks and screw-type helical assemblies of varying pitch (Figure3a) [34]. In another instance, a synthetic Permuted BMC – a version of PduA (BMC-H) that was engineered to introduce an artificial circular permutation – rearranged to form a cyclic homopentamer, despite retaining the BMC fold (Figure 3a) [53]. Interestingly, such structural polymorphism does not appear unique to Permuted BMCs. BMV shell proteins are understood to serve as the pentameric component required for (Gaussian) shell curvature and closure [12,13,54], yet the EutN protein (from the BMV family, PF03319) was found in two separate crystallographic experiments to be capable of forming cyclic homohexamers (PDB ID 2HD3 and 2Z9H) [55].

BMC-Ts also appear capable of alternative quaternary conformations (Figure 2.3b). Specifically, all structurally characterized members of the Permuted BMC-T(+) subset have been observed to assemble as stacked disks (a dimer of trimers), creating a large central cavity accessible by pores on opposite ends [37,42–45]. While their biological relevance has yet to be confirmed, their recurrence in multiple studies suggests their potential importance. Stacked disks are observed in crystal structures of recombinantly expressed and purified Permuted BMC-T(+) type proteins from alpha-carboxysomes (CsoS1D), beta-carboxysomes (CcmP), aminoacetone utilization MCPs and other MCPs of unknown function [37,42,43,52]. Double disks have also been observed by crystallography and cryo-EM in recombinantly purified mini MCP shells [44,56], further described below. One intriguing hypothesis is that they could serve as a gated airlock

system for transport [37]. Variability in the open and closed states of the pores of stacked disks have been noted in numerous studies. Some stacked disks have two open pores (PDB ID 4HT7), some have one open and one closed pore (PDB ID 3F56, 3FCH, 4HT5, 5LSR and 5V75) and some have two closed pores (PDB ID 3NWG, 5L39, 5LT5, 5V76 and 5SUH) [37,42– 45]. The presence of stacked disks with two open pores would counter an airlock mechanism, but crystallographic observations call for cautious interpretation on this issue, owing both to conditions of the crystalline state and challenges in adequately capturing important dynamic behavior.

Large recombinant assembly forms

Attempts to assemble larger protein species, including work to mix different shell proteins, date to the first BMC protein structural studies [4]. Subsequent efforts to develop experimental procedures and suitable combinations of shell proteins have led to remarkable successes in purifying and characterizing what can be described as miniaturized synthetic shells (Figure 4a). Several examples have been obtained, built from either a single component or multiple shell components and ranging in size from 130 Å to 400 Å in diameter [19,44,53,56,57]. These miniaturized synthetic MCP shells have helped to support and more finely articulate assembly principles that were formulated from studies on individual shell components. Of particular note, these synthetic miniaturized MCPs have confirmed the roles of BMV proteins as polyhedral vertices and the role of lateral associations between hexameric units through conserved interactions at their perimeters.

These miniaturized structures have also led to surprises. The first observed mini shell was obtained serendipitously from an engineering experiment wherein a synthetically permuted version of an otherwise ordinary BMC-H protein, PduA, formed a 130 Å dodecahedron from twelve cyclic homopentamers [53]. More deliberate assembly studies based on mixtures of various BMC and BMV proteins have produced a range of structures. A 6.5 MDa mini shell was constructed from one BMC-H, one BMV and three BMC- T proteins from an MCP of unknown

function. This 400 Å diameter, icosahedral shell (triangulation number T=9) highlighted the dynamic ability of different BMC-domain containing proteins to occupy different positions in the icosahedral shell [44,56]. Studies on constructing mini shells from beta- carboxysome shell proteins resulted in a variety of structures including 210 Å T=3, 245 Å T=4 and a 310 Å prolate T=4, Q=6 icosahedral mini shells. A broad diversity in shape, size and morphology was observed despite using only one BMC-H and one BMV protein [57]. The most recent study led to the structural characterization of a 250 Å T=4 shell constructed from GRM2 proteins. This work utilized three BMC-Hs, one BMV and numerous enzymatic proteins in varying combinations. In addition to the T=4 shell that was characterized in detail, this work revealed diversity in shape and size [19]. While they were not necessary for the formation of closed structures, different enzymes appeared to be hierarchically involved in the formation of larger particles, though their structures could not be resolved by cryo-EM [19].

The experimental studies on miniaturized shells have emphasized the importance of identifying a suitable composition of BMC (and BMV) paralogs for assembly. In some cases, the resolution of the structural studies has not fully distinguished the identities of similar paralogs in shells that contain complex mixtures of BMC proteins [19,44,56]. Thus, some uncertainties remain in modeling the key atomic interactions between components; models where distinctions between multiple BMC paralogs are ambiguous exhibit surface complementarity between BMC and BMV components with values that are somewhat lower than seen in well-resolved BMC-BMC interfaces (for example, 0.4 vs 0.6 or higher).

An important and unexpected observation from the structures of miniaturized synthetic MCPs concerns the orientation of the protein layer forming the shell. To date, all the cases on characterized mini shells show that BMC proteins are oriented with their concave faces oriented outward towards the cytosol. Early structural studies on BMC and BMV proteins offered both orientations (concave in or concave out) as possibilities [12], but biochemical and mutational

studies provided evidence to suggest that concave faces interact with internal enzymes [58,59], which would require that flat faces to be oriented outward towards the cytosol. Additional biochemical data could be vital for clarifying this outstanding issue. In order to elucidate the exact arrangements of shell proteins, higher resolution structural data on intact native MCPs produced in-situ and containing interior enzymes will be essential. Encapsulated enzymes are particularly critical, given the important roles that interior enzymes have been shown to play in organizing the external shell in some systems [19,60–62].

Several Cryo-electron tomography (cryo-ET) studies have begun to pursue in-situ MCP structural elucidation. Multiple cryo-ET studies on carboxysomes, which are the most geometrically regular of the MCP types, have confirmed their roughly icosahedral shape revealing nearly flat facets and identifiable edges, though detailed structural features of the individual shell proteins, including shell protein orientation, have not been resolved [63–65]. Moreover, some degree of order has been seen for the encapsulated RuBisCO molecules [63,64,66–68]. The metabolosome MCPs present even greater challenges owing to their more irregular, polymorphic shapes [69]. Even the degree to which their polyhedral architectures might be described within the broader scheme of irregular icosahedra remains unclear. Recent studies highlight the possible need to bring new kinds of analysis to this problem [62,70].

Conclusion and outlook

Bacterial microcompartments are extraordinary examples of how complex protein assemblies have evolved to provide subcellular organization and compartmentalization in bacterial cells. Their ability to form robust supramolecular architectures from a complex mixture of homologous shell proteins rivals similar phenomena seen in large viruses. A great deal of structural data has revealed nuances in topological variations, conformational flexibility and quaternary polymorphisms in MCP shell proteins, highlighting the role that duplication has played in supporting functional diversification. Studies on miniaturized MCPs have affirmed models for

larger scale shell architecture, though the absence of internal enzymes presents an important gap that will ultimately need to be bridged in order to understand native MCPs. In particular, the shells of metabolosome MCPs are considerably more irregular compared to the icosahedral and nearly icosahedral assembly models presently available. Cryo-ET studies pushed to higher resolution limits may be essential for achieving a fuller understanding of MCPs. In parallel, computational simulations could be informative regarding biophysical parameters that might govern assembly architecture. New large-scale assembly simulation methods are beginning to provide insights along this line [71,72].

Acknowledgements

This work was supported by NIH/NIAID Grant R01AI081146.

The authors declare no competing interests.

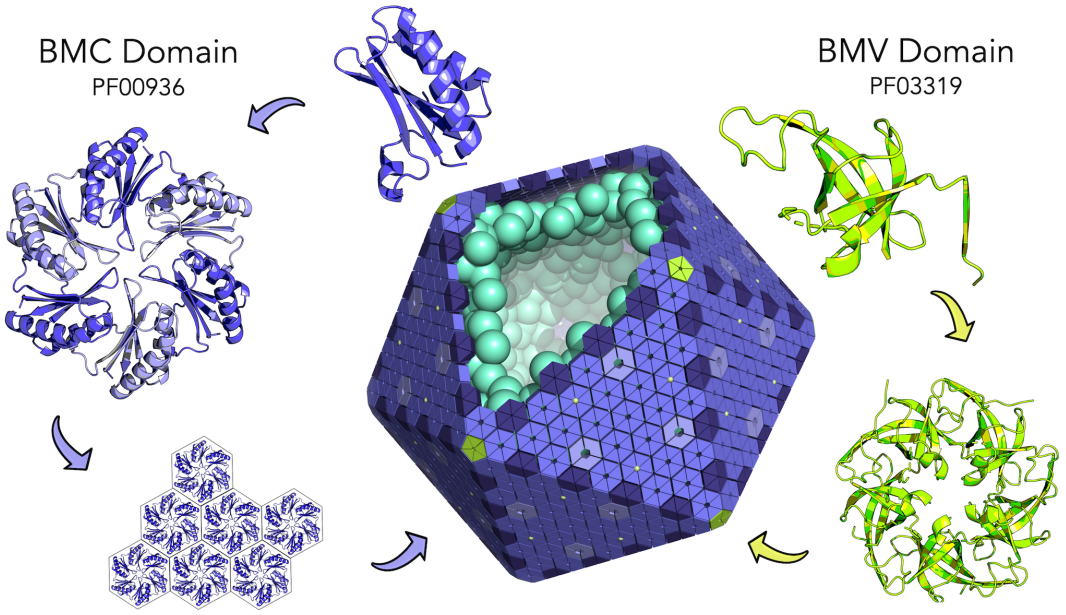
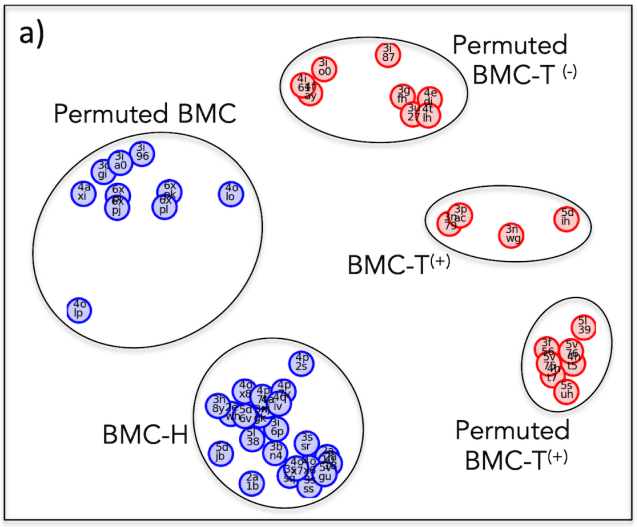


Figure 3.1

Assembly principles of bacterial microcompartments (MCPs). The BMC protein domain (purple) oligomerizes to form hexameric disks. Hundreds of hexamers tesselate side by side to form the nearly flat facets of the outer microcompartment shell (center). The BMV domain (chartreuse) is a distinctly different protein component, which oligomerizes to form pentamers at the polyhedral vertices of the microcompartment shell. Encapsulated enzymes are diagrammed in cyan. This idealized microcompartment shell is shown as a regular icosahedron. Most types of MCPs are polymorphic and less regular in shape.


a)
Permuted BMC-T (-)
Permuted BMC
BMC-T(+)
BMC-H
Permuted BMC-T(+)

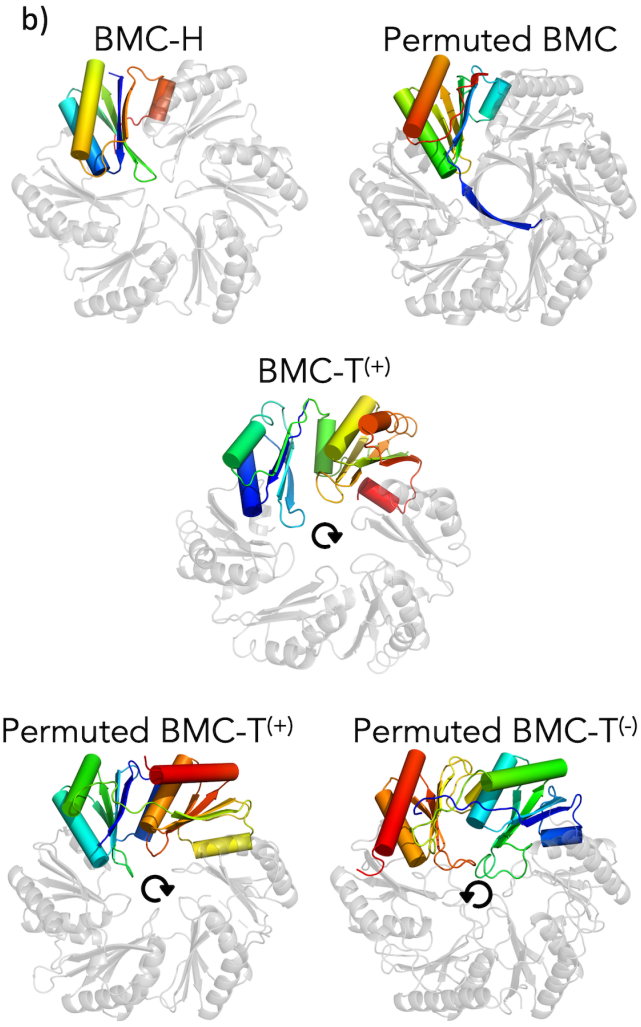

b)
BMC-H
Permuted BMC
BMC-T(+)
Permuted BMC-T(+)
Permuted BMC-T(-)

Figure 3.2

Different BMC shell proteins exhibit varied tertiary structures. (a) Based on comparisons of three-dimensional similarity with sequential ordering enforced, BMC shell proteins of known structure (PDB codes shown) cluster into five distinct populations, representing subtypes within the two major families: BMC-H which assemble as hexamers from a single BMC domain and BMC-T which assemble as trimers from two BMC domains. In the case of BMC-Ts, superscript notations denote whether the sequential domains are arranged clockwise (+) or counterclockwise (-) in a disk, as shown in panel b. (b) Rainbow-colored chains (N-terminus blue, C-terminus red) and arrows highlight the domain organizations in different BMC subtypes.

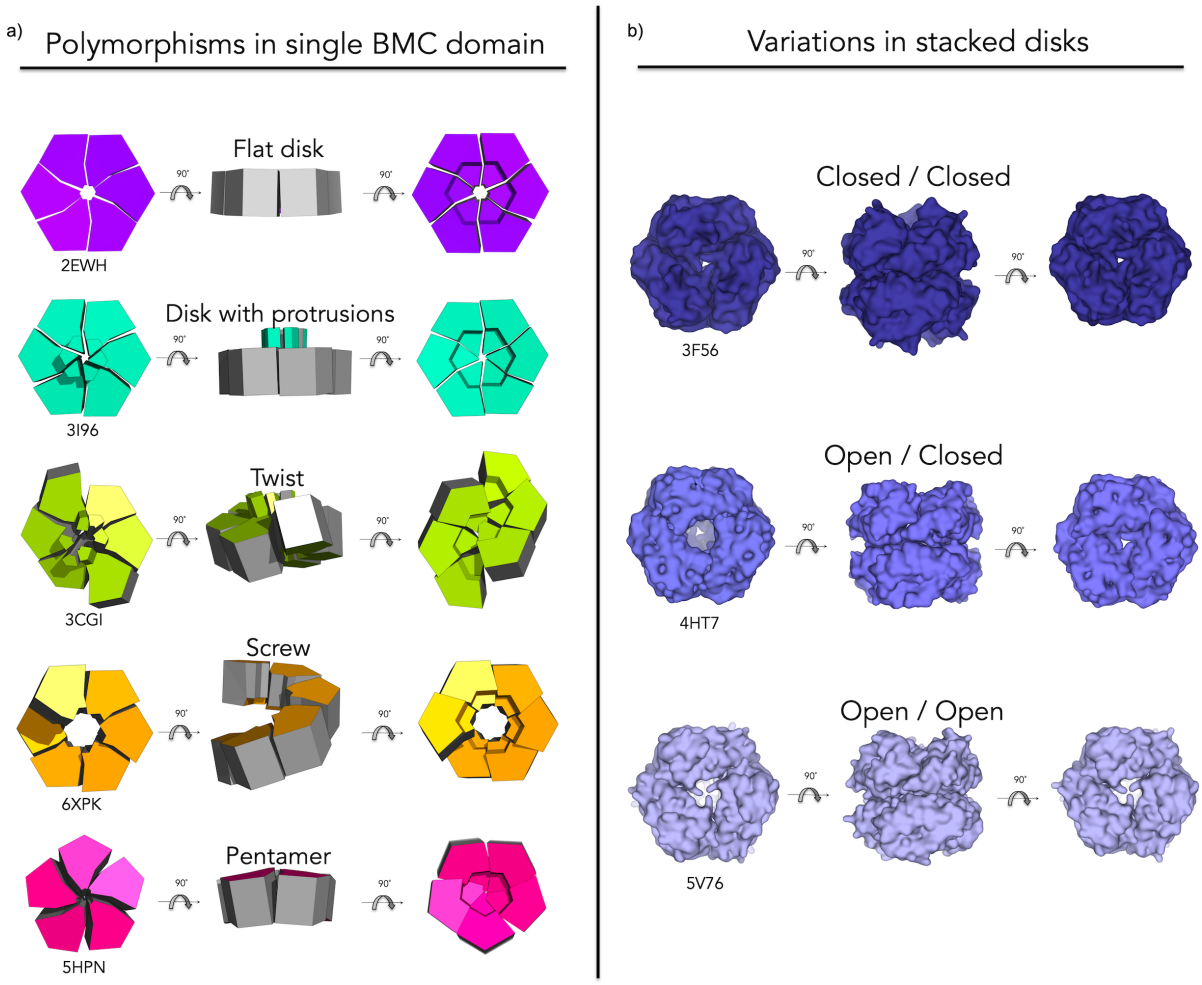


Figure 3.3

BMC shell proteins exhibit quaternary structure variations and several modes of flexibility. (a) Geometric renderings highlight variations that have been observed especially within the Permuted BMC protein type including: flat disks (purple), flat disks with beta-barrel protrusions (turquoise), twisted structures (chartreuse), six-fold screws (orange) and rearrangement to pentamers (magenta). (b) Proteins of the Permuted BMC-T(+) type are often observed as stacked disks with variations in pore openness/occlusion. Stacked disks have been reported with two closed pores (deep blue), one open and one closed pore (slate) or two open pores (light blue).

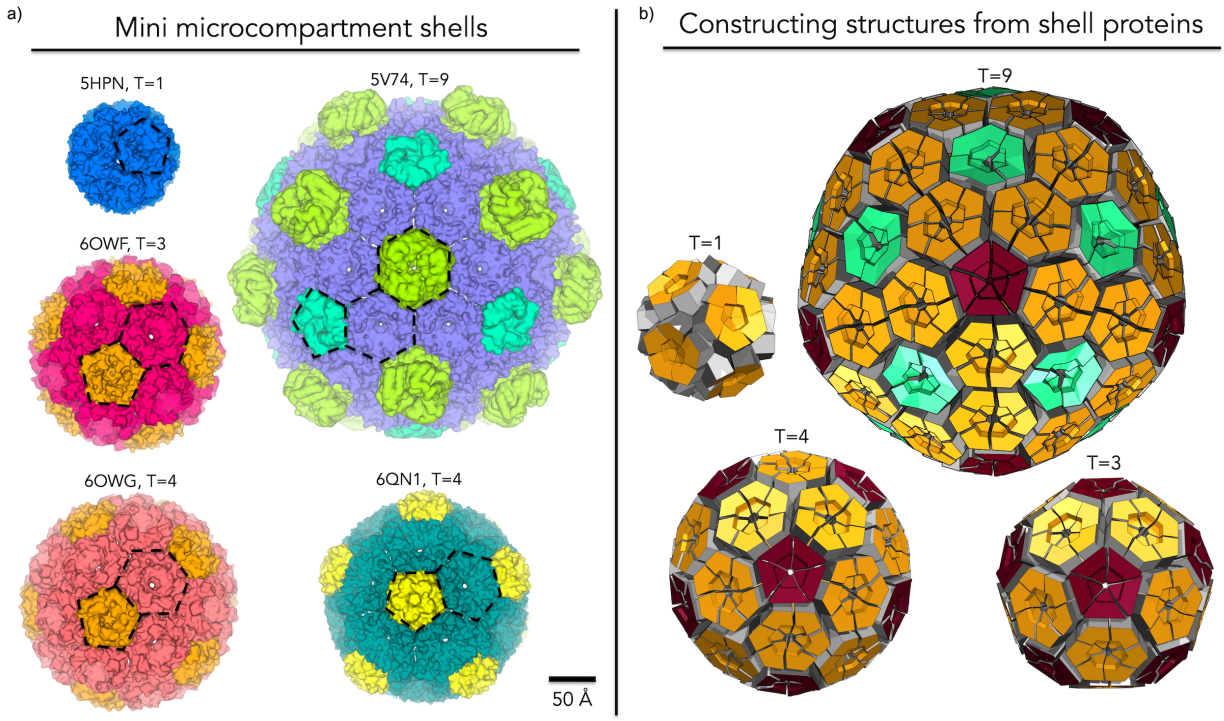


Figure 3.4

Gallery of miniaturized MCP shells highlighting diversity in shape, size and number of components. (a) First columns (top to bottom): PDB: 5HPN – Shell from an engineered circularly permuted BMC shell protein, PduA, which formed a pentamer; PDB: 6OWF – Shell constructed from one BMV and one BMC-H from a beta-carboxysome, T=3; PDB: 6OWG – Shell constructed from one BMV and one BMC-H from a beta-carboxysome, T=4. Second column (top to bottom): PDB: 5V74 – 6.5 MDa shell constructed from one BMV, one BMC-H and three types of presently indistinguishable BMC-T proteins from a Haliangium ochraceum MCP. PDB: 6QN1 – mini GRM2 shell constructed from one BMV and three presently indistinguishable BMC-Hs. (b) Geometric models representing different icosahedral triangulationpatternsobserved. BMV(red),BMC-H(orange),andBMC-T(aquamarine). IntheT=1case an engineered BMC-H protein rearranged to form pentameric units.

REFERENCES

1. Bobik TA, Lehman BP, Yeates TO: Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways. Molecular Microbiology 2015, 98:193–207.
2. Kerfeld CA, Aussignargues C, Zarzycki J, Cai F, Sutter M: Bacterial microcompartments. Nature Reviews Microbiology 2018, doi:10.1038/nrmicro.2018.10.
3. Stewart KL, Stewart AM, Bobik TA: Prokaryotic Organelles: Bacterial Microcompartments in E. coli and Salmonella. EcoSal Plus 2020, 9.
4. Kerfeld CA, Sawaya MR, Tanaka S, Nguyen CV, Phillips M, Beeby M, Yeates TO: Protein Structures Forming the Shell of Primitive Bacterial Organelles. Science 2005, 309:936–938.
5. Tsai Y, Sawaya MR, Cannon GC, Cai F, Williams EB, Heinhorst S, Kerfeld CA, Yeates TO: Structural Analysis of CsoS1A and the Protein Shell of the Halothiobacillus neapolitanus Carboxysome. PLoS Biol 2007, 5.
6. Dryden KA, Crowley CS, Tanaka S, Yeates TO, Yeager M: Two-dimensional crystals of carboxysome shell proteins recapitulate the hexagonal packing of three-dimensional crystals. Protein Sci 2009, 18:2629–2635.
7. Sutter M, Faulkner M, Aussignargues C, Paasch BC, Barrett S, Kerfeld CA, Liu L-N: Visualization of Bacterial Microcompartment Facet Assembly Using High-Speed Atomic Force Microscopy. Nano Lett 2016, 16:1590–1595.
8. Yeates TO, Jorda J, Bobik TA: The Shells of BMC-Type Microcompartment Organelles in Bacteria. Journal of Molecular Microbiology and Biotechnology 2013, 23:290–299.
9. Sinha S, Cheng S, Sung YW, McNamara DE, Sawaya MR, Yeates TO, Bobik TA: Alanine Scanning Mutagenesis Identifies an Asparagine–Arginine–Lysine Triad Essential to Assembly of the Shell of the Pdu Microcompartment. Journal of Molecular Biology 2014, 426:2328–2345.

10. Pang A, Frank S, Brown I, Warren MJ, Pickersgill RW: Structural Insights into Higher Order Assembly and Function of the Bacterial Microcompartment Protein PduA. J Biol Chem 2014, 289:22377–22384.
11. 11.Uddin I, Frank S, Warren MJ, Pickersgill RW: A Generic Self-Assembly Process in Microcompartments and Synthetic Protein Nanotubes. Small 2018, 14:1704020.
12. Tanaka S, Kerfeld CA, Sawaya MR, Cai F, Heinhorst S, Cannon GC, Yeates TO: Atomic-Level Models of the Bacterial Carboxysome Shell. Science 2008, 319:1083–1086.
13. Wheatley NM, Gidaniyan SD, Liu Y, Cascio D, Yeates TO: Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. Protein Sci 2013, 22:660–665.
14. Sutter M, Wilson SC, Deutsch S, Kerfeld CA: Two new high-resolution crystal structures of carboxysome pentamer proteins reveal high structural conservation of CcmL orthologs among distantly related cyanobacterial species. Photosynth Res 2013, 118:9–16.
15. Yeates TO, Kerfeld CA, Heinhorst S, Cannon GC, Shively JM: Protein-based organelles in bacteria: carboxysomes and related microcompartments. Nature Reviews Microbiology 2008, 6:681–691.
16. Kerfeld CA, Heinhorst S, Cannon GC: Bacterial microcompartments. Annu Rev Microbiol 2010, 64:391–408.
17. Cannon GC, Bradburne CE, Aldrich HC, Baker SH, Heinhorst S, Shively JM: Microcompartments in Prokaryotes: Carboxysomes and Related Polyhedra. Appl Environ Microbiol 2001, 67:5351– 5361.
18. Herring TI, Harris TN, Chowdhury C, Mohanty SK, Bobik TA: A Bacterial Microcompartment Is Used for Choline Fermentation by Escherichia coli 536. J Bacteriol 2018, 200:e00764-17.

19. Kalnins G, Cesle E-E, Jansons J, Liepins J, Filimonenko A, Tars K: Encapsulation mechanisms and structural studies of GRM2 bacterial microcompartment particles. Nature Communications 2020, 11.
20. Jorda J, Lopez D, Wheatley NM, Yeates TO: Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. Protein Science 2013, 22:179–195.
21. Axen SD, Erbilgin O, Kerfeld CA: A Taxonomy of Bacterial Microcompartment Loci Constructed by a Novel Scoring Method. PLOS Computational Biology 2014, 10:e1003898.
22. Ferlez B, Sutter M, Kerfeld CA: Glycyl Radical Enzyme-Associated Microcompartments: Redox- Replete Bacterial Organelles. mBio 2019, 10.
23. Petit E, LaTouf WG, Coppi MV, Warnick TA, Currie D, Romashko I, Deshpande S, Haas K, Alvelo- Maurosa JG, Wardman C, et al.: Involvement of a Bacterial Microcompartment in the Metabolism of Fucose and Rhamnose by Clostridium phytofermentans. PLOS ONE 2013, 8:e54337.
24. Thompson MC, Wheatley NM, Jorda J, Sawaya MR, Gidaniyan SD, Ahmed H, Yang Z, McCarty KN, Whitelegge JP, Yeates TO: Identification of a Unique Fe-S Cluster Binding Site in a Glycyl- Radical Type Microcompartment Shell Protein. J Mol Biol 2014, 426:3287–3304.
25. Zarzycki J, Sutter M, Cortina NS, Erb TJ, Kerfeld CA: In Vitro Characterization and Concerted Function of Three Core Enzymes of a Glycyl Radical Enzyme - Associated Bacterial Microcompartment. Sci Rep 2017, 7.
26. Lundin AP, Stewart KL, Stewart AM, Herring TI, Chowdhury C, Bobik TA: Genetic Characterization of a Glycyl Radical Microcompartment Used for 1,2-Propanediol Fermentation by Uropathogenic Escherichia coli CFT073. Journal of Bacteriology 2020, 202.
27. Mallette E, Kimber MS: Structure and Kinetics of the S -(+)-1-Amino-2-propanol Dehydrogenase from the RMM Microcompartment of Mycobacterium smegmatis. Biochemistry 2018, 57:3780– 3789.

28. Mallette E, Kimber MS: Structural and kinetic characterization of (S)-1-amino-2-propanol kinase from the aminoacetone utilization microcompartment of Mycobacterium smegmatis. J Biol Chem 2018, doi:10.1074/jbc.RA118.005485.
29. Ravcheev DA, Moussu L, Smajic S, Thiele I: Comparative Genomic Analysis Reveals Novel Microcompartment-Associated Metabolic Pathways in the Human Gut Microbiome. Front Genet 2019, 10.
30. Burley SK, Berman HM, Kleywegt GJ, Markley JL, Nakamura H, Velankar S: Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. In Protein Crystallography: Methods and Protocols. Edited by Wlodawer A, Dauter Z, Jaskolski M. Springer; 2017:627–641.
31. Ochoa JM, Bair K, Holton T, Bobik TA, Yeates TO: MCPdb: The Bacterial Microcompartment Database. bioRxiv 2021, doi:10.1101/2021.01.09.426059. *Under review
32. Crowley CS, Sawaya MR, Bobik TA, Yeates TO: Structure of the PduU Shell Protein from the Pdu Microcompartment of Salmonella. Structure 2008, 16:1324–1332.
33. Tanaka S, Sawaya MR, Yeates TO: Structure and Mechanisms of a Protein-Based Organelle in Escherichia coli. Science 2010, 327:81–84.
34. Ochoa JM, Nguyen VN, Nie M, Sawaya MR, Bobik TA, Yeates TO: Symmetry breaking and structural polymorphism in a bacterial microcompartment shell protein for choline utilization. Protein Science 2020, 29:2201–2212.
35. Tanaka S, Sawaya MR, Phillips M, Yeates TO: Insights from multiple structures of the shell proteins from the β-carboxysome. Protein Science 2009, 18:108–120.
36. Sagermann M, Ohtaki A, Nikolakakis K: Crystal structure of the EutL shell protein of the ethanolamine ammonia lyase microcompartment. Proc Natl Acad Sci U S A 2009, 106:8883–8887.
37. Klein MG, Zwart P, Bagby SC, Cai F, Chisholm SW, Heinhorst S, Cannon GC, Kerfeld CA:

38. Identification and Structural Analysis of a Novel Carboxysome Shell Protein with Implications for Metabolite Transport. Journal of Molecular Biology 2009, 392:319–333.
39. Crowley CS, Cascio D, Sawaya MR, Kopstein JS, Bobik TA, Yeates TO: Structural Insight into the Mechanisms of Transport across the Salmonella enterica Pdu Microcompartment Shell. J Biol Chem 2010, 285:37838–37846.
40. Pang A, Warren MJ, Pickersgill RW: Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe–4S cluster-binding site. Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr 2011, 67:91–96.
41. Aussignargues C, Pandelia M-E, Sutter M, Plegaria JS, Zarzycki J, Turmo A, Huang J, Ducat DC, Hegg EL, Gibney BR, et al.: Structure and Function of a Bacterial Microcompartment Shell Protein Engineered to Bind a [4Fe-4S] Cluster. J Am Chem Soc 2016, 138:5262–5270.
42. Thompson MC, Cascio D, Leibly DJ, Yeates TO: An allosteric model for control of pore opening by substrate binding in the EutL microcompartment shell protein. Protein Science 2015, 24:956–975.
43. 42.Mallette E, Kimber MS: A Complete Structural Inventory of the Mycobacterial Microcompartment Shell Proteins Constrains Models of Global Architecture and Transport. Journal of Biological Chemistry 2017, 292:1197–1210.
44. Cai F, Sutter M, Cameron JC, Stanley DN, Kinney JN, Kerfeld CA: The Structure of CcmP, a Tandem Bacterial Microcompartment Domain Protein from the β-Carboxysome, Forms a Subcompartment Within a Microcompartment. J Biol Chem 2013, 288:16055–16063.
45. Sutter M, Greber B, Aussignargues C, Kerfeld CA: Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. Science 2017, 356:1293–1297.

46. Larsson AM, Hasse D, Valegård K, Andersson I: Crystal structures of β-carboxysome shell protein CcmP: ligand binding correlates with the closed or open central pore. Journal of Experimental Botany 2017, 68:3857–3867.
47. Chowdhury C, Chun S, Pang A, Sawaya MR, Sinha S, Yeates TO, Bobik TA: Selective molecular transport through the protein shell of a bacterial microcompartment organelle. PNAS 2015, 112:2990–2995.
48. Faulkner M, Szabó I, Weetman SL, Sicard F, Huber RG, Bond PJ, Rosta E, Liu L-N: Molecular simulations unravel the molecular principles that mediate selective permeability of carboxysome shell protein. Scientific Reports 2020, 10:17501.
49. Slininger Lee MF, Jakobson CM, Tullman-Ercek D: Evidence for Improved Encapsulated Pathway Behavior in a Bacterial Microcompartment through Shell Protein Engineering. ACS Synth Biol 2017, 6:1880–1891.
50. Park J, Chun S, Bobik TA, Houk KN, Yeates TO: Molecular Dynamics Simulations of Selective Metabolite Transport across the Propanediol Bacterial Microcompartment Shell. J Phys Chem B 2017, 121:8149–8154.
51. Mangan NM, Flamholz A, Hood RD, Milo R, Savage DF: pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism. PNAS 2016, 113:E5354–E5362.
52. Jakobson CM, Tullman-Ercek D, Slininger MF, Mangan NM: A systems-level model reveals that 1,2-Propanediol utilization microcompartments enhance pathway flux through intermediate sequestration. PLOS Computational Biology 2017, 13:e1005525.
53. Larsson AM, Hasse D, Valegård K, Andersson I: Crystal structures of β-carboxysome shell protein CcmP: ligand binding correlates with the closed or open central pore. Journal of Experimental Botany 2017, 68:3857–3867.

54. Jorda J, Leibly DJ, Thompson MC, Yeates TO: Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein. Chem Commun 2016, 52:5041–5044.
55. Cai F, Menon BB, Cannon GC, Curry KJ, Shively JM, Heinhorst S: The Pentameric Vertex Proteins Are Necessary for the Icosahedral Carboxysome Shell to Function as a CO 2 Leakage Barrier. PLOS ONE 2009, 4:e7521.
56. Forouhar F, Kuzin A, Seetharaman J, Lee I, Zhou W, Abashidze M, Chen Y, Yong W, Janjua H, Fang Y, et al.: Functional insights from structural genomics. Journal of structural and functional genomics 2007, 8:37–44.
57. Greber BJ, Sutter M, Kerfeld CA: The Plasticity of Molecular Interactions Governs Bacterial Microcompartment Shell Assembly. Structure 2019, 27:749-763.e4.
58. Sutter M, Laughlin TG, Sloan NB, Serwas D, Davies KM, Kerfeld CA: Structure of a Synthetic β- Carboxysome Shell. Plant Physiology 2019, 181:1050–1058.
59. Fan C, Cheng S, Liu Y, Escobar CM, Crowley CS, Jefferson RE, Yeates TO, Bobik TA: Short N- terminal sequences package proteins into bacterial microcompartments. PNAS 2010, 107:7509–7514.
60. Thompson MC, Crowley CS, Kopstein J, Bobik TA, Yeates TO: Structure of a bacterial microcompartment shell protein bound to a cobalamin cofactor. Acta Crystallographica Section F Structural Biology Communications 2014, 70:1584–1590.
61. Chen AH, Afonso B, Silver PA, Savage DF: Spatial and Temporal Organization of Chromosome Duplication and Segregation in the Cyanobacterium Synechococcus elongatus PCC 7942. PLOS ONE 2012, 7:e47837.
62. Cameron JC, Wilson SC, Bernstein SL, Kerfeld CA: Biogenesis of a Bacterial Organelle: The Carboxysome Assembly Pathway. Cell 2013, 155:1131–1140.

63. Juodeikis R, Lee MJ, Mayer M, Mantell J, Brown IR, Verkade P, Woolfson DN, Prentice MB, Frank S, Warren MJ: Effect of metabolosome encapsulation peptides on enzyme activity, coaggregation, incorporation, and bacterial microcompartment formation. MicrobiologyOpen 2020, 9:e1010.
64. Schmid MF, Paredes AM, Khant HA, Soyer F, Aldrich HC, Chiu W, Shively JM: Structure of Halothiobacillus neapolitanus Carboxysomes by Cryo-electron Tomography. Journal of Molecular Biology 2006, 364:526–535.
65. Iancu CV, Morris DM, Dou Z, Heinhorst S, Cannon GC, Jensen GJ: Organization, Structure, and Assembly of α-Carboxysomes Determined by Electron Cryotomography of Intact Cells. Journal of Molecular Biology 2010, 396:105–117.
66. Dai W, Fu C, Raytcheva D, Flanagan J, Khant HA, Liu X, Rochat RH, Haase-Pettingell C, Piret J, Ludtke SJ, et al.: Visualizing virus assembly intermediates inside marine cyanobacteria. Nature 2013, 502:707–710.
67. Dai W, Chen M, Myers C, Ludtke SJ, Pettitt BM, King JA, Schmid MF, Chiu W: Visualizing Individual RuBisCO and Its Assembly into Carboxysomes in Marine Cyanobacteria by Cryo-Electron Tomography. Journal of Molecular Biology 2018, 430:4156–4167.
68. Iancu CV, Ding HJ, Morris DM, Dias DP, Gonzales AD, Martino A, Jensen GJ: The Structure of Isolated Synechococcus Strain WH8102 Carboxysomes as Revealed by Electron Cryotomography. Journal of Molecular Biology 2007, 372:764–773.
69. Dai W, Fu C, Raytcheva D, Flanagan J, Khant HA, Liu X, Rochat RH, Haase-Pettingell C, Piret J, Ludtke SJ, et al.: Visualizing Virus Assembly Intermediates Inside Marine Cyanobacteria. Nature 2013, 502:707–710.
70. Kennedy NW, Hershewe JM, Nichols TM, Roth EW, Wilke CD, Mills CE, Jewett MC, Tullman-Ercek D: Apparent size and morphology of bacterial microcompartments varies with technique. PLOS ONE 2020, 15:e0226395.

71. Bag S, Prentice MB, Liang M, Warren MJ, Roy Choudhury K: Classification of polyhedral shapes from individual anisotropically resolved cryo-electron tomography reconstructions. BMC Bioinformatics 2016, 17:234.
72. Mohajerani F, Hagan MF: The role of the encapsulated cargo in microcompartment assembly. PLOS Computational Biology 2018, 14:e1006351.
73. Mohajerani F, Sayer E, Neil C, Inlow K, Hagan MF: Mechanisms of scaffold-mediated microcompartment assembly and size-control. bioRxiv 2020, doi:10.1101/2020.10.14.338509. *Under review

Chapter 4: Symmetry breaking and structural polymorphism in a bacterial microcompartment shell protein for choline utilization

Received: 13 July 2020 | Revised: 28 August 2020 | Accepted: 1 September 2020
DOI: 10.1002/pro.3941

ARTICLE

THE PROTEIN SOCIETY WILEY

Symmetry breaking and structural polymorphism in a bacterial microcompartment shell protein for choline utilization

Jessica M. Ochoa[1] | **Vy N. Nguyen**[2] | **Mengxiao Nie**[2] | **Michael R. Sawaya**[3] | **Thomas A. Bobik**[4] | **Todd O. Yeates**[1,2,3]

[1]UCLA-Molecular Biology Institute, University of California, Los Angeles (UCLA), California, Los Angeles

[2]Department of Chemistry and Biochemistry, University of California, Los Angeles (UCLA), California, Los Angeles

[3]UCLA-DOE Institute of Genomics and Proteomics, University of California, Los Angeles (UCLA), California, Los Angeles

[4]Department of Biochemistry, Biophysics and Molecular Biology; Iowa State University, Ames, Iowa

Correspondence
Todd O. Yeates, UCLA Department of Chemistry 611 Charles E. Young Drive East Los Angeles, CA 90095.
Email: yeates@mbi.ucla.edu

Funding information
National Institute of Allergy and Infectious Diseases, Grant/Award Number: NIAID R01AI081146; National Institute of General Medical Sciences, Grant/Award Number: P30 GM124165; National Institutes of Health, Grant/Award Number: S10OD021527; Office of Science, Grant/Award Number: DE-AC02-06CH11357

Abstract

Bacterial microcompartments are protein-based organelles that carry out specialized metabolic functions in diverse bacteria. Their outer shells are built from several thousand protein subunits. Some of the architectural principles of bacterial microcompartments have been articulated, with lateral packing of flat hexameric BMC proteins providing the basic foundation for assembly. Nonetheless, a complete understanding has been elusive, partly owing to polymorphic mechanisms of assembly exhibited by most microcompartment types. An earlier study of one homologous BMC shell protein subfamily, EutS/PduU, revealed a profoundly bent, rather than flat, hexameric structure. The possibility of a specialized architectural role was hypothesized, but artifactual effects of crystallization could not be ruled out. Here we report a series of crystal structures of an orthologous protein, CutR, from a glycyl-radical type choline-utilizing microcompartment from the bacterium *Streptococcus intermedius*. Depending on crystal form, expression construct, and minor mutations, a range of novel quaternary architectures was observed, including two spiral hexagonal assemblies. A new graphical approach helps illuminate the variations in BMC hexameric structure, with results substantiating the idea that the EutS/PduU/CutR subfamily of BMC proteins may endow microcompartment shells with flexible modes of assembly.

KEYWORDS

bacterial microcompartment, bacterial organelle, carboxysome, choline, glycyl radical, polymorphism, shell protein, symmetry breaking

1 | INTRODUCTION

Nearly 20% of bacteria species produce giant protein-based organelles that are used for carrying out sensitive metabolic reactions in a sequestered cellular environment.[1] These extraordinary structures, known as bacterial microcompartments or MCPs (or alternatively BMCs), encapsulate distinct enzyme types in different bacteria.[1–4] Generally, a small metabolic intermediate (either volatile, toxic, or both) is produced within the MCP and is further metabolized before it can escape into the cytosol or out of the cell entirely.[2,5–7] These key metabolic intermediates include CO_2 for the case of the carboxysome, the founding member of the MCP family, while various aldehyde intermediates occur in other types of MCPs, including the type under investigation in the

present study. Comprised of thousands of protein subunits, MCPs are among the largest and most mechanistically sophisticated protein assemblies to have evolved in nature.[8,9]

A protein shell, reminiscent of a viral capsid, is the hallmark of MCPs (Figure 1). Different MCP types, even those with divergent metabolic functions, are all assembled from homologous shell proteins. The bulk of the MCP shell is composed of proteins containing a roughly 100-amino acid BMC protein domain (Pfam PF00936). The recognition that this protein family (first identified by Shively and coworkers[10]) is encoded across diverse bacterial operons has provided a bioinformatics basis for exploring MCP function and evolution.[1,4,11,12] It is now understood that BMC proteins assemble to form nearly flat faces of the MCP shell, while a second minor and distinct protein (known as a BMV protein, Pfam PF03319) forms vertices.[13–17] Multiple paralogs of BMC proteins, typically ranging from two to seven in number, are present in different systems. Different BMC protein paralogs offer modularity in the shell. Some BMC proteins provide pores for substrate and product diffusion across the shell, some are believed to bind and organize interior enzymes, some may serve specialized architectural roles, while others may have yet unrecognized functions.

Structural studies on MCP proteins have shed light on architectural and functional mechanisms, with the basic principles emerging from crystallographic studies beginning fifteen years ago.[18,19] BMC proteins of the canonical type form flat hexameric units; these are the building blocks for shell assembly. Typically, BMC hexamers possess narrow pores for molecular transport. These hexamers tile side by side to form extended facets of nearly solid protein perforated by small holes. A remarkable feature of BMC shell proteins is the wide range of structural rearrangements that have occurred through evolution, thereby conferring distinct properties on multiple BMC paralogs within an operon. Notable variations include circular permutations of the chain to produce BMC versions with termini in different locations,[20] versions bearing tandem BMC domains,[16,21–23] versions bearing iron–sulfur clusters and other small molecules in their central pores,[24–27] and versions fused to other protein domains.[23] Biochemical, genetic, and engineering studies have provided additional insights into the prospective roles of different BMC shell proteins,[16,28–33] but much remains unknown about separate and distinct (or even redundant) functions that might be attributed to different BMC paralogs. Functional distinctions between BMC shell proteins may furthermore vary between different MCP types, based on differences in metabolic purpose, BMC paralog composition, and MCP architecture; for example, some MCP shells, like the carboxysome, appear much more geometrically regular (i.e., icosahedral) (Figure 1d) than those that metabolize various organic metabolites (Figure 1e).

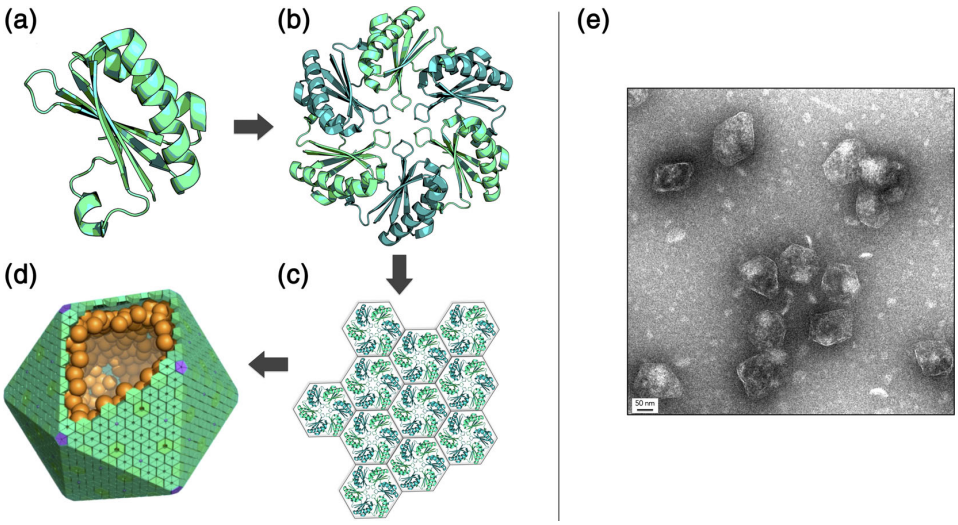


FIGURE 1 Microcompartment shells are composed of homologous hexameric proteins. For all bacterial microcompartments, the canonical BMC domain (a) oligomerizes to form the traditional flat hexamer (b). These hexameric shell proteins tessellate (c) to form the nearly flat faces of bacterial microcompartments shells. (d) An idealized model of a microcompartment with hexameric BMC proteins that form the face (teal), pentameric BMV proteins that form the vertices (purple), and internal encapsulated enzymes (orange). (e) Negative stain EM of purified Pdu microcompartments (scale bar: 50 nm)

One type of BMC shell protein paralog of special interest is EutS/PduU (from the ethanolamine utilization and propanediol utilization MCPs, respectively). Proteins from this subfamily are a permuted version of the canonical BMC domain. By virtue of their permuted topology, EutS/PduU paralogs present distinct features in the form of a protruding, N-terminal six-stranded parallel beta-barrel that occludes the central pore. Remarkably, an early crystallographic study of EutS from *E. coli* revealed a unique structure in which the hexamer had undergone an extraordinary twisting deformation of approximately 40°,[23] a stark contrast to other BMC hexamers that are very nearly flat, a general expectation for cyclic homooligomeric assemblies. This peculiar observation invoked the possibility of an important architectural role. However, other structures of this paralog (e.g., PduU and even a point mutant of EutS) revealed a typical flat hexameric assembly,[13,20] leaving uncertainty about the meaning and significance of the dramatic departure from the typical BMC structure. The relevance versus artifactual nature of the structural variation observed in EutS has not been revisited since that initial observation a decade ago. In the present work, we undertook a series of crystal structure investigations on an orthologous protein, CutR, from a different type of MCP from *Streptococcus intermedius*. We present further observations of major polymorphism in this shell protein subfamily and discuss implications for a specialized architectural role in certain bacterial metabolic organelles.

2 | RESULTS

2.1 | CutR crystallized as five unique structures

In order to investigate the structural polymorphism of a permuted BMC shell protein, we undertook a series of crystallographic studies of CutR, a BMC shell protein from the type 1 choline microcompartment (choline utilization or Cut MCP) from *Streptococcus intermedius*. When an initial structure revealed a completely unexpected, and likely artificial, dimeric structure, further mutagenesis and structural studies were pursued to dissect the possible effects of individual amino acids and terminal purification tags. We report five crystal structures obtained from four sequence variants. These structural forms include one novel dimer, two traditional flat hexamers, and two novel six-fold screws of varying pitch. Geometric and computational analyses helped illuminate details of the observed quaternary forms.

The first structure we obtained for CutR revealed a novel dimeric arrangement not previously observed in other BMC proteins (Figure 2). A model was refined to a resolution of 1.8 Å with final R_{work}/R_{free} values of 0.176/0.210. Interestingly, this unexpected dimeric form of CutR crystallized from gel-filtration fractions whose elution positions were consistent with that of a BMC hexamer (a molecular weight of approximately 80 kDa based on standard curves). In this unexpected structure, the two monomeric subunits came together, exchanging terminal segments in a fashion characteristic of domain

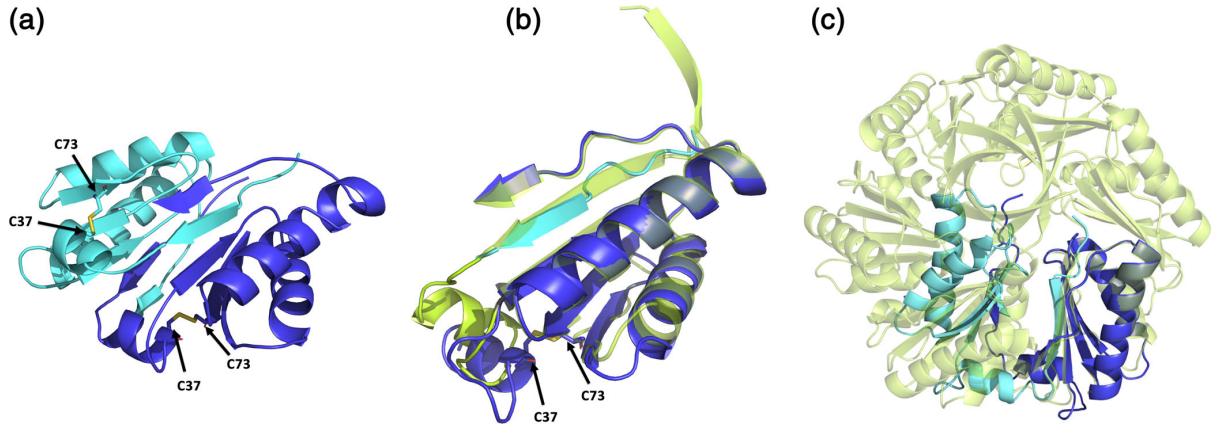


FIGURE 2 A comparison of a novel CutR dimeric form to the traditional flat hexameric form of the same protein. (a) Cartoon representation of the observed novel dimeric form. A disulfide bridge forms between C37 and C73, pulling the N-terminus from its traditional position in the BMC domain, causing residues 18–28 to occupy the adjacent monomer by domain swapping. (b) Overlay of a monomer from the hexameric form (lime green) with residues 18–28 of one monomer (cyan) and residues 29–116 from the second monomer (deep blue), with emphasis on regions that differ. (c) Superimposition of one chain from the dimer with one chain from the flat hexamer shows that severe steric clashes would be caused by the presence of the dimeric form in the context of the hexameric assembly

swapping.[34] Moreover, we observed an unexpected intramolecular disulfide bond between C37 and C73 (Figure 2a). The disulfide bond pulls the first N-terminal beta-strand (residues G18-A28) out of its anticipated location and into a position contacting the adjacent monomer (Figure 2a). This domain swap preserves the overall BMC tertiary structure. The N-terminal beta-strand from one monomer (residues G18-A28) packs alongside the remaining residues (A29-S116) of the second monomer, forming intermolecular interactions that results in a structure resembling the traditional BMC fold [0.50 Å RMSD backbone deviation (Figure 2b)].

While this dimeric architecture of CutR presented a new and intriguing structure, it could not be reconciled with the quaternary hexameric arrangement understood for BMC shell proteins or their packing in a layer. Aligning one monomer from the dimeric form with one monomer from the hexameric form results in severe steric clashes of their quaternary structures (Figure 2c). Notwithstanding the effects of an apparently adventitious disulfide bond, the unusual quaternary structure that we observed prompted further structural studies to look for potential modes of polymorphic assembly. We thus conducted structural studies on sequence variants of the protein as well as structural studies under reducing conditions. This work led to multiple structures with additional novel quaternary arrangements.

Motivated by the unconventional dimeric form, we pursued mutagenesis work to determine whether the circularly permuted CutR was capable of forming a traditional flat BMC hexamer. We obtained two distinct crystal forms of CutR, with subsequent structural analysis revealing canonical flat hexamers. The first crystal form referred to as Hexamer 1 (Figure 3a) crystallized in space group C2 and provided an atomic structure to a resolution of 2.6 Å with R_{work}/R_{free} values of 0.201/0.242. We also solved the structure of a second crystal form, referred to as Hexamer 2 (Figure 3b). Hexamer 2 crystallized in space group $P4_22_12$ and provided a structure to a resolution of 1.5 Å with R_{work}/R_{free} values of 0.161/0.184. We obtained numerous crystals with varied morphologies including flat hexagonal plates and octahedral bipyramidal forms. Importantly, two separate mutants of CutR each gave rise to traditional flat hexagonal BMC structures. The first mutant retained an N-terminal His-6 tag and had a C37A mutation. The second mutant contained a cleavable His-6 tag and a K66A mutation, instead of the original K66D mutation (Table S1). Hexamer 2 was the most well behaved, presumably owing to the non-polar residue at position 66 (discussed subsequently) located at the edge and to the lack of obstruction from the N-terminal His-6 tag. This version purified with relative ease and formed numerous diffraction quality crystals within 1 month of setting up broad screen crystallization trials. Crystals that formed in 2 M ammonium sulfate, 0.2 M lithium sulfate, and 0.1 M CAPS/sodium hydroxide pH 10.5 had the strongest diffraction pattern and the highest resolution, ultimately

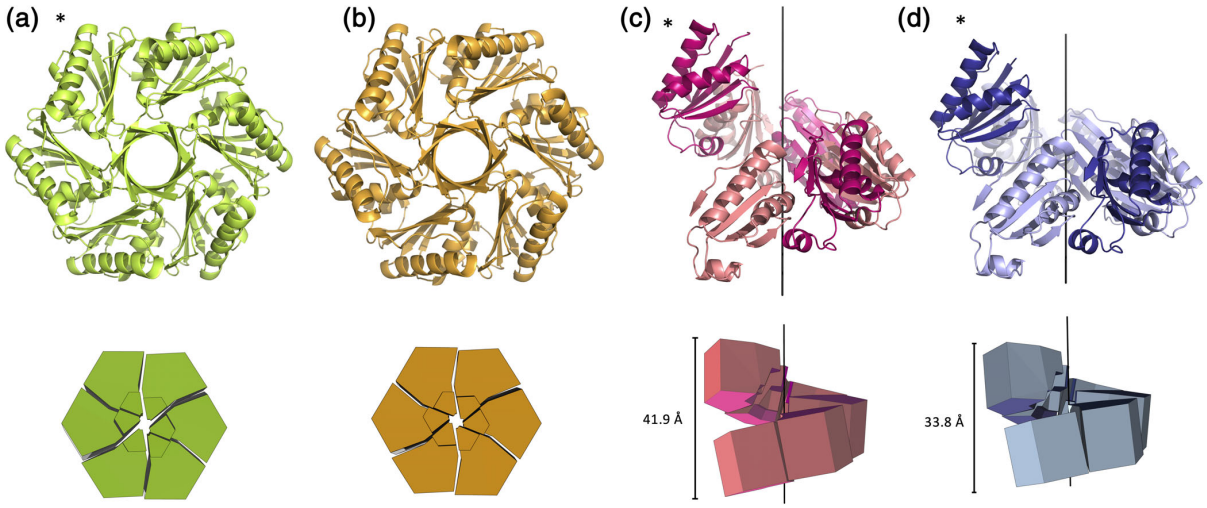


FIGURE 3 Cartoon and geometric representation of four CutR polymorphs. CutR crystallized as a traditional flat BMC hexamer (a and b) and formed two distinct screws with varying pitch (c and d). The flat hexamers are viewed down the six-fold axis of symmetry (a and b) and the screws are displayed on their sides to show their pitch (c and d). Geometric representations reveal that in the flat form, Hexamer 1 (a) and Hexamer 2 (b) both deviate slightly from perfect C6 symmetry. Screw 1 (c) has a pitch of 41.9 Å and Screw 2 (d) has a pitch of 33.8 Å. Asterisks indicate constructs that crystalized while retaining an N-terminal His-6 tag

giving rise to the Hexamer 2 structure. As with other circularly-permuted BMC domains, specifically PduU[20] (PDB ID 3CGI) and EutS[23] (PDB ID 3I96), we observed a beta-barrel on the flat face created by the protruding N-termini of the six chains (Figure 3a,b). The interior of the beta-barrel is decorated with the side chains of alternating residues I11, Q13, and S15. As with PduU,[20] the bulky side chains of the interior-lining residues are unable to occupy symmetry-equivalent positions. Specifically, the six instances of Q13 adopt two conformations: pointing up toward the N-terminus and down toward the center of the hexameric disk.

We also determined the structure of two novel BMC domain-based structures in which the quaternary structure formed a six-fold screw axis. The first hereinafter referred to as Screw 1, came from the same CutR construct that gave rise to the dimer, which still contains a cysteine at position 37, an aspartic acid at position 66 (significance explained below) and an N-terminal His-6 tag (Table S1). We solved this structure to 2.8 Å with R_{work}/R_{free} values of 0.231/0.267. Screw 1 is a right-handed screw with a pitch of 41.9 Å (Figure 3c). This screw has a 6 sub 1 axis of helical symmetry. Looking down the axis, we observe that the structure forms an apparent hexamer that except for its pitch in the z-axis, has a similar morphology to Hexamer 2 (Figure S1). Measuring from C-alpha to C-alpha of corresponding positions, Screw1 and Hexamer 2 have diameters with maxima of 69.2 and 71.1 Å, respectively (Figure S1). The unexpected nature of the screw form prompted further studies.

Owing to purification challenges from initial work, including aggregation and dynamic oligomerization, and because we observed novel and unexpected structures, we sought to purify an additional mutant version, CutR_D66A, under reducing conditions. CutR_D66A was purified in the presence of TCEP to disrupt the disulfide bridge observed in the initial dimeric structure and has a point mutation at residue 66. In the native form, this residue is normally occupied by a conserved lysine, which participates in intermolecular salt bridges to support the lateral tessellation of BMC hexamers in the shell (Figure 1c); this lysine has been widely mutated in laboratory studies to allow isolation of individual hexamers.[35,36] We hypothesized that the aspartic acid in this version played a significant role in our inability to purify well-behaved, monodisperse species of the CutR construct. This new CutR_D66A mutant retained the His-6 affinity tag and was purified with relative ease. Surprisingly, this construct gave rise to a second structure having a six-fold screw axis, hereinafter referred to as Screw 2. Like Screw 1, Screw 2 forms a right-handed screw, though its pitch of 33.8 Å is considerably shorter than that of Screw 1. We solved the structure to 3.3 Å with R_{work}/R_{free} values of 0.219/0.265. Like the previously observed screw, Screw 2 also creates an apparent hexamer when looking down the six-fold axis of symmetry with a similar diameter of 70.1 Å (Figure S1). Both screw structures were obtained with an intact N-terminal His-6 tag.

2.2 | Analysis of quaternary polymorphisms in CutR

We found value in a new graphical approach for visualizing structural arrangements in the BMC shell protein family. Our prior studies have emphasized the canonical packing of BMC subunits into roughly flat hexameric units. BMC subunits typically fit together like six slightly twisted pie pieces with a central depression commonly found on one side of the full hexamer. Furthermore, as noted in this work, the permuted BMC family bears a small protruding beta-barrel on the opposite side of the hexameric disk. Based on the well-established packing of canonical BMC hexamers, we established a coordinate system for mapping a simple pie-shaped structure onto the conserved secondary structural elements of a BMC protomer. With this mapping established, it was possible to generate diverse packing diagrams to illustrate the wide-ranging quaternary arrangements observed in the permuted BMC family (Figure 4). This was helpful for interpreting the nature of the nuanced variations and for complementing numerical calculations such as angular rotations and atomic coordinate shifts.

Previous studies of the circularly-permuted BMC proteins PduU (PDB ID 3CGI) and the bent EutS (PDB ID 3I96) have been interesting study cases of the BMC domain. Both maintain the traditional BMC domain fold but comparing adjacent monomer pairs reveals remarkable flexibility. We established a system for evaluating the angular rotations and shifts observed in the EutS structures. We overlapped the A subunit of a given structure with the A subunit of a canonical BMC hexamer, and then evaluated the difference between the adjacent B subunits from the two structures being compared (Figure 4). These comparisons are summarized here and in Table 1. As anticipated, traditional flat hexamers from this permuted BMC family, including PduU and the CutR hexamers from this work, have RMSD values of less than 1 Å, while a comparison of the flat hexamers to EutS (bent) yields RMSD values between 15 and 18 Å. Likewise, the rotation angles required to align corresponding B monomers of flat permuted BMC hexamers were less than 2°, while alignment of the corresponding B monomer from any flat hexamer to the corresponding B monomer from EutS required up to 53° of rotation in order to

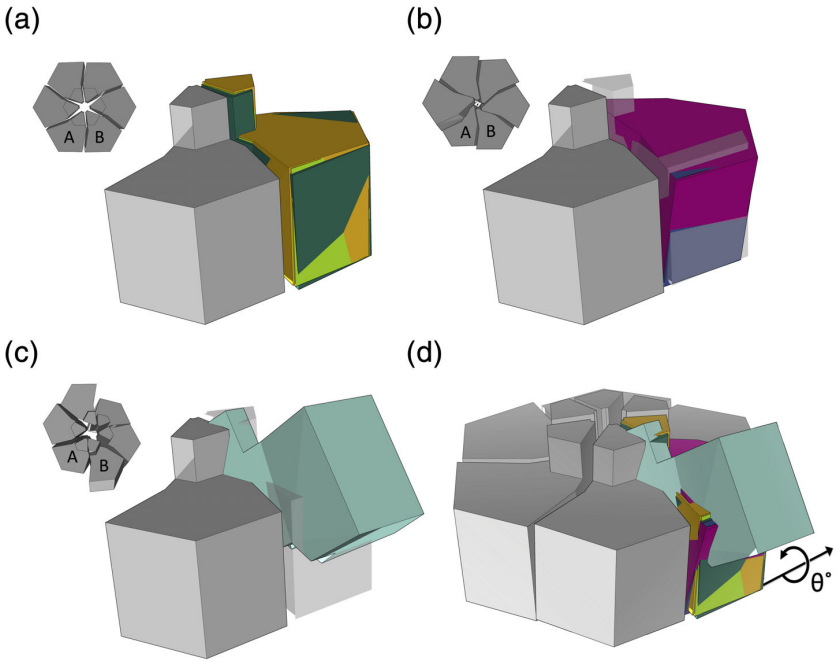


FIGURE 4 Summary of the quaternary structure deformations observed in the CutR/EutS/PduU subfamily of BMC shell proteins. (a) Geometric representation of adjacent pairs of permuted BMC shell protein subunits from three flat or very nearly flat hexameric structures (CutR Hexamer 1, CutR Hexamer 2, and PduU) after aligning their A subunits (gray). The relative position of the B subunits is shown for CutR Hexamer 1 (lime green), CutR Hexamer 2 (orange), and PduU (forest green). (b) Similar representations for the two screw forms (CutR Screw 1 (magenta) and CutR Screw 2 (purple) relative to the flat CutR form (gray). (c) A similar representation of the bent EutS structure (teal, PDB ID 3I96) relative to the flat CutR form (gray). (d) Overlay of all A monomers relative to the flat hexamer with the B monomers colored as follows: Cut R flat Hexamer 1 (lime green), Cut R flat Hexamer 2 (orange), PduU (forest green), CutR Screw 1 (magenta), CutR Screw 2 (purple), and EutS (teal). The rotation of adjacent subunits in alternate assembly forms, as calculated in Table 1, is diagrammed

TABLE 1 Comparison of structural deviations across the permuted BMC family

–	PduU	Hex1	Hex2	Screw1	Screw2	EutS
PduU	–	0.5 Å	0.7 Å	3.5 Å	2.8 Å	15.8 Å
Hex1	1.6°	–	0.2 Å	3.5 Å	2.7 Å	15.6 Å
Hex2	1.7°	0.6°	–	3.4 Å	2.7 Å	15.4 Å
Screw1	14.4°	16.5°	16.7°	–	0.8 Å	17.1 Å
Screw2	11.7°	13.6°	13.9°	2.9°	–	17.1 Å
EutS	42.0°	40.4°	40.4°	55.9°	52.9°	–

Note: RMSD values (upper right) shown are based on adjacent AB monomers for four hexameric CutR polymorphs, along with PduU and EutS. The values describe backbone deviations in corresponding B subunits after superimposing A subunits. The rotation angle required to align the B monomers for any given pair of adjacent monomers is reported in the bottom left.

achieve the optimal overlap (Table 1). Unsurprisingly, a comparison of the two CutR screws shows relatively minor coordinate deviations (0.8 Å RMSD) and angular differences (2.9°), as the two screw forms vary mainly in pitch.

Motivated by the variability observed between the structures of CutR polymorphs, we checked for evidence of dynamic behavior in solution studies. Using another BMC-domain protein (EutL, a tandem BMC-domain protein from *Clostridium perfringens*) as a control, we compared elution profiles of recursive size exclusion chromatography runs of the various CutR constructs (Figure S2). While recursive size exclusion of the EutL control resulted in a single sharp peak, recursive size

exclusion of the hexameric peak from CutR constructs resulted in either broad peaks or multiple peaks, supporting a tendency to reequilibrate between multiple conformations or assembly forms in solution.

Finally, the variable packing modes in CutR motivated a study of atomic interactions at the subunit interfaces. A general finding was that for most of the alternate packing variations observed, the bending modes between subunits can occur without substantial perturbations to key atomic interactions. The exceptions to this are the (presumptively artifactual) dimeric form, and the highly bent hexameric form initially observed by Tanaka et al.[23] Structural variation in the latter case, in addition to the bending motion, is accompanied by a twist of roughly 40°, which brings alternative secondary structure packing interactions into play. For the various flat and screw hexamers observed for CutR, major disruptions are not observed. The retention of a basic hexameric (flat or screw) shape despite high flexibility raises a question as to whether a small number of particularly strong interactions in CutR might be important for integrity. A potential role for the beta-barrel in that regard was noted previously.[20] Our analysis showed that while the extended portions of the beta-barrel are disrupted in the screw forms, other interactions near the central regions and at the interfaces of the hexamer could be important. Using manual evaluation and the PDBePISA server,[37] we determined that residues S50, E55, D62, S81, Y101, K115, and S116 may play key stabilizing roles. A hydrogen bond network involving these key residues is present in both the flat hexamer and in the screw forms of CutR (Figure 5).

3 | DISCUSSION

Since first determining the structures of PduU[20] and EutS,[23] the single-domain permuted BMC family has been an interesting but under-studied special class of the BMC family. In this work, we sought to investigate another orthologue from this class, CutR from *Streptococcus intermedius*. We conclude that the circularly-permuted BMC protein CutR has an innate and dynamic capacity to sample alternative quaternary forms, with its canonical C6 symmetry breaking down in various ways. Based on these findings, we hypothesize that the permuted BMC family may play a key role in conferring heterogeneity and flexibility in certain types of bacterial microcompartments.

In this work, we elucidated five crystal structures from four CutR constructs, observing a dimer, two flat hexamers, and two six-fold screws of varying pitch. In addition to variations driven by different crystallization conditions and crystal packing forces, it is likely that variations in the different constructs played important roles in dictating the diverse outcomes. The presence or absence of cysteine (C37), the presence or absence of a charged residue in the traditional edge-lysine position (K66), and the presence or absence of an N-terminal His-6 purification tag (Table S1) are all relevant

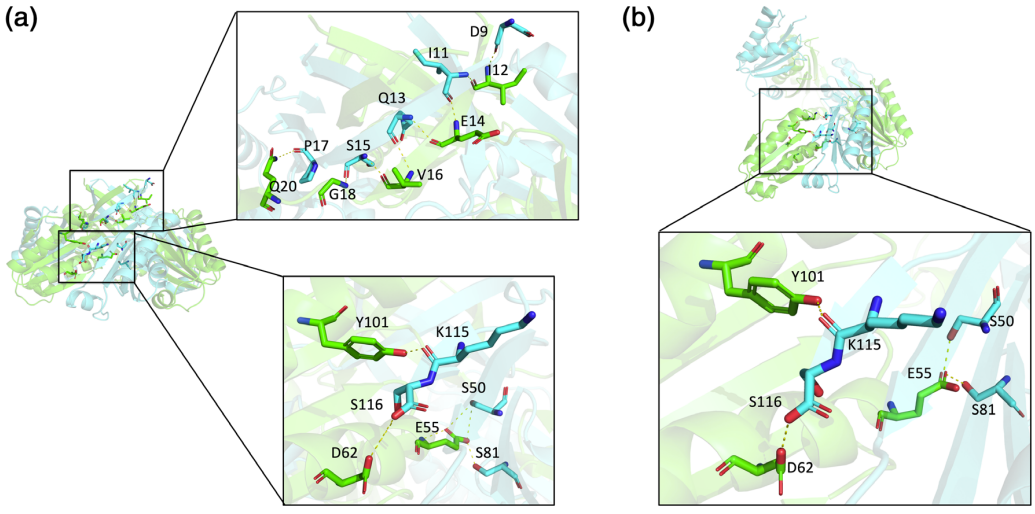


FIGURE 5 The extensive hydrogen bond network of the beta-barrel from a flat hexamer compared to the hydrogen bond network of the screw. (a) The flat hexameric form has a beta-barrel with an extensive hydrogen bond network. There are several key residues essential for maintaining the overall hexameric architecture. (b) In the screw form, the hydrogen bond network of the beta-barrel is largely lost, but other key intermolecular interactions near the pore are retained

variations. Certain correlations between constructs and quaternary forms are notable. The formation of a screw was only observed in structures that retained an N-terminal His-6 tag (Table S1). However, the presence of an N-terminal purification tag did not force the formation of a helical arrangement, as evidenced by the formation of flat Hexamer 1. A further intriguing question concerns how (or at what point) helical architectures are formed. We note that the recombinant protein that gave rise to both screw forms, based on size exclusion chromatography, had an estimated molecular weight consistent with a discrete hexamer (roughly 80 kDa). This argues for ordinary (non-helical) hexameric forms being predominant, yet retaining the flexibility to undergo considerable distortions and interface disruptions under diverse, relatively mild, and solution conditions. Significantly, the screws retained key intermolecular interactions that are shared with the flat hexameric form (Figure 5).

The vast majority of homomeric protein assemblies are arranged in a symmetric fashion. The reasons for this were articulated as early as Crick and Watson,[38] and have been expanded upon and surveyed many times since.[39–41] The evident tendency toward asymmetry, therefore, puts CutR and its orthologues in a unique category. Symmetry breaking in protein assemblies has been described in other areas. In particular, it has been explored in the context of molecular motors where alternating configurations occur during catalytically-driven motions[42] as well as in viral capsids.[43] In the latter category, as explained by Casper and Klug, alternative conformations that break symmetry are required to accommodate different packing environments.[44] The unusual polymorphism seen in CutR may help explain the heterogeneity and structural irregularity seen in many types of bacterial microcompartments.

Carboxysomes are among the most extensively studied bacterial microcompartments, and they are typically the most regular in shape. EM studies have shown that carboxysome microcompartments exhibit a relatively high degree of homogeneity and geometric regularity.[45–51] In a few cases, miniaturized versions of microcompartment shells have been produced from synthetic operons. Those have proven to assemble with high levels of order, even obeying icosahedral symmetry, with examples including a beta-carboxysome shell, a GRM2 microcompartment shell, and a shell from a microcompartment of unknown function,[15,30,31] but native forms of these MCPs likely diverge in important ways. Unlike their more symmetrical carboxysome counterparts, other native microcompartments have been found to have polymorphic structures[52,53] (Figure 1e). These include several MCP types, such as Pdu, Eut, and Cut. Sometimes classified as "metabolosomes", these MCPs degrade propanediol, ethanolamine, choline, aminoalcohols, and other small metabolites.[1–5,16] Circularly permuted BMC paralogs (PduU, EutS, CutR) are found in several MCPs of the metabolosome type, but as far as we know these circularly permuted paralogs are not present in carboxysome microcompartments, nor have they been included in the synthetic miniaturized MCP shells that have proven to be geometrically regular. Some insight is available from genetic studies in the Pdu system, which metabolizes 1,2-propanediol. While the permuted shell protein PduU is not required for the formation of intact microcompartment shells (as it is a relatively low-abundance component of the shell), PduU deletion mutants showed growth defects, with an increase in lag time while growing to higher cell density compared to their wildtype counterparts.[28] The dynamic nature of permuted BMC proteins, highlighted in the present study, helps explain the impaired growth rates of PduU deletion mutants, considering that dynamic behavior could be important for degradation, recycling, and seeding new microcompartment shells.

The highly twisted EutS structure originally reported by Tanaka[13] and the CutR polymorphs presented here may only be a small population of the potential quaternary structure variations that the permuted BMC family is capable of forming. Beyond a potential role in dynamics, increased shell flexibility could allow for packing a greater variety of internal enzymes in non-carboxysomal MCPs, or more variable enzyme stoichiometries. The processes that govern MCP assembly and disassembly remain only partially understood,[15,54–57] and flexibility could be important for those processes. The flexibility observed in this protein family invites additional questions about microcompartment structure, evolution, and function. Investigation of additional orthologues could add further insights.

4 | MATERIALS AND METHODS

4.1 | Negative stain electron microscopy

Intact Pdu microcompartments were purified from *S. enterica* serovar Typhimurium LT2 and imaged with negative stain electron microscopy as previously described by Havemann et al. and Sinha et al.[58,59] Briefly, *S. enterica* serovar Typhimurium LT2 was grown in 400 ml of growth medium containing 1X NCE, 1 mM $MgSO_4$, 0.5% succinate, and 0.6% 1,2-PD. Cultures were grown at 37°C overnight, shaking. Cells were harvested by centrifugation (4,000*g* for 15 min) and washed once with Buffer A, containing 50 mM Tris–HCl (pH 8.0),

500 mM KCl, 12.5 mM $MgCl_2$, and 1.5% 1,2-PD. The cells were resuspended in a lysis buffer containing a mixture of 40% Buffer A and 60% B-PER II (Thermo Scientific) supplemented with Pierce Protease Inhibitor tablets (Thermo Scientific), Lysozyme (Sigma-Aldrich), and DNase I (Sigma-Aldrich) and incubated at 30°C for 1 hr to lyse. Cell debris was removed by centrifugation at 12,000*g* for 5 min, and intact Pdu MCPs were pelleted by spinning at 20,000*g* for 40 min. Pdu MCPs were resuspended in a buffer containing 50 mM Tris–HCl (pH 8.0), 50 mM KCl, 5 mM $MgCl_2$, and 1% 1,2-PD to a concentration of 0.5 mg/ml. Pdu MCPs were placed on Formvar/Carbon 400 mesh copper grids (Ted Pella), washed twice with 0.1% 1,2-PD and stained with 5 μl of 2% uranyl acetate. Pdu MCPs were imaged using an FEI Tecnai T12 transmission electron microscope.

4.2 | Cloning, protein expression, and purification

We initially purchased a codon-optimized gBlock Gene Fragment of CutR, a EutS homolog from *Streptococcus intermedius*, from Integrated DNA Technologies (IDT). This first construct was ordered with a non-cleavable N-terminal His-6 tag and a K66D mutation. Previous work has demonstrated that mutating this edge lysine facilitates recombinant BMC protein expression and purification. We hypothesized that maintaining a charged amino acid in this position would achieve the same facile purification while maintaining structural integrity. We inserted the gBlock into the pET24a expression vector using *NdeI* and *HindIII* restriction endonuclease sites. For mutagenesis work, we also purchased primers from IDT. Primers were designed based on the previously published Quikchange protocol.[60] We explored a variety of mutants including C37A, to disrupt the observed disulfide bridge, D66A to further facilitate expression and purification, and insertion of a TEV-cleavage site (ENLYFQG) just after the His-6 tag, in order to assess the role of the His-6 tag in crystal packing.

We used the BL21(DE3) *E. coli* expression system (New England Biolabs) to express recombinant protein. Briefly, we used 6 ml 1 mg/L ampicillin supplemented LB overnight cultures to inoculate 1 L of autoinduction media. Cells were grown in TB medium supplemented with kanamycin and 5,052 autoinduction sugars.[61] Cultures were grown at 37°C for 6 hr and then 18°C overnight and subsequently harvested by centrifugation for 15 min at 5,000*g*. For a recombinant protein that resulted in the dimeric form, cells were grown in TB medium supplemented with 1 mg/L kanamycin. At OD_{600} = 1.0–1.2, we used 1 mM of IPTG to induce expression and proceeded to grow cells at 25°C overnight.

Cell pellets were resuspended in 20 mM Tris pH 8.0, 300 mM NaCl, supplemented with protease inhibitor cocktail tablets (Pierce Protease Inhibitor Tablets, EDTA-Free, Thermo Scientific), lysozyme (Sigma), DNase (Sigma), and RNase (Sigma). We lysed the resuspended pellets with sonication over ice and clarified the lysate by centrifugation (15,000*g* for 30 min). The clarified lysate was applied over a pre-equilibrated Nickel IMAC gravity column (HisPur Ni-NTA Resin, Thermo Scientific) and His-tagged samples were eluted using 20 mM Tris pH 8.0, 300 mM NaCl, 300 mM imidazole, and 10% glycerol. Samples that contained a TEV-cleavage site were subject to TEV-protease and dialyzed overnight at 4°C into 50 mM Tris pH 8.0 and 150 mM NaCl. All proteins were subject to the second round of purification using gel-filtration chromatography (Superdex 200 10/300 GL, GE Healthcare) and eluted with a buffer containing 50 mM Tris pH 8.0 and 150 mM NaCl. Fractions from peaks that corresponded to a molecular weight of approximately 80 kDa were pooled together. We verified the presence and purity of protein throughout the purification process using denaturing SDS-PAGE. We also sought to investigate the role and significance of the observed disulfide bridge by using reducing agents. In one iteration of our purification procedure, we added 10 mM tris (2-carboxyethyl)phosphine (TCEP) to purification buffers. Following the two-step purification, proteins were subsequently concentrated to approximately 30 mg/ml using 10KDa MWCO Amicon Ultra concentrators (Millipore) and syringe-filtered through 0.22 μM filters (Millipore). Following concentration and filtration, this protein was used for crystallization experiments.

4.3 | Crystallization

We obtained diffraction quality crystals using a TTP Labtech robotic mosquito and the hanging-drop vapor diffusion method at the UCLA Crystallization facility. We observed several crystal forms in numerous conditions, the most prominent being flat hexagonal sheets and bi-pyramidal octahedrons. We collected diffraction data for the CutR dimer that crystallized in 0.1 M Potassium thiocyanate, 30% w/v PEG 2000. Flat Hexamer 1, which contained a C37A mutation, crystallized in 2.0 M ammonium sulfate, 0.1 M BIS-Tris, pH 5.5. Flat Hexamer 2, with a cleaved His-6 tag, crystallized in 2 M ammonium sulfate, 0.2 M lithium sulfate, and 0.1 M CAPS/Sodium hydroxide pH 10.5. Screw 1 crystallized in 10% (w/v) PEG 3000, 200 mM sodium chloride, and 100 mM sodium phosphate dibasic/Citric acid pH 4.2. Finally, Screw

2, which was purified in the presence of reducing agents and contained the D66A mutation, crystallized in 0.2 M potassium sodium tartrate tetrahydrate, 0.1 M Bis-Tris propane, 8.52% w/v PEG 3350.

4.4 | Data collection, structure determination, and refinement

X-ray diffraction datasets were collected at the Advanced Photon Source in Chicago on beamlines 24-ID-C and 24-ID-E (NE-CAT) and then indexed, integrated, and scaled using XDS/XSCALE.[62] We used PHASER[63] to first solve the structure of the CutR dimer by molecular replacement using a EutS homolog from *Clostridium difficile* (PDB ID 4AXI) as a reference model. The model was built using COOT[64] and refined using PHENIX.[65] We also used Refmac[66] and Buster[67] in subsequent rounds of refinement. The CutR dimer was refined to a final model with R_{work}/R_{free} of 0.176/0.210 at a resolution of 1.8 Å. Subsequent structures were solved by molecular replacement using one monomer of the CutR dimer as a reference using similar model-building and refinement strategies. We also utilized the PDBePISA server[37] to computationally assess the interfaces created by these various structures.

ACKNOWLEDGMENTS

This work was funded by an award from the National Institutes of Health, NIAID R01AI081146 (TAB and TOY). Diffraction studies were supported by research conducted at the Northeastern Collaborative Access Team beamlines (NECAT), which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). JMO is a Howard Hughes Medical Institute Gilliam Fellow. The Eiger 16 M detector on 24-ID-E is funded by an NIH-ORIP HEI grant (S10OD021527). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. We are grateful for the support of the NECAT staff. The authors thank Duilio Cascio and Michael Collazo for crystallographic help and advice.

AUTHOR CONTRIBUTIONS

Jessica Ochoa: Conceptualization; data curation; formal analysis; investigation; methodology; software; visualization; writing-original draft; writing-review and editing. **Vy Nguyen:** Data curation; formal analysis; investigation; validation; visualization; writing-original draft; writing-review and editing. **Mengxiao Nie:** Data curation; formal analysis; investigation. **Michael R. Sawaya:** Data curation; formal analysis; investigation; methodology; validation; writing-review and editing. **Thomas Bobik:** Conceptualization; funding acquisition; investigation; project administration; supervision; writing-review and editing. **Todd Yeates:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; software; supervision; validation; visualization; writing-original draft; writing-review and editing.

CONFLICT OF INTEREST

The authors declare no competing interests.

Data Availability

Coordinates and diffraction data are deposited at the PDB under codes: 6XPH, 6XPI, 6XPJ, 6XPK, and 6XPL.

ORCID

Todd O. Yeates https://orcid.org/0000-0001-5709-9839

REFERENCES

1. Jorda J, Lopez D, Wheatley NM, Yeates TO. Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. Protein Sci. 2013;22:179–195.
2. Chowdhury C, Sinha S, Chun S, Yeates TO, Bobik TA. Diverse bacterial microcompartment organelles. Microbiol Mol Biol Rev. 2014;78:438–468.
3. Ravcheev DA, Moussu L, Smajic S, Thiele I. Comparative genomic analysis reveals novel microcompartment-associated metabolic pathways in the human gut microbiome. Front Genet. 2019;10:636.
4. Axen SD, Erbilgin O, Kerfeld CA. A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. PLoS Comput Biol. 2014;10:e1003898.
5. Bobik TA, Lehman BP, Yeates TO. Bacterial microcompartments: Widespread prokaryotic organelles for isolation and optimization of metabolic pathways. Mol Microbiol. 2015;98:193–207.
6. Kerfeld CA, Heinhorst S, Cannon GC. Bacterial microcompartments. Annu Rev Microbiol. 2010;64:391–408.
7. Kerfeld CA, Aussignargues C, Zarzycki J, Cai F, Sutter M. Bacterial microcompartments. Nat Rev Microbiol. 2018;16:277–290.
8. Yeates TO, Crowley CS, Tanaka S. Bacterial microcompartment organelles: Protein shell structure and evolution. Annu Rev Biophys. 2010;39:185–205.
9. Yeates TO, Thompson MC, Bobik TA. The protein shells of bacterial microcompartment organelles. Curr Opin Struct Biol. 2011;21:223–231.
10. English RS, Lorbach SC, Qin X, Shively JM. Isolation and characterization of a carboxysome shell gene from Thiobacillus neapolitanus. Mol Microbiol. 1994;12:647–654.

11. Bobik TA, Xu Y, Jeter RM, Otto KE, Roth JR. Propanediol utilization genes (pdu) of salmonella typhimurium: Three genes for the propanediol dehydratase. J Bacteriol. 1997;179:6633–6639.
12. Beeby M, Bobik TA, Yeates TO. Exploiting genomic patterns to discover new supramolecular protein assemblies. Protein Sci. 2009;18:69–79.
13. Tanaka S, Kerfeld CA, Sawaya MR, et al. Atomic-level models of the bacterial carboxysome shell. Science. 2008;319: 1083–1086.
14. Wheatley NM, Gidaniyan SD, Liu Y, Cascio D, Yeates TO. Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. Protein Sci. 2013;22:660–665.
15. Sutter M, Greber B, Aussignargues C, Kerfeld CA. Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. Science. 2017;356:1293–1297.
16. Mallette E, Kimber MS. A complete structural inventory of the mycobacterial microcompartment shell proteins constrains models of global architecture and transport. J Biol Chem. 2017; 292:1197–1210.
17. Sutter M, Wilson SC, Deutsch S, Kerfeld CA. Two new high-resolution crystal structures of carboxysome pentamer proteins reveal high structural conservation of CcmL orthologs among distantly related cyanobacterial species. Photosynth Res. 2013; 118:9–16.
18. Kerfeld CA, Sawaya MR, Tanaka S, et al. Protein structures forming the shell of primitive bacterial organelles. Science. 2005; 309:936–938.
19. Tsai Y, Sawaya MR, Cannon GC, et al. Structural analysis of CsoS1A and the protein shell of the Halothiobacillus neapolitanus carboxysome. PLoS Biol. 2007;5:e144.
20. Crowley CS, Sawaya MR, Bobik TA, Yeates TO. Structure of the PduU shell protein from the Pdu microcompartment of salmonella. Structure. 2008;16:1324–1332.
21. Klein MG, Zwart P, Bagby SC, et al. Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. J Mol Biol. 2009;392:319–333.
22. Sagermann M, Ohtaki A, Nikolakakis K. Crystal structure of the EutL shell protein of the ethanolamine ammonia lyase microcompartment. Proc Natl Acad Sci U S A. 2009;106: 8883–8887.
23. Tanaka S, Sawaya MR, Yeates TO. Structure and mechanisms of a protein-based organelle in *Escherichia coli*. Science. 2010; 327:81–84.
24. Crowley CS, Cascio D, Sawaya MR, Kopstein JS, Bobik TA, Yeates TO. Structural insight into the mechanisms of transport across the *Salmonella enterica* Pdu microcompartment shell. J Biol Chem. 2010;285:37838–37846.
25. Pang A, Warren MJ, Pickersgill RW. Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe–4S cluster-binding site. Acta Crystallogr. 2011;D67:91–96.
26. Aussignargues C, Pandelia M-E, Sutter M, et al. Structure and function of a bacterial microcompartment shell protein engineered to bind a [4Fe-4S] cluster. J Am Chem Soc. 2016; 138:5262–5270.
27. Samborska B, Kimber MS. A dodecameric CcmK2 structure suggests β-carboxysomal shell facets have a double-layered organization. Structure. 2012;20:1353–1362.
28. Cheng S, Sinha S, Fan C, Liu Y, Bobik TA. Genetic analysis of the protein shell of the microcompartments involved in coenzyme B12-dependent 1,2-propanediol degradation by salmonella. J Bacteriol. 2011;193:1385–1392.
29. Parsons JB, Lawrence AD, McLean KJ, Munro AW, Rigby SEJ, Warren MJ. Characterisation of PduS, the pdu metabolosome corrin reductase, and evidence of substructural organisation within the bacterial microcompartment. PLOS ONE. 2010;5: e14009.
30. Sutter M, Laughlin TG, Sloan NB, Serwas D, Davies KM, Kerfeld CA. Structure of a synthetic β-carboxysome shell. Plant Physiol. 2019;181:1050–1058.
31. Kalnins G, Cesle E-E, Jansons J, Liepins J, Filimonenko A, Tars K. Encapsulation mechanisms and structural studies of GRM2 bacterial microcompartment particles. Nat Commun. 2020;11:388.
32. Lee MJ, Mantell J, Hodgson L, et al. Engineered synthetic scaffolds for organizing proteins within the bacterial cytoplasm. Nat Chem Biol. 2018;14:142–147.
33. Kinney JN, Axen SD, Kerfeld CA. Comparative analysis of carboxysome shell proteins. Photosynth Res. 2011;109:21–32.
34. Bennett MJ, Schlunegger MP, Eisenberg D. 3D domain swapping: A mechanism for oligomer assembly. Protein Sci. 1995;4: 2455–2468.
35. Sinha S, Cheng S, Sung YW, et al. Alanine scanning mutagenesis identifies an asparagine-arginine-lysine triad essential to assembly of the shell of the Pdu microcompartment. J Mol Biol. 2014;426:2328–2345.
36. Pang A, Frank S, Brown I, Warren MJ, Pickersgill RW. Structural insights into higher order assembly and function of the bacterial microcompartment protein PduA. J Biol Chem. 2014; 289:22377–22384.
37. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007;372:774–797.
38. Crick FHC, Watson JD. Structure of small viruses. Nature. 1956;177:473–475.
39. André I, Strauss CEM, Kaplan DB, Bradley P, Baker D. Emergence of symmetry in homooligomeric biological assemblies. Proc Natl Acad Sci U S A. 2008;105:16148–16152.
40. Goodsell DS, Olson AJ. Structural symmetry and protein function. Annu Rev Biophys Biomol Struct. 2000;29:105–153.
41. Cannon KA, Ochoa JM, Yeates TO. High-symmetry protein assemblies: Patterns and emerging applications. Curr Opin Struct Biol. 2019;55:77–84.
42. Sun SX, Wang H, Oster G. Asymmetry in the F1-ATPase and its implications for the rotational cycle. Biophys J. 2004;86: 1373–1384.
43. Harrison SC. The familiar and the unexpected in structures of icosahedral viruses. Curr Opin Struct Biol. 2001;11:195–199.
44. Caspar DLD, Klug A. Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol. 1962; 27:1–24.
45. Shively JM, Ball FL, Kline BW. Electron microscopy of the carboxysomes (polyhedral bodies) of Thiobacillus neapolitanus. J Bacteriol. 1973;116:1405–1411.
46. Cannon GC, Bradburne CE, Aldrich HC, Baker SH, Heinhorst S, Shively JM. Microcompartments in prokaryotes: Carboxysomes and related polyhedra. Appl Environ Microbiol. 2001;67:5351–5361.
47. Price GD, Coleman JR, Badger MR. Association of carbonic anhydrase activity with carboxysomes isolated from the

cyanobacterium *Synechococcus* PCC7942. Plant Physiol. 1992; 100:784–793.
48. Long BM, Badger MR, Whitney SM, Price GD. Analysis of carboxysomes from *Synechococcus* PCC7942 reveals multiple rubisco complexes with carboxysomal proteins CcmM and CcaA. J Biol Chem. 2007;282:29323–29335.
49. Iancu CV, Morris DM, Dou Z, Heinhorst S, Cannon GC, Jensen GJ. Organization, structure, and assembly of α-carboxysomes determined by electron cryotomography of intact cells. J Mol Biol. 2010;396:105–117.
50. Dai W, Chen M, Myers C, et al. Visualizing individual RuBisCO and its assembly into carboxysomes in marine cyanobacteria by cryo-electron tomography. J Mol Biol. 2018;430: 4156–4167.
51. Dai W, Fu C, Raytcheva D, et al. Visualizing virus assembly intermediates inside marine cyanobacteria. Nature. 2013;502: 707–710.
52. Bag S, Prentice MB, Liang M, Warren MJ, Roy Choudhury K. Classification of polyhedral shapes from individual anisotropically resolved cryo-electron tomography reconstructions. BMC Bioinformatics. 2016;17:234.
53. Kennedy NW, Hershewe JM, Nichols TM, et al. Apparent size and morphology of bacterial microcompartments varies with technique. PloS One. 2020;15:e0226395.
54. Bonacci W, Teng PK, Afonso B, et al. Modularity of a carbon-fixing protein organelle. Proc Natl Acad Sci U S A. 2012;109: 478–483.
55. Hill NC, Tay JW, Altus S, Bortz DM, Cameron JC. Life cycle of a cyanobacterial carboxysome. Sci Adv. 2020;6: eaba1269.
56. Cameron JC, Wilson SC, Bernstein SL, Kerfeld CA. Biogenesis of a bacterial organelle: The carboxysome assembly pathway. Cell. 2013;155:1131–1140.
57. Perlmutter JD, Mohajerani F, Hagan MF. Many-molecule encapsulation by an icosahedral shell. Elife. 2016;5:e14078.
58. Havemann GD, Bobik TA. Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar typhimurium LT2. J Bacteriol. 2003;185:5086–5095.
59. Sinha S, Cheng S, Fan C, Bobik TA. The PduM protein is a structural component of the microcompartments involved in coenzyme B12-dependent 1,2-propanediol degradation by salmonella. J Bacteriol. 2012;194:1912–1918.
60. Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol. 2008;8:91.
61. Studier FW. Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif. 2005;41:207–234.
62. Kabsch W. XDS. Acta Crystallogr. 2010;D66:125–132.
63. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Cryst. 2007;40:658–674.
64. Emsley P, Cowtan K. Coot: Model-building tools for molecular graphics. Acta Crystallogr. 2004;D60:2126–2132.
65. Adams PD, Grosse-Kunstleve RW, Hung LW, et al. PHENIX: Building new software for automated crystallographic structure determination. Acta Crystallogr. 2002;D58:1948–1954.
66. Winn MD, Ballard CC, Cowtan KD, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr. 2011; D67:235–242.
67. Bricogne G, Blanc E, Brandl M, et al. BUSTER version 2.10.3. Cambridge, United Kingdom: Global Phasing Ltd; 2019.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Ochoa JM, Nguyen VN, Nie M, Sawaya MR, Bobik TA, Yeates TO. Symmetry breaking and structural polymorphism in a bacterial microcompartment shell protein for choline utilization. *Protein Science*. 2020;29: 2201–2212. https://doi.org/10.1002/pro.3941

Supplementary Information for:

Symmetry Breaking and Structural Polymorphism in a Bacterial Microcompartment Shell Protein for Choline Utilization

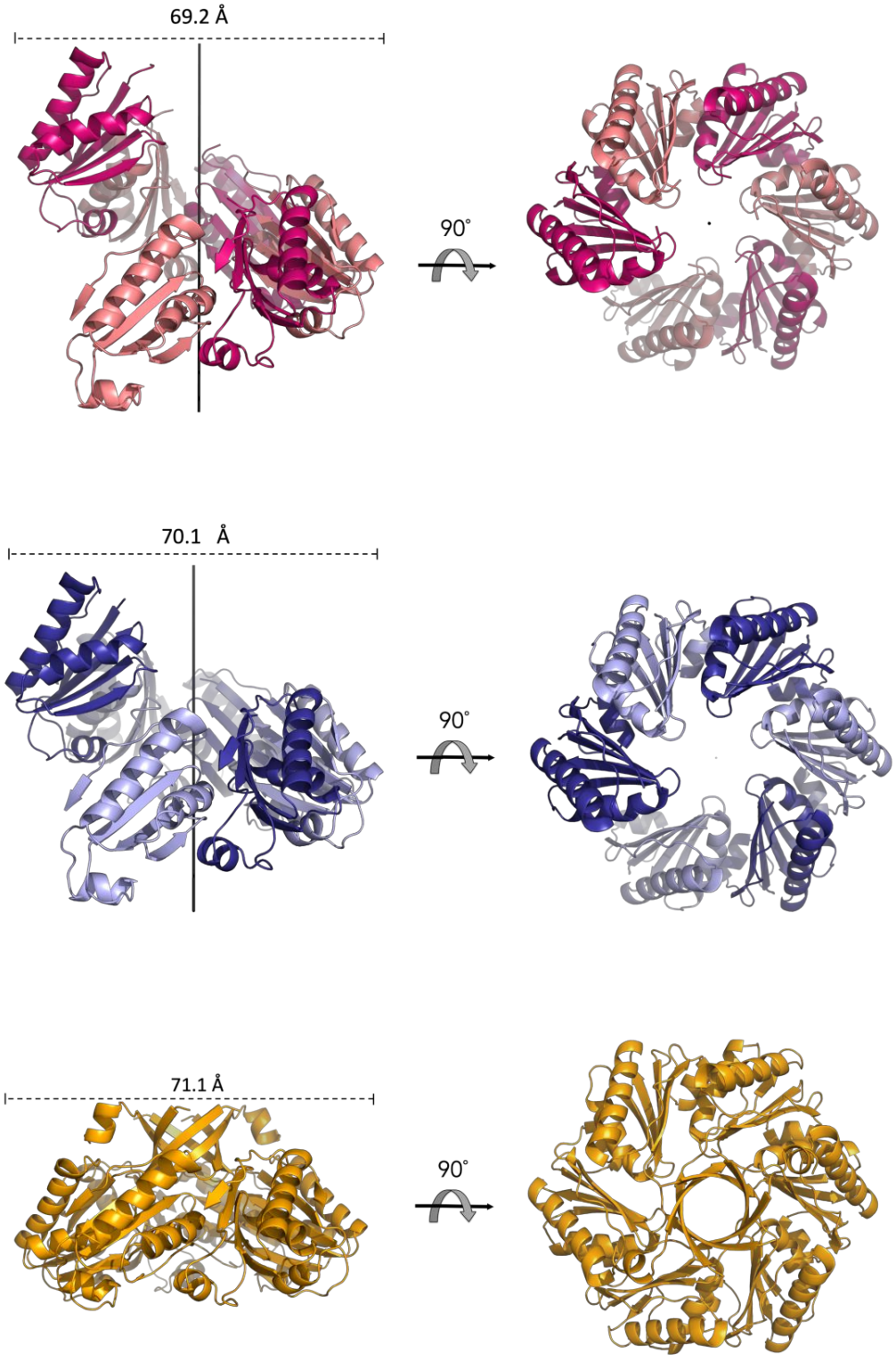
Jessica M. Ochoa, Vy N. Nguyen, Mengxiao Nie, Michael R. Sawaya, Thomas A. Bobik and Todd O. Yeates

Contents:

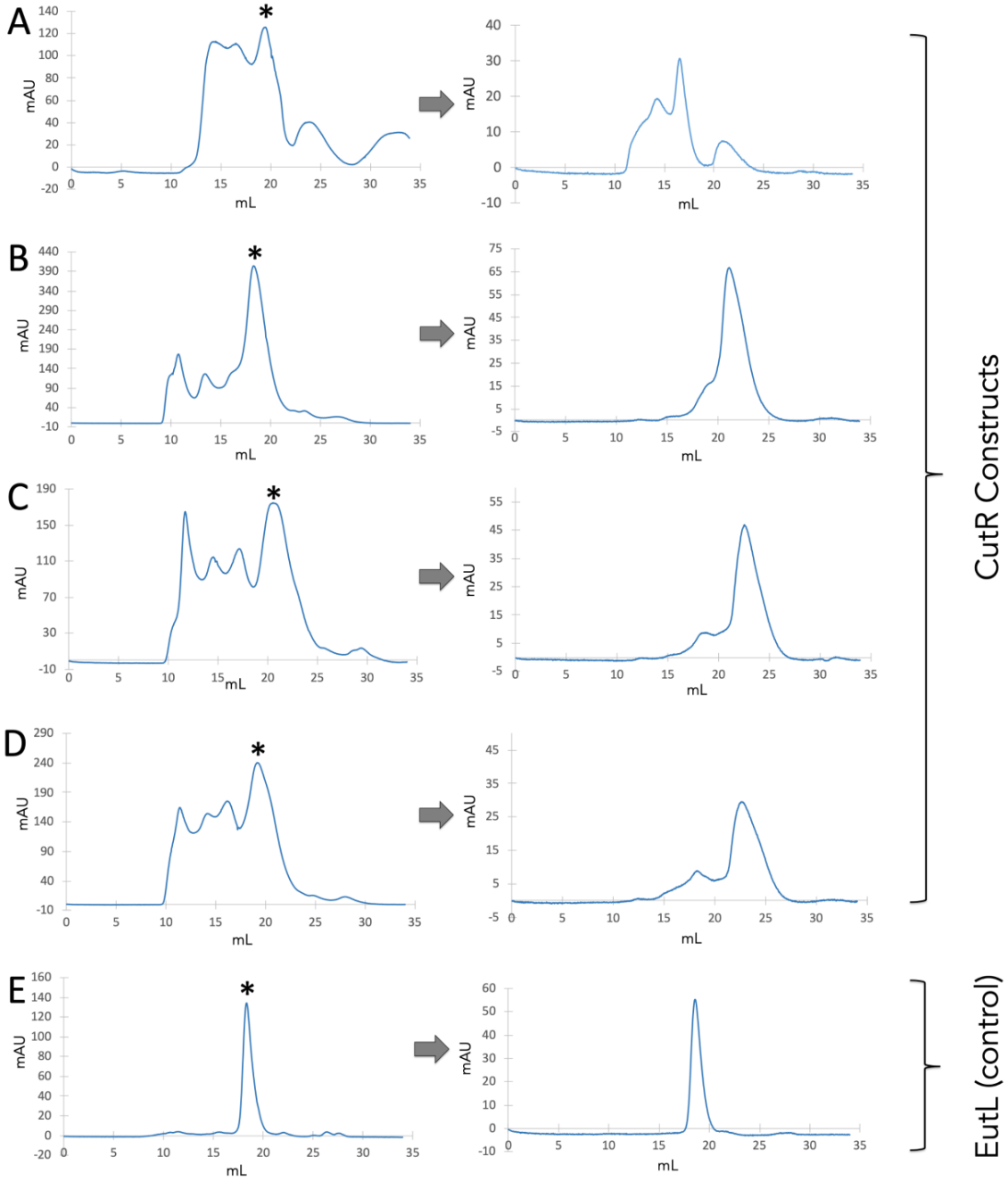
Two supplementary figures

Two supplementary tables

One Supplementary Citation



Supplementary Figure 1. Comparison of the screw and flat polymorphs of CutR. When looking down the six-fold axis of symmetry of Screw 1 (magenta) and Screw 2 (purple), CutR forms a similar architecture to the traditional flat hexamer (orange), possessing similar diameters.



Supplementary Figure 2. Size exclusion profiles and recursive runs for the CutR polymorphs and EutL, a natural tandem BMC-domain construct used as a control. After initial size exclusion chromatography (SEC) (left panels), hexameric peaks (denoted by an asterisk) of CutR_K66D (A), CutR_K66A (B), CutR_C37A (C), CutR_TEV (D) and for EutL (E) were concentrated and subject to an additional round of SEC. SEC profiles of those asterisk-denoted peaks are shown in the corresponding right panels.

Supplementary Table 1. CutR and polymorph sequences

```
CutR                    -------MIEELGKIDRIIQESVPGKQITLAHVIAAPIEAVYECLGVDHEGAIGVVSLTP 53
Hexamer2(CutR_TEV)      ------GMIEELGKIDRIIQESVPGKQITLAHVIAAPIEAVYECLGVDHEGAIGVVSLTP 54
Hexamer1(CutR_C37A)     MHHHHHHMIEELGKIDRIIQESVPGKQITLAHVIAAPIEAVYEALGVDHEGAIGVVSLTP 60
Dimer(CutR_K66D)        MHHHHHHMIEELGKIDRIIQESVPGKQITLAHVIAAPIEAVYECLGVDHEGAIGVVSLTP 60
Screw1(CutR_K66D)       MHHHHHHMIEELGKIDRIIQESVPGKQITLAHVIAAPIEAVYECLGVDHEGAIGVVSLTP 60
Screw2(CutR_K66A)       MHHHHHHMIEELGKIDRIIQESVPGKQITLAHVIAAPIEAVYECLGVDHEGAIGVVSLTP 60
                               **********************************.***************

CutR                    NETAIIAADIAGKAANIDICFVDRFTGSVMFSGDIQSVETSLEDILEYFKNSLGFSTVPL 113
Hexamer2(CutR_TEV)      NETAIIAADIAGAAANIDICFVDRFTGSVMFSGDIQSVETSLEDILEYFKNSLGFSTVPL 114
Hexamer1(CutR_C37A)     NETAIIAADIAGAAANIDICFVDRFTGSVMFSGDIQSVETSLEDILEYFKNSLGFSTVPL 120
Dimer(CutR_K66D)        NETAIIAADIAGDAANIDICFVDRFTGSVMFSGDIQSVETSLEDILEYFKNSLGFSTVPL 120
Screw1(CutR_K66D)       NETAIIAADIAGDAANIDICFVDRFTGSVMFSGDIQSVETSLEDILEYFKNSLGFSTVPL 120
Screw2(CutR_K66A)       NETAIIAADIAGAAANIDICFVDRFTGSVMFSGDIQSVETSLEDILEYFKNSLGFSTVPL 120
                        ************ ********************************************

CutR                    TKS 116
Hexamer2(CutR_TEV)      TKS 117
Hexamer1(CutR_C37A)     TKS 123
Dimer(CutR_K66D)        TKS 123
Screw1(CutR_K66D)       TKS 123
Screw2(CutR_K66A)       TKS 123
                        ***
```

Supplementary Table 2. X-ray Diffraction and Atomic Refinement Statistics

PDB ID	6XPH	6XPI	6XPJ	6XPK	6XPL
Paper name	Dimer	Hexamer 1	Hexamer 2	Screw 1	Screw 2
Data collection					
Space group	$P4_332$	C2	$P4_22_12$	$P6_1$	$P6_1$
Cell dimensions					
a, *b*, *c* (Å)	109.66, 109.66, 109.66	135.36, 76.14, 67.81	79.29, 79.29, 100.79	61.83, 61.83, 41.93	64.91, 64.91, 33.78
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 119.71, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00
Resolution (Å)	77.54-1.80 (1.85-1.80)	63.91-2.60 (2.67-2.60)	62.32-1.50 (1.54-1.50)	53.55-2.80 (2.87-2.80)	56.21-3.30 (3.50-3.30)
R_{merge}	0.079 (1.39)	0.073 (1.44)	0.059 (0.855)	0.113 (0.954)	0.234 (1.34)
I / σ (I)	49.24 (4.71)	10.31 (1.04)	23.7 (3.08)	15.46 (2.22)	4.63 (0.98)
$CC_{1/2}$	100.0 (94.9)	99.9 (74.7)	99.9 (86.0)	99.9 (72.2)	99.2 (51.1)
Completeness (%)	100.0 (100.0)	92.3 (83.1)	99.9 (99.3)	99.4 (95.8)	99.3 (99.5)
Redundancy	75.7 (75.2)	4.02 (4.14)	13.0 (12.7)	9.51 (9.38)	4.68 (4.54)
Refinement					
Resolution (Å)	1.80	2.60	1.50	2.80	3.30
No. reflections	19336 (1386)	17276 (311)	46821 (3373)	2074 (145)	1129 (79)
R_{work} / R_{free}	0.176/0.210 (0.237/0.273)	0.201/0.242 (0.286/0.492)	0.161/0.184 (0.207/0.245)	0.231/0.267 (0.293/0.284)	0.219/0.265 (0.334/0.296)
Molecules per asymmetric unit	2	6	3	1	1
No. atoms					
Protein	1506	4808	2536	734	720
Ligand/ion	21	0	20	0	0
Water	128	0	201	2	0
B-factors					
Protein	35.9	105.5	21.8	67.4	134.5
Ligand/ion	56.1	-	58.0	-	-
Water	41.5	-	29.8	77.5	-
R.m.s. deviations					
Bond lengths (Å)	0.014	0.008	0.010	0.002	0.005
Bond angles (°)	1.8	1.1	1.6	1.2	1.3
Ramachandran statistics (%)					
Most favorable	100	98.3	99.7	93.9	96.9
Allowed	0	1.4	0.3	6.1	3.1
Outliers	0	0.3	0	0	0

Data for each structure were collected from a single crystal. *Values in parentheses are for the highest-resolution shell.

Supplementary citation

S1. Sievers F, Higgins DG (2014) Clustal Omega. Current Protocols in Bioinformatics 48:3.13.1-3.13.16.

Chapter 5: MCPdb: The bacterial microcompartment database

PLOS ONE

RESEARCH ARTICLE

MCPdb: The bacterial microcompartment database

Jessica M. Ochoa[1], Kaylie Bair[2], Thomas Holton[3], Thomas A. Bobik[4], Todd O. Yeates[1,2,3]*

1 UCLA Molecular Biology Institute, University of California Los Angeles, Los Angeles, California, United States of America, **2** UCLA-DOE Institute for Genomics and Proteomics, University of California Los Angeles, Los Angeles, California, United States of America, **3** UCLA Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, California, United States of America, **4** Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, Iowa, United States of America

* yeates@mbi.ucla.edu

Citation: Ochoa JM, Bair K, Holton T, Bobik TA, Yeates TO (2021) MCPdb: The bacterial microcompartment database. PLoS ONE 16(3): e0248269. https://doi.org/10.1371/journal.pone.0248269

Editor: Yanbin Yin, University of Nebraska-Lincoln, UNITED STATES

Received: December 9, 2020

Accepted: February 24, 2021

Published: March 29, 2021

Data Availability Statement: The data are provided in a public repository at https://mcpdb.mbi.ucla.edu/.

Funding: T.O.Y and T.A.B were supported by NIH/NIAID Grant R01AI081146-08, from the National Institutes of Health (https://www.nih.gov/) J.M.O. was supported by a Gilliam Fellowship from the Howard Hughes Medical Institute (https://www.hhmi.org/science-education/programs/gilliam-fellowships-advanced-study) These funders did not play any role in the study.

Abstract

Bacterial microcompartments are organelle-like structures composed entirely of proteins. They have evolved to carry out several distinct and specialized metabolic functions in a wide variety of bacteria. Their outer shell is constructed from thousands of tessellating protein subunits, encapsulating enzymes that carry out the internal metabolic reactions. The shell proteins are varied, with single, tandem and permuted versions of the PF00936 protein family domain comprising the primary structural component of their polyhedral architecture, which is reminiscent of a viral capsid. While considerable amounts of structural and biophysical data have been generated in the last 15 years, the existing functionalities of current resources have limited our ability to rapidly understand the functional and structural properties of microcompartments (MCPs) and their diversity. In order to make the remarkable structural features of bacterial microcompartments accessible to a broad community of scientists and non-specialists, we developed MCPdb: The Bacterial Microcompartment Database (https://mcpdb.mbi.ucla.edu/). MCPdb is a comprehensive resource that categorizes and organizes known microcompartment protein structures and their larger assemblies. To emphasize the critical roles symmetric assembly and architecture play in microcompartment function, each structure in the MCPdb is validated and annotated with respect to: (1) its predicted natural assembly state (2) tertiary structure and topology and (3) the metabolic compartment type from which it derives. The current database includes 163 structures and is available to the public with the anticipation that it will serve as a growing resource for scientists interested in understanding protein-based metabolic organelles in bacteria.

Introduction

Bacterial microcompartments (MCPs or alternatively BMCs), are supramolecular structures found in approximately 20% of bacteria across numerous phyla [1, 2]. These giant protein-based structures have evolved to serve organelle-like functions, with different MCP types

Competing interests: No authors have competing interests.

encapsulating distinct enzymes in order to carry out specific metabolic processes in a sequestered environment within the cell interior [3–6]. MCPs are known to carry out diverse metabolic processes; their unifying functional feature is that they provide a mechanism for bacteria to perform certain multistep reactions in a way that retains metabolic intermediates inside the MCP. The co-localization of sequentially acting enzymes housed inside the MCP helps optimize metabolic flux while limiting alternative side reactions. Importantly, MCPs help prevent the efflux of toxic and/or volatile intermediates into the cytosol [3, 7, 8]. Bacterial microcompartments can be broadly classified into two major categories: carboxysomes and metabolosomes. Carboxysomes are the founding members of the MCPs. They enhance CO_2 fixation in bacteria by encapsulating two sequentially acting enzymes–carbonic anhydrase and ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) [4, 9, 10]. Bicarbonate (in addition to ribulose-bisphosphate) is the substrate that enters the carboxysome via diffusion across the shell; CO_2 is the key intermediate, which is produced by carbonic anhydrase and must be consumed by RuBisCO prior to escape. By contrast, metabolosomes use an assortment of key enzymes to metabolize a variety of substrates including 1,2-propanediol for the propanediol utilization (PDU) MCP and ethanolamine for the ethanolamine utilization (EUT) [4, 7, 8, 10]. Other microcompartments utilize glycyl-radical chemistry (GRM MCPs) and can be further divided into subclasses based on their substrates and signature enzymes, including the glycyl-radical propanediol (Grp) MCP, the choline utilization (Cut) MCP and an additional GRM type that utilizes fucose and rhamnose [11–15]. Lastly, there are MCPs that have been more recently discovered whose metabolic functions are still emerging, including the RMM/Aaum MCP and the Etu MCP. Several recent structures of both BMC and BMV (bacterial microcompartment vertex) proteins have been determined for an MCP first called RMM (for *Rhodococcus and Mycolicibacterium Microcompartment)* and then renamed Aaum (for its apparent role in amino acetone utilization [1, 5, 14, 16, 17]. Additionally, the Etu MCP, or the ethanol utilization microcompartment, has been observed in *Clostridium kluyveri* and has had one of its shell proteins characterized [18, 19].

Despite their functional diversity, bacterial microcompartments are now understood to be structurally similar. Constructed entirely of proteins, the outer microcompartment shell is composed of thousands of homologous tessellating shell proteins belonging to the BMC protein family [20–22], whose structures were first elucidated in 2005 [23, 24]. The canonical BMC protein domain (Fig 1) oligomerizes to form hexameric disks with central pores for the (presumably) diffusive influx of metabolic substrates and the efflux of products. The hexameric disks pack laterally to form the nearly flat facets of the intact shell, while pentameric BMV proteins form the vertices of these large, polyhedral structures (Fig 1) [15, 20]. Any single microcompartment type is composed of multiple paralogs of the BMC protein, with different paralogs offering distinct structural properties. This roughly 100-amino acid domain (Pfam PF00936) remains the primary key for exploring and discovering new types of microcompartments, and has been extensively studied and characterized [15, 21, 22, 25–32]. Structural studies have revealed major topologically distinct variations of the BMC protein domain. The canonical form is the BMC-H shell protein; it is the most abundant, contains a single BMC domain and forms a cyclic homohexamer (Fig 2A) [23]. An alternate topological form of lesser-understood function occurs in the form of permuted BMC proteins [29]. These contain a single, essentially intact BMC domain with a circular permutation. This circular permutation results in a reordering of the amino acid sequence but a similar overall BMC protein fold (Fig 2B), with some of these structures revealing a high degree of flexibility and symmetry-breaking [28, 31]. The BMC-T (*T* stands for tandem) category of proteins consists of two tandem repeats of the BMC domain. BMC-Ts are cyclic trimers that form pseudohexamers (Fig 2C) whose overall shape closely resemble a canonical BMC hexamer [28, 33, 34]. Further variations

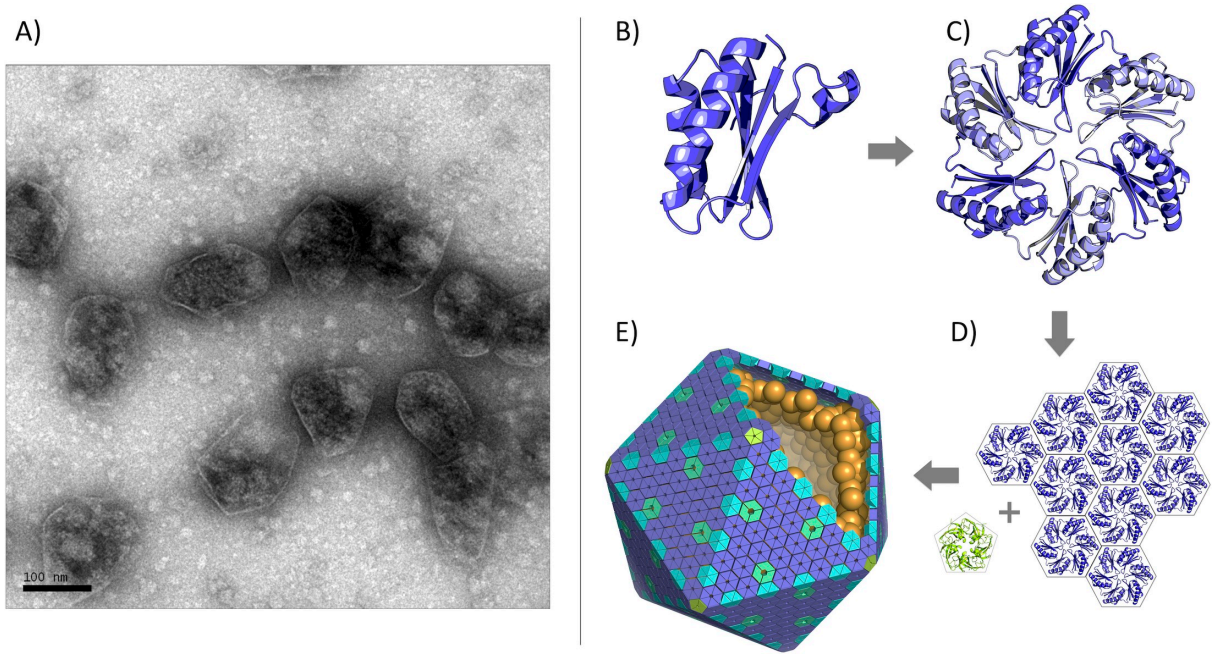


Fig 1. Bacterial microcompartments (MCPs) are large proteinaceous assemblies that function as metabolic organelles. (A) Negative stain electron micrograph of purified Pdu MCPs (scale bar: 50 nm). MCP shells are assembled primarily from proteins belonging to the BMC family (B), which are hexameric or trimeric pseudohexamers (C). (D) Hexameric and pseudohexameric BMC shell proteins pack laterally to form the facets while pentameric BMV proteins (lime green) of unrelated structure form the vertices (D). (E) An idealized model of a microcompartment with external shell proteins and encapsulated enzymes. Most natural MCP shells are not as geometrically regular as depicted here by the icosahedral architecture.

https://doi.org/10.1371/journal.pone.0248269.g001

exist within the BMC-T type, with some also exhibiting circular permutations. In some cases, BMC-T shell proteins have been shown to undergo large conformational changes between closed and open pore states, with critical implications for regulated transport [28, 33–36]. Moreover, some BMC-Ts and even some BMC-H shell proteins have been found to bind iron-sulfur clusters in their central pores [12, 30, 37]. Finally, BMV proteins (sometimes referred to as BMC-P) are cyclic homopentamers that form the vertices of bacterial microcompartments (Fig 2D) [15, 20, 38, 39]. These are based on the Pfam03319 protein domain, which is entirely unrelated in sequence and structure from the BMC protein domain family. The sophisticated mechanistic features of MCPs emphasize their qualification as true organelles in bacteria, built from proteins rather than a lipid bilayer.

Notwithstanding their wide distribution and the extensive investigation into their outer shells, microcompartments remain only partially understood. To date, more than 150 bacterial microcompartment-related structures have been characterized and deposited in the Protein Data Bank (PDB) (Fig 3). Various items of information about each structure–organism, amino acid sequence, functional name, etc.–are generally available, but other critical insights about structure and function are difficult to sort out from the raw data as it is typically presented, and this challenge is especially true for non-experts that have minimal familiarity with the PDB protein structure database. Because understanding quaternary structure–i.e. protein assembly states–is especially critical to understanding elements of MCP function, we viewed the challenges associated with identifying natural assembly forms as a major barrier for novices trying to generate and understand the natural biological forms of MCP shell proteins. We have

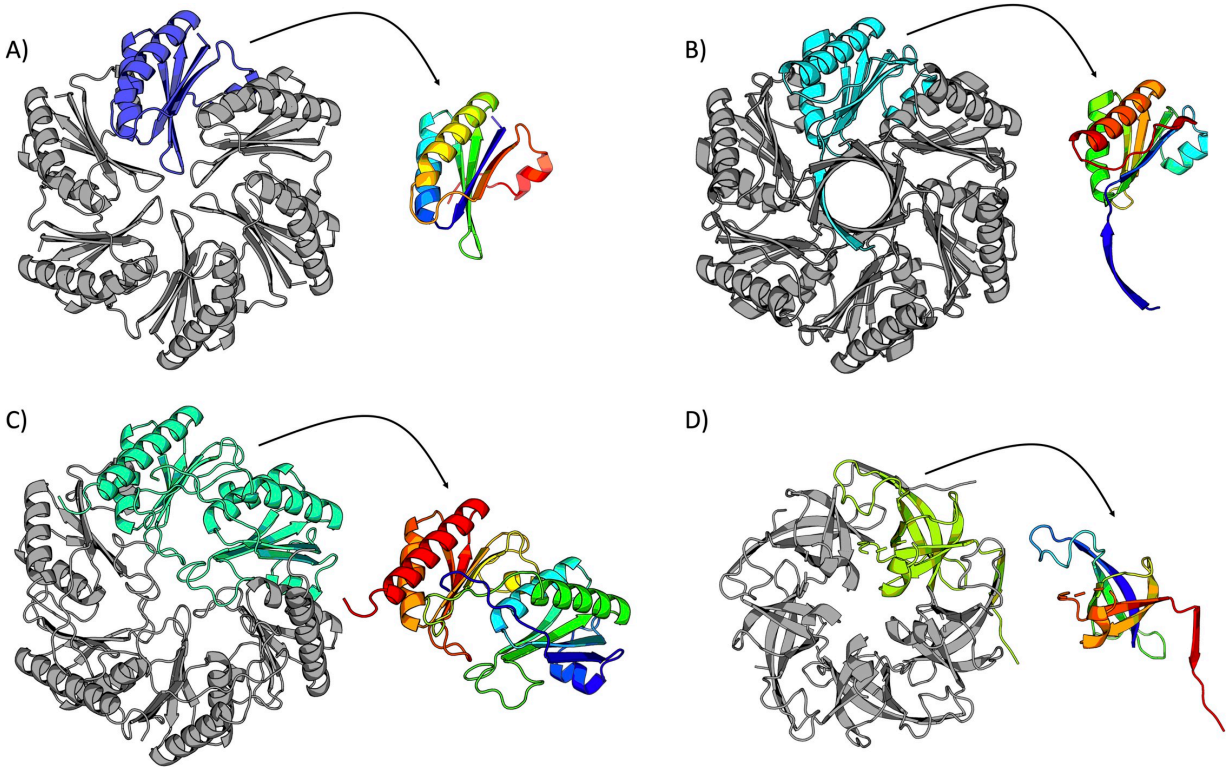


Fig 2. Cartoon representations of four bacterial microcompartment shell proteins. A single monomer is highlighted and presented in the context of the biological assembly, with a color-ramped (blue = N-terminus; red = C-terminus) version of the monomer adjacent to each structure. (A) A representative hexameric BMC shell protein (BMC-H) (PDB 2EWH) [24]. (B) A representative permuted BMC shell protein (PDB 6XPI) [31]. (C) A representative trimeric BMC shell protein (BMC-T) (PDB 3I82) [28]. (D) A representative BMV shell protein (PDB 4I7A) [15].

https://doi.org/10.1371/journal.pone.0248269.g002

also addressed MCP-specific aspects of form and function that are not easily discerned from raw structure files. Discrimination of diverse topological forms of BMC proteins is also provided. This is usually non-obvious from raw structural data files, and these structural variations often relate to important properties of the pores (e.g. 'open' or 'closed'), which are routes for metabolite transport.

A distinct class of prokaryotic nanocompartments, known as encapsulins, has also come under recent investigation. Like MCPs, encapsulins are protein-based compartments from diverse prokaryotes that facilitate compartmentalization and cellular organization [40, 41]. They are icosahedral shells between 25–42 nm in diameter and capable of encapsulating one or more cargo proteins [40–42]. Encapsulin proteins are distinct from the BMC and BMV proteins of MCPs; structural similarity indicates that the encapsulin protein shares a common ancestor with the capsid proteins from the HK97 family of viruses [41]. Encapsulin shells generally require only a single protomer, which self-assembles to form the outer shell [41, 43, 44]. A growing body of research has demonstrated that encapsulins are capable of mitigating oxidative stress and functioning as iron storage containers [41, 43–46].

A growing appreciation of the uniqueness and biological importance of MCPs and other nanocompartments, an expanding body of data on their shell proteins, and current paucity of

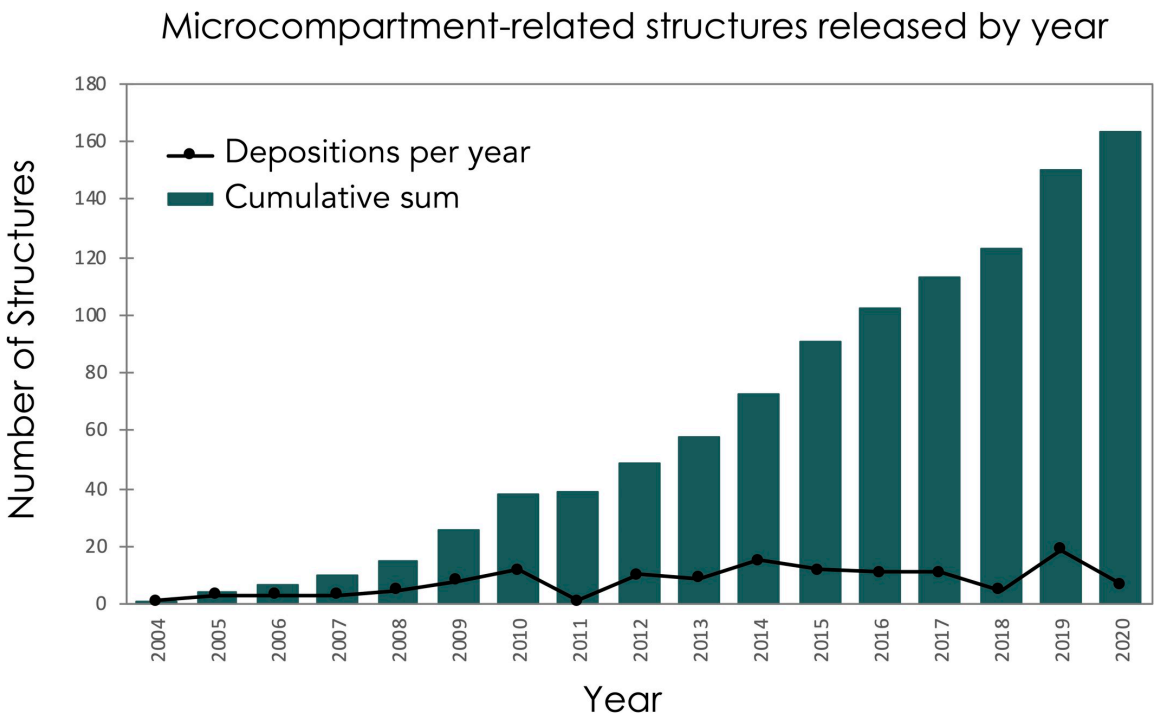


Fig 3. Growth over time of known microcompartment-related structures. There are currently 163 microcompartment and encapsulin-related protein structures deposited in the PDB. Structures were identified by using "microcompartment," "carboxysome" and "encapsulin" as search terms in the PDB. The amino acid sequence of a few representative BMCs and BMVs were also used to ensure we identified all microcompartment shell proteins that have been deposited in the PDB.

https://doi.org/10.1371/journal.pone.0248269.g003

systematic annotation, motivated the development of a centralized database to address these knowledge gaps. Making bacterial microcompartments more accessible to not only structural experts but to a broader scientific audience should help advance this growing field of biology. Here, we describe the development of a novel database, MCPdb: The Bacterial Microcompartment Database (https://mcpdb.mbi.ucla.edu/). While metabolic compartments based on the common BMC protein architecture are the main focus of this database, we also make connections to other systems by including structural information on encapsulins. We collected all known bacterial microcompartment protein structures and assembled a novel online tool that provides users with simplified searching capabilities, structural and biophysical annotations and multiple visualization avenues for examining microcompartment biological assemblies. Most importantly, all structures in MCPdb have been validated–that is to say, quaternary structures have been manually confirmed using human-expertise-based curation.

Materials and methods

Data collection and curation

MCPdb is built by extracting relevant data from the Protein Data Bank [47] and UniProt [48]. We compiled a list of 163 bacterial microcompartment and encapsulin-related structures. A preliminary list of relevant structures was obtained using keyword searches through the PDB web server (https://www.rcsb.org/). An initial search of the term *microcompartment* yielded

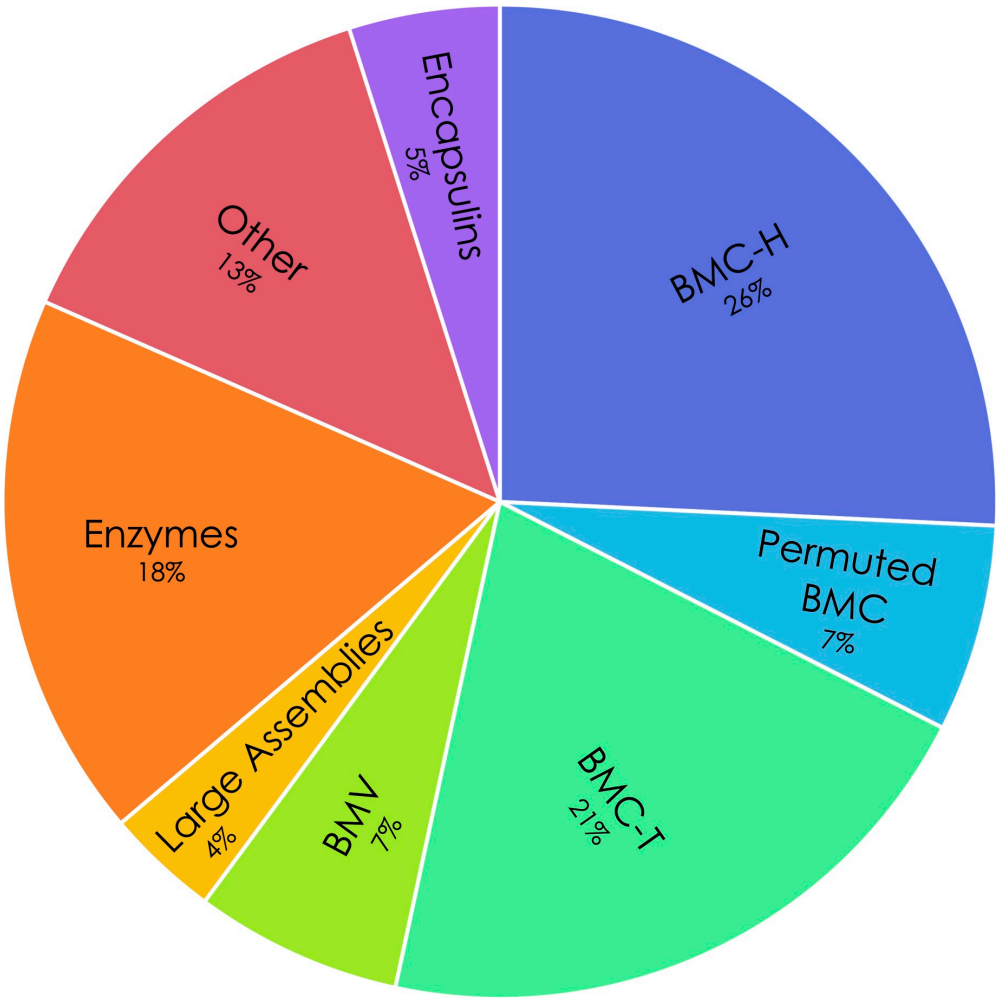


Fig 4. Distribution of protein structure types in the MCPdb. More than 60% of all structures are microcompartment BMC shell proteins (BMC-H, permuted BMC, BMC-T) or pentamers (BMV), with larger icosahedral assemblies comprising 4%, internal enzymes comprising 18% and other microcompartment associated proteins comprising 13%. Encapsulin structures make up the remaining 5% of the MCPdb.

https://doi.org/10.1371/journal.pone.0248269.g004

112 structures that required manual validation and verification, resulting in a total of 98 structures related to MCPs. In order to curate a more comprehensive list, we performed searches of structures using the amino acid sequences of representative BMC-H, BMC-T, BMV and permuted BMC structures. With the addition of several structures from the unrelated encapsulins, MCPdb presently consists of 163 structures (Fig 4). We performed an HTML-based query to collect relevant information including structure resolution, deposition authors and citations (Fig 5). After generating a master list of PDB IDs, we curated their corresponding amino acid sequences obtained from UniProt. A total of 91 unique UniProt IDs gives rise to the 163 separate PDB structures.

Upon collecting all relevant data from the PDB and UniProt, we assigned a series of classifications and annotations to each structure. While individual PDB IDs were used as a key for pertinent structural information, the UniProt IDs were used to provide additional protein details (Fig 5). Each structure in the database has been assigned an *MCP Type*, *MCP Classification*, *Protein Type and Topology*, and *Observed Assembly Form* (Fig 6). *MCP Type* broadly categorizes each structure as a carboxysome, a metabolosome or an encapsulin, while *MCP Classification* provides more details about the microcompartment based on its metabolic function, distinguishing between alpha/beta carboxysomes, and the different metabolosome types including the propanediol utilization MCP, ethanolamine utilization MCP and others. We likewise categorized each structure by intrinsic characteristics including *Protein Type and Topology* and *Observed Assembly Form*. While the *Protein Type and Topology* are inherent and describe the type of protein for a given structure (*i.e.* BMC-H, BMC-T, BMV, etc.), the *Observed Assembly Form* describes the experimental crystal packing (in some cases) and presumptive quaternary architectures.

SQL tables were created to link PDB IDs, UniProt IDs and annotations. In order to construct our database, we utilized a Linux server running Ubuntu 20.04 LTS and MySQL version 5.7. CSV files of the PDB data, UniProt data and annotations were converted into SQL tables with the construction of a linker table to join the tables in the query and to establish the one-to-many relationships between PDB IDs and UniProt IDs (Fig 7). One UniProt can be associated with numerous PDBs (i.e. if the same protein has been structurally characterized in the context of multiple experiments) and one PDB can be associated with numerous UniProts (i.e. if the structure characterized is comprised of proteins of more than one identity). We then generated a series of PHP scripts to query the data and populate our website content. Structures on MCPdb are organized and called by their four-character PDB ID.

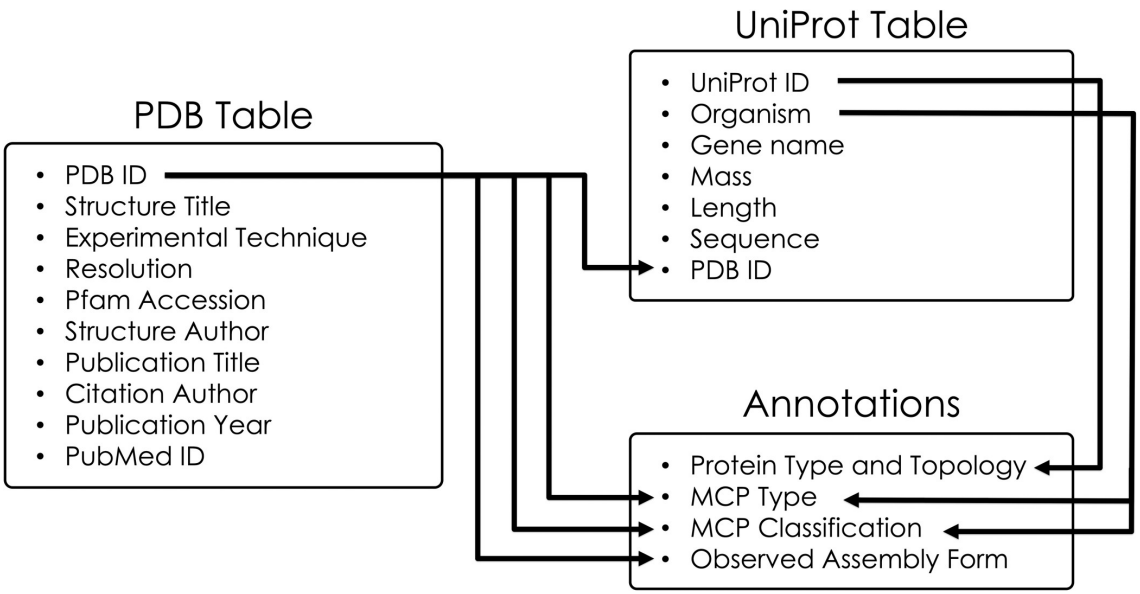


Fig 5. Data sources and annotations for entries in the MCPdb. Key structural information from the PDB as well as associated protein information from UniProt are used to describe each entry. The PDB is the primary link to UniProt IDs. The PDB data file provides information about the Observed Assembly Form for the protein, and UniProt provides information from which the protein topology (e.g. circular permutations and domain duplications) can be discerned. These data sources and the literature are used to annotate the MCP functional type and subclassification.

https://doi.org/10.1371/journal.pone.0248269.g005

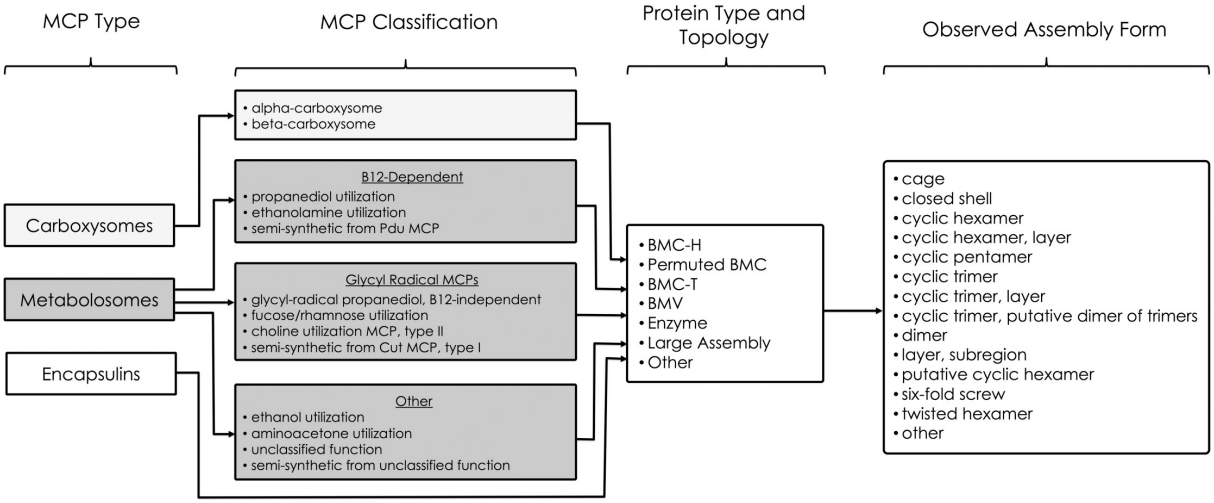


Fig 6. MCPdb entry annotations. MCP Type indicates the broad metabolic category. MCP Classification further distinguishes between the different metabolic subtypes of carboxysomes and metabolosomes. Protein Type and Topology describe properties inherent to the protein tertiary structure. Lastly, Observed Assembly Form describes protein quaternary characteristics of the experimentally described structure.

https://doi.org/10.1371/journal.pone.0248269.g006

File curation and preparation

In order to construct a centralized microcompartment database, we extracted and compiled relevant files including PDBs, biological assemblies, and FASTA amino acid sequence files

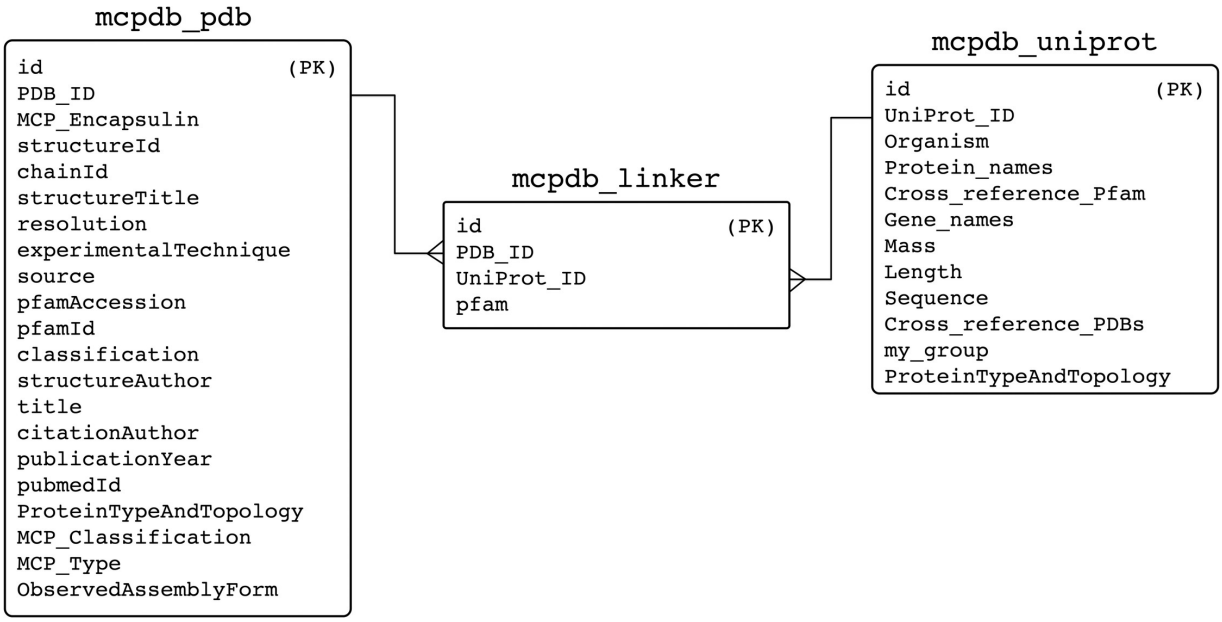


Fig 7. Entity relationship diagram of the MCPdb as a MySQL database. Boxes show the primary data sources and the linker table used to join the tables in the queries. Primary keys (PK) have been identified.

https://doi.org/10.1371/journal.pone.0248269.g007

with the goal of providing these files to the end user. We also sought to provide users with numerous modes of interacting with each structure. To appeal to experts and novices alike, we incorporated: (1) an interactive 3D viewer for rapid structure interrogation, (2) ready-to-use PyMOL graphics sessions for streamlined figure preparation, and (3) images for quickly viewing and interpreting structures while browsing the database.

All files are housed on our permanent institutional web server using the PDB ID as the primary identifier. With the master list of 163 PDB IDs, we utilized a wget command to pull atomic coordinates of all structures in the form of.pdb and.cif files onto our Linux server, which are available to our users as downloads. We were also able to retrieve nearly 60% of the correctly named and trimmed biological assemblies using the program PISA [49]. The biological assemblies that were generated by PISA and migrated to our server were validated for accuracy. The remaining structures whose biological assemblies could not be successfully generated with PISA required manual intervention; the need for this step highlights one of the key utilities of the database. Because a structure file may contain multiple sets of coordinates for the same set of atoms (that are distinguished by unique models), we used PyMOL to create new.pdb files in which we assigned a unique chain ID to each chain so that these biological assemblies can be easily loaded, free of multiple objects and multiple states. These validated biological assemblies have been cleaned to exclude most small molecules judged to not reflect biological function (e.g. crystallization buffer molecules, etc.). In a few select cases, the natural biological assembly form of a BMC protein remains uncertain (some BMC-T trimers tend to occur in structural studies in the form of two stacked disks). In those cases, users can access alternate assembly forms. MCPdb also provides relevant sequence information in the form of. fasta and.txt files. FASTA-formatted sequences for each structure were retrieved from the PDB; these reflect the actual sequence of the experimentally characterized protein, which can include mutations and the addition of protein purification tags. The native, unmodified protein amino acid sequences (.txt) are extracted from the UniProt data using a PHP query. These files are also available as downloads (Fig 8).

We incorporated an interactive 3D viewer that enables users to dynamically engage with most of the MCPdb structures without the need to download additional molecular visualization software (Fig 8). The mutation position imaging toolbox (MuPIT) is a browser-based visualization application originally designed for novice structure investigators [50]. By integrating this unmodified software into our database, we provide users the opportunity to quickly visualize structures of interests on desktop and mobile-based browsers. Users may view ribbon, line and stick models of each structure. About 8% of structures were too large for the interactive viewer (some contain as many as 540 protein chains), in which cases the server offers movies (created in PyMOL) that dynamically change views and toggle through ribbon and surface renderings of the structure of interest. The short movies of these structures are played in the browser and can also be downloaded and saved locally.

Additionally, we provide ready-to-use PyMOL session files (.pse) as optional downloads. These are functional even for the largest of the structures. After curating and manually validating our library of biological assemblies, we prepared a series of PyMOL sessions (Fig 8). For each structure, we provide users with a cartoon and surface representation of each structure. Structures are colored such that users can rapidly distinguish between multiple polypeptide chains. Surface representations are semi-transparent for easy viewing. We have also pre-loaded short movies so that users are immediately presented with a rotating view of the selected structure upon launching the PyMOL session. Rendering surface representations of large structures, including cages and closed shells, is computationally taxing and can crash PyMOL under some computer user configurations. To overcome these challenges, we employed various lesser-known PyMOL strategies. By reducing the surface quality and altering the Gaussian

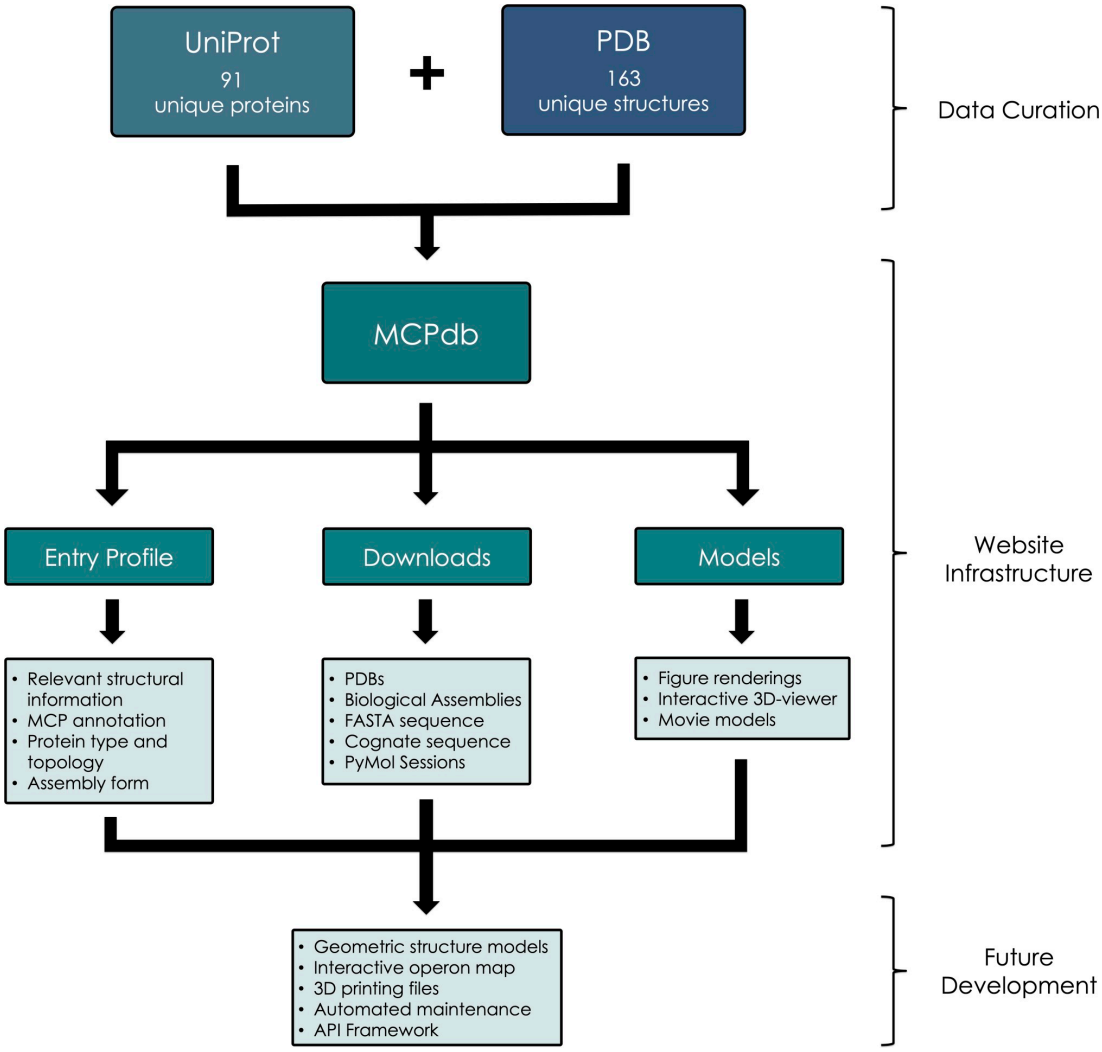


Fig 8. Flow chart depicting data curation and generation, website infrastructure and goals for future development.

https://doi.org/10.1371/journal.pone.0248269.g008

resolution option in the Fourier filtering representation prior to generating isosurface maps, we were able to create surface representations, even for the largest structures, which are visually informative while requiring significantly reduced computing power. Our uniform PyMOL sessions create an effortless way for novice PyMOL users to interact with each of the 163 structures and a simple way of preparing accurate and illustrative figures. Lastly, we generated a series of three figure-ready images (.pngs) for users to scroll through as they are browsing a structure on MCPdb (Fig 8). Based on specifically crafted PyMOL sessions, we exported a series of views as.pngs and added these as sliders to each entry in the MCPdb. We additionally created and included N to C-terminus color-ramped cartoon diagrams of the asymmetric unit of each structure in the image slider.

Website construction

Following database curation, we generated a user-friendly browser interface. The MCPdb infrastructure was created using WordPress, HTML, CSS and JavaScript. We utilized the WordPress graphical user interface (GUI) to build the landing page and accessory information pages. We used HTML, CSS and JavaScript to generate a template page that displays select information for each structure. We replicated and auto-populated data fields in this template for each of the 163 structures using a PHP script. We also created a series of queries to provide our users with seamless and intuitive search features. Infrastructure for simple searches based on key words and more complex filtering searches were also created using PHP.

Results and discussion

Database description

MCPdb is available at https://mcpdb.mbi.ucla.edu, a permanent institutional URL managed by the UCLA-DOE Institute for Genomics and Proteomics. MCPdb was created in order to compile and consolidate structures related to MCPs, and encapsulin structures that are known at this time. More importantly, MCPdb was designed to provide users with readily available structural and biophysical annotations as well as validated biological assemblies. The current version of our database pulls together 163 structures from the PDB (comprising 91 unique UniProts) and a collection of curated files and utilities that are available for in-browser viewing and download (Fig 8). Downloads include PDB files, biological assembly structures (pdb format), files for biological and experimental protein sequences, and ready-to-use PyMOL session files. Users can view rendered images of each structure or interact with them in 3D within the browser. In alignment with our philosophy of introducing new users to the field, MCPdb is freely available and optimized for accessibility on desktops, tablets and mobile devices (Fig 9).

Web interface

MCPdb provides a simple and interactive framework for users to explore bacterial microcompartments and encapsulins. Upon navigating to the home page, users are presented with a brief database description and provided with links that navigate to a summary page, search page and a quick-start guide. As users explore MCPdb, they are introduced to high-level information about MCPs and their characteristic shell proteins. As they navigate to an individual entry page, users are provided with images of the structure and relevant annotations including *MCP Type*, *MCP Classification*, *Protein Type and Topology* and *Observed Assembly Form* (Fig 9). Users can scroll down for additional information related to the structure and authorship, they can download validated structure files and ready-made PyMOL sessions or they can view the structure in 3D within their browser. We additionally provide a *Get Connected* page to allow users to request assistance and provide feedback.

Comparison to other databases

The MCPdb combines data available from other sources, including the PDB [30] and UniProt [31], with curation and substantial post-processing. The various curation and postprocessing protocols add considerable value compared to currently available data repositories. Presentation of correct biological assembly states is often a challenge for structures obtained by crystallographic methods, and as noted above this is a critical aspect of interrogating MCP structure and function. Vital information, and search capacity, is also provided concerning metabolic function types and unique topological features in the BMC protein family; these features relate to functional differences in their assembly and their roles in molecular transport. There are

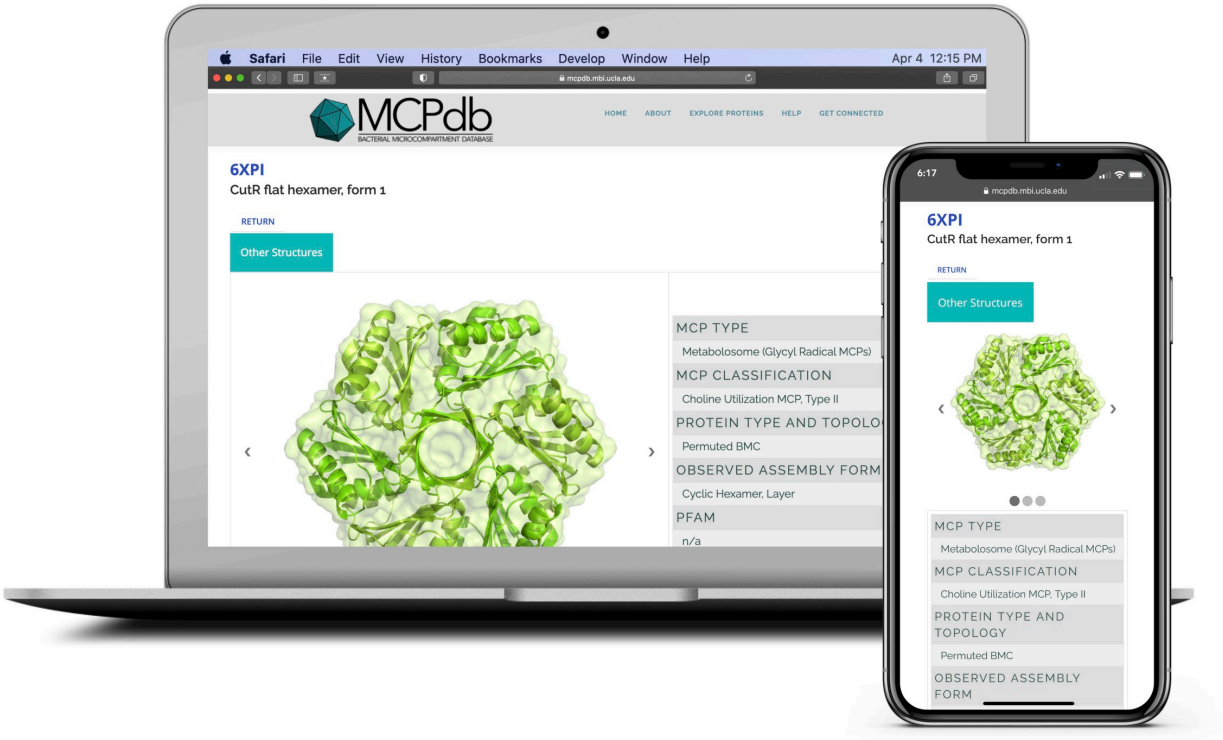


Fig 9. Screenshot and example of a structure profile on the MCPdb interface. MCPdb has been optimized for use on desktops, tablets and mobile devices.

https://doi.org/10.1371/journal.pone.0248269.g009

some parallels between MCPs and viral capsids, and indeed the need for a database that curates biological assembly forms for viral capsids was recognized some years ago with the development of the VIPERdb database [51]. Similarly, other specialized databases that systematically collect, annotate and process structures using expert curation, including KLIFS (Kinase–Ligand Interaction Fingerprints and Structures), have provided researchers with valuable feature-rich resources [52]. The MCPdb answers an analogous need for bacterial microcompartments.

Curation has been applied to remove complicating accessory data (e.g. bound buffer molecules, conflicting polypeptide chain names, etc.), which might otherwise confuse non-expert users. The integration with multiple modes of visualization, tailored where necessary according to size, will facilitate the graphical display and dissemination of information on these special biological systems. Attention has been given to providing simple methods of display to serve the broadest community of users.

Conclusions and future prospects

The MCPdb currently houses 163 microcompartment protein and encapsulin-related structures. Access to validated biological structures as well as structural and biophysical annotations is necessary for well-informed scientific investigation surrounding MCPs. As a relatively new field, the structural biology of MCPs is an area of growing scientific and bioengineering interest [53–64]. Not only a tool for experts in the field, the MCPdb provides novices and young

students the opportunity to learn about and explore bacterial MCPs. The exceptional biological role of MCPs as protein-based organelles makes them an attractive subject for young scientists, as they challenge the textbook paradigm that eukaryotic cells possess mechanistically complex subcellular organelles while bacterial cells do not.

As the body of structural data on MCPs grows, increased automation will be required to keep the database current. Ongoing developments will involve methods to periodically survey the PDB for new microcompartment and encapsulin-related structures, and their associated data files. Additionally, further efforts will expand the types of information and utilities available on the database. Subsequent versions will introduce an interactive operon map for exploring the operon structure and genomic context of BMC shell proteins and their associated encapsulated enzymes. We are also working to provide users with further geometric representations of the structures, electrostatic potentials, pore properties and graphics files for 3D printing, as well as an API framework to extend the functionality of MCPdb for future users. These capabilities will further facilitate access to the field of MCPs for basic and applied research.

Acknowledgments

The authors thank Alex Lisker, Duilio Cascio, Michael Sawaya, and Kyle Meador for help and advice.

Author Contributions

Conceptualization: Jessica M. Ochoa, Todd O. Yeates.

Data curation: Jessica M. Ochoa, Kaylie Bair, Thomas Holton.

Funding acquisition: Thomas A. Bobik.

Methodology: Thomas Holton.

Project administration: Todd O. Yeates.

Software: Jessica M. Ochoa, Kaylie Bair, Thomas Holton.

Validation: Jessica M. Ochoa, Kaylie Bair, Thomas A. Bobik, Todd O. Yeates.

Visualization: Jessica M. Ochoa.

Writing – original draft: Jessica M. Ochoa, Kaylie Bair, Todd O. Yeates.

Writing – review & editing: Jessica M. Ochoa, Thomas A. Bobik, Todd O. Yeates.

References

1. Jorda J, Lopez D, Wheatley NM, Yeates TO. Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. Protein Science. 2013; 22: 179–195. https://doi.org/10.1002/pro.2196 PMID: 23188745
2. AbdulRahman F. The Distribution of Polyhedral Bacterial Microcompartments Suggests Frequent Horizontal Transfer and Operon Reassembly. Journal of Phylogenetics & Evolutionary Biology. 2013;01. https://doi.org/10.4172/2329-9002.1000118
3. Chowdhury C, Sinha S, Chun S, Yeates TO, Bobik TA. Diverse Bacterial Microcompartment Organelles. Microbiol Mol Biol Rev. 2014; 78: 438–468. https://doi.org/10.1128/MMBR.00009-14 PMID: 25184561
4. Kerfeld CA, Heinhorst S, Cannon GC. Bacterial microcompartments. Annu Rev Microbiol. 2010; 64: 391–408. https://doi.org/10.1146/annurev.micro.112408.134211 PMID: 20825353

5. Axen SD, Erbilgin O, Kerfeld CA. A Taxonomy of Bacterial Microcompartment Loci Constructed by a Novel Scoring Method. PLOS Computational Biology. 2014; 10: e1003898. https://doi.org/10.1371/journal.pcbi.1003898 PMID: 25340524

6. Ravcheev DA, Moussu L, Smajic S, Thiele I. Comparative Genomic Analysis Reveals Novel Microcompartment-Associated Metabolic Pathways in the Human Gut Microbiome. Front Genet. 2019;10. https://doi.org/10.3389/fgene.2019.00010 PMID: 30815010

7. Bobik TA, Lehman BP, Yeates TO. Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways. Molecular Microbiology. 2015; 98: 193–207. https://doi.org/10.1111/mmi.13117 PMID: 26148529

8. Kerfeld CA, Erbilgin O. Bacterial microcompartments and the modular construction of microbial metabolism. Trends in Microbiology. 2015; 23: 22–34. https://doi.org/10.1016/j.tim.2014.10.003 PMID: 25455419

9. Cannon GC, Bradburne CE, Aldrich HC, Baker SH, Heinhorst S, Shively JM. Microcompartments in Prokaryotes: Carboxysomes and Related Polyhedra. Appl Environ Microbiol. 2001; 67: 5351–5361. https://doi.org/10.1128/AEM.67.12.5351-5361.2001 PMID: 11722879

10. Yeates TO, Kerfeld CA, Heinhorst S, Cannon GC, Shively JM. Protein-based organelles in bacteria: carboxysomes and related microcompartments. Nature Reviews Microbiology. 2008; 6: 681–691. https://doi.org/10.1038/nrmicro1913 PMID: 18679172

11. Petit E, LaTouf WG, Coppi MV, Warnick TA, Currie D, Romashko I, et al. Involvement of a Bacterial Microcompartment in the Metabolism of Fucose and Rhamnose by Clostridium phytofermentans. PLOS ONE. 2013; 8: e54337. https://doi.org/10.1371/journal.pone.0054337 PMID: 23382892

12. Thompson MC, Wheatley NM, Jorda J, Sawaya MR, Gidaniyan SD, Ahmed H, et al. Identification of a Unique Fe-S Cluster Binding Site in a Glycyl-Radical Type Microcompartment Shell Protein. J Mol Biol. 2014; 426: 3287–3304. https://doi.org/10.1016/j.jmb.2014.07.018 PMID: 25102080

13. Lundin AP, Stewart KL, Stewart AM, Herring TI, Chowdhury C, Bobik TA. Genetic Characterization of a Glycyl Radical Microcompartment Used for 1,2-Propanediol Fermentation by Uropathogenic *Escherichia coli* CFT073. Metcalf WW, editor. Journal of Bacteriology. 2020;202. https://doi.org/10.1128/JB.00017-20 PMID: 32071097

14. Zarzycki J, Sutter M, Cortina NS, Erb TJ, Kerfeld CA. In Vitro Characterization and Concerted Function of Three Core Enzymes of a Glycyl Radical Enzyme—Associated Bacterial Microcompartment. Sci Rep. 2017; 7. https://doi.org/10.1038/srep42757 PMID: 28202954

15. Wheatley NM, Gidaniyan SD, Liu Y, Cascio D, Yeates TO. Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. Protein Sci. 2013; 22: 660–665. https://doi.org/10.1002/pro.2246 PMID: 23456886

16. Mallette E, Kimber MS. Structural and kinetic characterization of (S)-1-amino-2-propanol kinase from the aminoacetone utilization microcompartment of Mycobacterium smegmatis. J Biol Chem. 2018; 293: 19909–19918. https://doi.org/10.1074/jbc.RA118.005485 PMID: 30361441

17. Mallette E, Kimber MS. Structure and Kinetics of the *S*-(+)-1-Amino-2-propanol Dehydrogenase from the RMM Microcompartment of *Mycobacterium smegmatis*. Biochemistry. 2018; 57: 3780–3789. https://doi.org/10.1021/acs.biochem.8b00464 PMID: 29757625

18. Lurz R, Mayer F, Gottschalk G. Electron microscopic study on the quaternary structure of the isolated particulate alcohol-acetaldehyde dehydrogenase complex and on its identity with the polygonal bodies of Clostridium kluyveri. Arch Microbiol. 1979; 120: 255–262. https://doi.org/10.1007/BF00423073

19. Heldt D, Frank S, Seyedarabi A, Ladikis D, Parsons JB, Warren MJ, et al. Structure of a trimeric bacterial microcompartment shell protein, EtuB, associated with ethanol utilization in Clostridium kluyveri. Biochem J. 2009; 423: 199–207. https://doi.org/10.1042/BJ20090780 PMID: 19635047

20. Tanaka S, Kerfeld CA, Sawaya MR, Cai F, Heinhorst S, Cannon GC, et al. Atomic-Level Models of the Bacterial Carboxysome Shell. Science. 2008; 319: 1083–1086. https://doi.org/10.1126/science.1151458 PMID: 18292340

21. Yeates TO, Thompson MC, Bobik TA. The protein shells of bacterial microcompartment organelles. Current Opinion in Structural Biology. 2011; 21: 223–231. https://doi.org/10.1016/j.sbi.2011.01.006 PMID: 21315581

22. Yeates TO, Jorda J, Bobik TA. The Shells of BMC-Type Microcompartment Organelles in Bacteria. Journal of Molecular Microbiology and Biotechnology. 2013; 23: 290–299. https://doi.org/10.1159/000351347 PMID: 23920492

23. Kerfeld CA, Sawaya MR, Tanaka S, Nguyen CV, Phillips M, Beeby M, et al. Protein Structures Forming the Shell of Primitive Bacterial Organelles. Science. 2005; 309: 936–938. https://doi.org/10.1126/science.1113397 PMID: 16081736

43. Giessen TW, Silver PA. Widespread distribution of encapsulin nanocompartments reveals functional diversity. Nature Microbiology. 2017; 2: 1–11. https://doi.org/10.1038/nmicrobiol.2017.29 PMID: 28263314

44. McHugh CA, Fontana J, Nemecek D, Cheng N, Aksyuk AA, Heymann JB, et al. A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. EMBO J. 2014; 33: 1896–1911. https://doi.org/10.15252/embj.201488566 PMID: 25024436

45. Gabashvili AN, Chmelyuk NS, Efremova MV, Malinovskaya JA, Semkina AS, Abakumov MA. Encapsulins—Bacterial Protein Nanocompartments: Structure, Properties, and Application. Biomolecules. 2020; 10. https://doi.org/10.3390/biom10060966 PMID: 32604934

46. Contreras H, Joens MS, McMath LM, Le VP, Tullius MV, Kimmey JM, et al. Characterization of a Mycobacterium tuberculosis Nanocompartment and Its Potential Cargo Proteins. J Biol Chem. 2014; 289: 18279–18289. https://doi.org/10.1074/jbc.M114.570119 PMID: 24855650

47. Burley SK, Berman HM, Kleywegt GJ, Markley JL, Nakamura H, Velankar S. Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. In: Wlodawer A, Dauter Z, Jaskolski M, editors. Protein Crystallography: Methods and Protocols. New York, NY: Springer; 2017. pp. 627–641. https://doi.org/10.1007/978-1-4939-7000-1_26 PMID: 28573592

48. Consortium TU. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019; 47: D506–D515. https://doi.org/10.1093/nar/gky1049 PMID: 30395287

49. Krissinel E, Henrick K. Inference of Macromolecular Assemblies from Crystalline State. Journal of Molecular Biology. 2007; 372: 774–797. https://doi.org/10.1016/j.jmb.2007.05.022 PMID: 17681537

50. Niknafs N, Kim D, Kim R, Diekhans M, Ryan M, Stenson PD, et al. MuPIT interactive: webserver for mapping variant positions to annotated, interactive 3D structures. Hum Genet. 2013; 132: 1235–1243. https://doi.org/10.1007/s00439-013-1325-0 PMID: 23793516

51. Carrillo-Tripp M, Shepherd CM, Borelli IA, Venkataraman S, Lander G, Natarajan P, et al. VIPERdb2: an enhanced and web API enabled relational database for structural virology. Nucleic Acids Research. 2009; 37: D436–D442. https://doi.org/10.1093/nar/gkn840 PMID: 18981051

52. Kanev GK, de Graaf C, Westerman BA, de Esch IJP, Kooistra AJ. KLIFS: an overhaul after the first 5 years of supporting kinase research. Nucleic Acids Research. 2021; 49: D562–D569. https://doi.org/10.1093/nar/gkaa895 PMID: 33084889

53. Parsons JB, Frank S, Bhella D, Liang M, Prentice MB, Mulvihill DP, et al. Synthesis of Empty Bacterial Microcompartments, Directed Organelle Protein Incorporation, and Evidence of Filament-Associated Organelle Movement. Molecular Cell. 2010; 38: 305–315. https://doi.org/10.1016/j.molcel.2010.04.008 PMID: 20417607

54. Lawrence AD, Frank S, Newnham S, Lee MJ, Brown IR, Xue W-F, et al. Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. ACS Synth Biol. 2014; 3: 454–465. https://doi.org/10.1021/sb4001118 PMID: 24933391

55. Huber I, Palmer DJ, Ludwig KN, Brown IR, Warren MJ, Frunzke J. Construction of Recombinant Pdu Metabolosome Shells for Small Molecule Production in Corynebacterium glutamicum. ACS Synth Biol. 2017; 6: 2145–2156. https://doi.org/10.1021/acssynbio.7b00167 PMID: 28826205

56. Wagner HJ, Capitain CC, Richter K, Nessling M, Mampel J. Engineering bacterial microcompartments with heterologous enzyme cargos. Engineering in Life Sciences. 2017; 17: 36–46. https://doi.org/10.1002/elsc.201600107 PMID: 32624727

57. Fang Y, Huang F, Faulkner M, Jiang Q, Dykes GF, Yang M, et al. Engineering and Modulating Functional Cyanobacterial CO2-Fixing Organelles. Front Plant Sci. 2018; 9. https://doi.org/10.3389/fpls.2018.00739 PMID: 29922315

58. Hagen A, Sutter M, Sloan N, Kerfeld CA. Programmed loading and rapid purification of engineered bacterial microcompartment shells. Nature Communications. 2018; 9: 2881. https://doi.org/10.1038/s41467-018-05162-z PMID: 30038362

59. Lee MJ, Mantell J, Brown IR, Fletcher JM, Verkade P, Pickersgill RW, et al. De novo targeting to the cytoplasmic and luminal side of bacterial microcompartments. Nature Communications. 2018; 9: 3413. https://doi.org/10.1038/s41467-018-05922-x PMID: 30143644

60. Huang J, Ferlez BH, Young EJ, Kerfeld CA, Kramer DM, Ducat DC. Functionalization of Bacterial Microcompartment Shell Proteins With Covalently Attached Heme. Front Bioeng Biotechnol. 2020; 7. https://doi.org/10.3389/fbioe.2019.00432 PMID: 31993414

61. Jakobson CM, Chen Y, Slininger MF, Valdivia E, Kim EY, Tullman-Ercek D. Tuning the Catalytic Activity of Subcellular Nanoreactors. Journal of Molecular Biology. 2016; 428: 2989–2996. https://doi.org/10.1016/j.jmb.2016.07.006 PMID: 27427532

24. Tsai Y, Sawaya MR, Cannon GC, Cai F, Williams EB, Heinhorst S, et al. Structural Analysis of CsoS1A and the Protein Shell of the Halothiobacillus neapolitanus Carboxysome. PLoS Biol. 2007; 5. https://doi.org/10.1371/journal.pbio.0050144 PMID: 17518518

25. Yeates TO, Crowley CS, Tanaka S. Bacterial Microcompartment Organelles: Protein Shell Structure and Evolution. Annual Review of Biophysics. 2010; 39: 185–205. https://doi.org/10.1146/annurev.biophys.093008.131418 PMID: 20192762

26. Bobik TA, Xu Y, Jeter RM, Otto KE, Roth JR. Propanediol utilization genes (pdu) of Salmonella typhimurium: three genes for the propanediol dehydratase. J Bacteriol. 1997; 179: 6633–6639. https://doi.org/10.1128/jb.179.21.6633-6639.1997 PMID: 9352910

27. Morgan Beeby, Bobik Thomas A., Yeates Todd O. Exploiting genomic patterns to discover new supramolecular protein assemblies. Protein Science. 2008; 18: 69–79. https://doi.org/10.1002/pro.1 PMID: 19177352

28. Tanaka S, Sawaya MR, Yeates TO. Structure and Mechanisms of a Protein-Based Organelle in Escherichia coli. Science. 2010; 327: 81–84. https://doi.org/10.1126/science.1179513 PMID: 20044574

29. Crowley CS, Sawaya MR, Bobik TA, Yeates TO. Structure of the PduU Shell Protein from the Pdu Microcompartment of Salmonella. Structure. 2008; 16: 1324–1332. https://doi.org/10.1016/j.str.2008.05.013 PMID: 18786396

30. Crowley CS, Cascio D, Sawaya MR, Kopstein JS, Bobik TA, Yeates TO. Structural Insight into the Mechanisms of Transport across the Salmonella enterica Pdu Microcompartment Shell. J Biol Chem. 2010; 285: 37838–37846. https://doi.org/10.1074/jbc.M110.160580 PMID: 20870711

31. Ochoa JM, Nguyen VN, Nie M, Sawaya MR, Bobik TA, Yeates TO. Symmetry breaking and structural polymorphism in a bacterial microcompartment shell protein for choline utilization. Protein Science. 2020; 29: 2201–2212. https://doi.org/10.1002/pro.3941 PMID: 32885887

32. Sutter M, Greber B, Aussignargues C, Kerfeld CA. Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. Science. 2017; 356: 1293–1297. https://doi.org/10.1126/science.aan3289 PMID: 28642439

33. Cai F, Sutter M, Cameron JC, Stanley DN, Kinney JN, Kerfeld CA. The Structure of CcmP, a Tandem Bacterial Microcompartment Domain Protein from the β-Carboxysome, Forms a Subcompartment Within a Microcompartment. J Biol Chem. 2013; 288: 16055–16063. https://doi.org/10.1074/jbc.M113.456897 PMID: 23572529

34. Sagermann M, Ohtaki A, Nikolakakis K. Crystal structure of the EutL shell protein of the ethanolamine ammonia lyase microcompartment. Proc Natl Acad Sci U S A. 2009; 106: 8883–8887. https://doi.org/10.1073/pnas.0902324106 PMID: 19451619

35. Thompson MC, Cascio D, Leibly DJ, Yeates TO. An allosteric model for control of pore opening by substrate binding in the EutL microcompartment shell protein. Protein Science. 2015; 24: 956–975. https://doi.org/10.1002/pro.2672 PMID: 25752492

36. 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. 2010; 192: 6056–6063. https://doi.org/10.1128/JB.00652-10 PMID: 20851901

37. Pang A, Warren MJ, Pickersgill RW. Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe–4S cluster-binding site. Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr. 2011; 67: 91–96. https://doi.org/10.1107/S0907444910050201 PMID: 21245529

38. Mallette E, Kimber MS. Structural and kinetic characterization of (S)-1-amino-2-propanol kinase from the aminoacetone utilization microcompartment of Mycobacterium smegmatis. J Biol Chem. 2018; jbc.RA118.005485. https://doi.org/10.1074/jbc.RA118.005485 PMID: 30361441

39. Keeling TJ, Samborska B, Demers RW, Kimber MS. Interactions and structural variability of β-carboxysomal shell protein CcmL. Photosynth Res. 2014; 121: 125–133. https://doi.org/10.1007/s11120-014-9973-z PMID: 24504539

40. Nichols RJ, Cassidy-Amstutz C, Chaijarasphong T, Savage DF. Encapsulins: molecular biology of the shell. Critical Reviews in Biochemistry and Molecular Biology. 2017; 52: 583–594. https://doi.org/10.1080/10409238.2017.1337709 PMID: 28635326

41. Sutter M, Boehringer D, Gutmann S, Günther S, Prangishvili D, Loessner MJ, et al. Structural basis of enzyme encapsulation into a bacterial nanocompartment. Nature Structural & Molecular Biology. 2008; 15: 939–947. https://doi.org/10.1038/nsmb.1473 PMID: 19172747

42. Giessen TW, Orlando BJ, Verdegaal AA, Chambers MG, Gardener J, Bell DC, et al. Large protein organelles form a new iron sequestration system with high storage capacity. Elife. 2019; 8. https://doi.org/10.7554/eLife.46070 PMID: 31282860

62. Quin MB, Perdue SA, Hsu S-Y, Schmidt-Dannert C. Encapsulation of multiple cargo proteins within recombinant Eut nanocompartments. Appl Microbiol Biotechnol. 2016; 100: 9187–9200. https://doi.org/10.1007/s00253-016-7737-8 PMID: 27450681

63. Lee MJ, Mantell J, Hodgson L, Alibhai D, Fletcher JM, Brown IR, et al. Engineered synthetic scaffolds for organizing proteins within the bacterial cytoplasm. Nature Chemical Biology. 2018; 14: 142–147. https://doi.org/10.1038/nchembio.2535 PMID: 29227472

64. Zhang G, Johnston T, Quin MB, Schmidt-Dannert C. Developing a Protein Scaffolding System for Rapid Enzyme Immobilization and Optimization of Enzyme Functions for Biocatalysis. ACS Synthetic Biology. 2019; 8: 1867–1876. https://doi.org/10.1021/acssynbio.9b00187 PMID: 31305981

Chapter 6: Structural characterization of hexamerix shell proteins from two types of choline utilisation bacterial microcomaprtments

Structural characterization of hexameric shell proteins from two types of choline utilization bacterial microcompartments

Jessica M. Ochoa[a], Oscar Mijares[b], Andrea A. Acosta[b], Xavier Escoto[c], Nancy Leon-Rivera[b], Joanna D. Marshall[c], Michael R. Sawaya[d] and Todd O. Yeates[cef]*

[a]UCLA-Molecular Biology Institute, University of California Los Angeles, 611 Charles E. Young Drive East, Los Angeles, California, 90095, United States

[b]Whittier College, 13406 E Philadelphia St, Whittier, California, 90602, United States

[c]Department of Chemistry and Biochemistry, University of California Los Angeles, 611 Charles E. Young Drive East, Los Angeles, California, 90095, United States

[d]UCLA-DOE Institute of Genomics and Proteomics, University of California Los Angeles, 611 Charles E. Young Drive East, Los Angeles, California, 90095, United States

[e]UCLA-DOE Institute of Genomics and Proteomics, University of California Los Angeles, Los Angeles, California, 90095, United States

[f]UCLA-Molecular Biology Institute, University of California Los Angeles, Los Angeles, California, 90095, United States

Correspondence email: yeates@mbi.ucla.edu

Funding Information

National Institutes of Health, NIAID (grant No. R01AI081146); National Institute of General Medical Sciences from the National Institutes of Health (award No. P30 GM124165 to Northeastern Collaborative Access Team beamlines (NECAT)); NIH-ORIP HEI (grant No. S10OD021527 to Eiger 16 M detector); Department of Energy Office of Science (grant No. DE-FC02-02ER63421 to Oscar Mijares, Andrea A. Acosta, Xavier Escoto, Nancy Leon-Rivera); U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of

Science by Argonne National Laboratory under Con-tract (award No. DE-AC02-06CH11357 to The Advanced Photon Source).

Synopsis

Several crystal structures are analyzed for proteins that form the shells of bacterial microcompartment organelles. These new structures are from a microcompartment where the metabolic substrate choline diffuses across the shell protein pore.

Abstract

Bacterial microcompartments are large supramolecular structures comprised of an outer proteinaceous shell that encapsulates various enzymes in order to optimize metabolic processes. The outer shells of bacterial microcompartments are made of several thousand protein subunits, generally forming hexameric building blocks based on the canonical bacterial microcompartment (BMC) domain. Among the diverse metabolic types of bacterial microcompartments, those that use glycyl radical enzymes to metabolize choline have not been adequately structurally characterized. Here, we report six structures of the hexameric shell proteins from the type I and type II choline utilization microcompartments. Sequence and structure analysis reveals electrostatic surface properties that are shared between the four types of shell proteins described here.

Keywords

bacterial microcompartment; bacterial organelle; BMC; shell protein; choline; glycyl radical

Introduction

Bacterial microcompartments, or MCPs, are proteinaceous organelle-like structures found in nearly 20% of bacteria (Jorda et al., 2013). These supramolecular structures are roughly 100 – 400 nm in diameter and have evolved to carry out a variety of specialized metabolic functions within the confines of a sequestered environment (Chowdhury et al., 2014; Kerfeld et al., 2010; Axen et al., 2014; Stewart et al., 2021). MCPs are comprised of a variety of signature enzymes encased within a proteinaceous outer shell that prevents the efflux of toxic or volatile intermediates during various multistep enzymatic reactions (Bobik et al., 2015; Kerfeld et al., 2018). While MCPs carry out diverse metabolic functions, including CO2 fixation and substrate degradation (including 1,2-propanediol and ethanolamine), they are unified by their structural similarity; MCPs are built from divergent but homologous families of structural proteins (Kerfeld et al., 2010; Cannon et al., 2001; Bobik et al., 2015; Kerfeld & Erbilgin, 2015; Yeates et al., 2008). The outer MCP shell is primarily composed of the canonical BMC (bacterial microcompartment) domain (Pfam 00936), a roughly 100-residue domain that serves as the functional building block for the larger MCP shell (Fig.1) (Tanaka et al., 2008; Yeates et al., 2011, 2013; Kerfeld et al., 2005; Tsai et al., 2007). A single BMC domain oligomerizes to form hexameric disks, which bear central pores for substrate and product transports, and which pack laterally to form the nearly flat facets of the outer MCP shell; pentameric proteins form the vertices of the polyhedral shell (Fig. 6.1) (Tanaka et al., 2008; Wheatley et al., 2013; Sutter et al., 2017). For all MCPs, multiple paralogs (between two and seven) of different BMC shell proteins are present.

Recent work has revealed two new types of BMC gene clusters in a wide variety of bacteria with subsequent work demonstrating the requirement for these microcompartment genes for choline utilization (Craciun & Balskus, 2012; Campo et al., 2015). The first type (type I) has approximately 20 genes while the second (type II) has approximately 16 genes, with both types encoding multiple paralogs of the canonical BMC domain (Campo et al., 2015). Both types of

Choline Utilization (Cut) MCPs use a glycyl radical choline TMA (trimethylamine) lyase to catalyze the conversion of choline to acetaldehyde and TMA (Herring et al., 2018; Craciun & Balskus, 2012). The type II Cut MCP was recently characterized in Escherichia coli 536 (Herring et al., 2018). First, choline diffuses across the outer MCP shell to enter the lumen and is subsequently converted to acetaldehyde and TMA via TMA lyase and its internal activating enzyme (Craciun & Balskus, 2012; Herring et al., 2018). Acetaldehyde is subsequently converted to acetyl-CoA and ethanol while additional downstream products are recycled back into the cell (Craciun & Balskus, 2012; Herring et al., 2018). The choline substrate of Cut MCPs is particularly intriguing because, unlike the substrates of other well-characterized MCPs, including ethanolamine (for Eut MCPs), 1,2-propanediol (for Pdu MCPs) and bicarbonate (for Carboxysomes), choline carries a fixed (non-titratable) positive charge. This prompted us to ask whether there are notable differences in surface electrostatics between the various homologs that might affect substrate recruitment to the MCP.

To date, there have been a limited number of structural studies that describe BMC shell proteins derived from Cut MCPs, types I and II (Ochoa et al., 2020; Kalnins et al., 2020). Thus, we sought to structurally characterize a library of BMC shell proteins from the type I and type II Cut MCPs and investigate differences in surface electrostatics by comparing paralogs and point mutants. Here, we describe the structure of one type I Cut MCP BMC shell protein, CutN, and three type II Cut MCP BMC shell proteins and their mutants; CmcA, CmcB and CmcC. CutN is a BMC shell protein from Streptococcus intermedius and CmcA, CmcB and CmcC are BMC shell proteins from Escherichia coli 536.

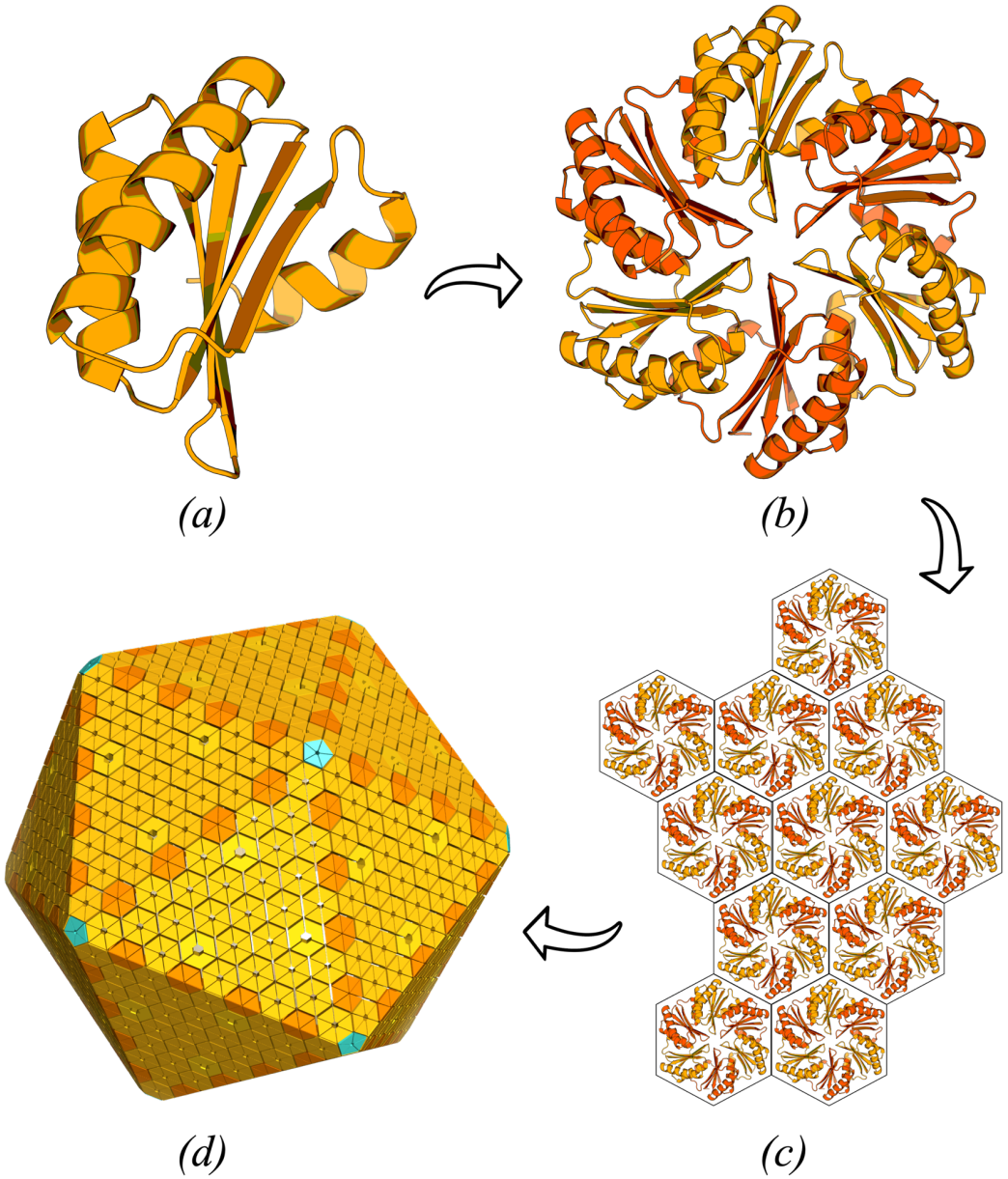


Figure 6.1

The BMC domain and bacterial microcompartment assembly principles. *(a)* All bacterial microcompartments contain a canonical BMC domain (Pfam 00936) that oligomerizes to form hexamers *(b)*. *(b)* Hexameric shell proteins interact along their edges to form flat facets *(c)* that ultimately comprise the bulk of the microcompartment (orange), while pentameric bacterial microcompartment vertex proteins (teal), belonging to an evolutionary unrelated protein family, cap the vertices *(d)*. *(d)* An idealized rendering of an icosahedral bacterial microcompartment. Microcompartments are generally more irregular in shape.

Materials and methods

Macromolecule production

For all structures, we purchased codon-optimized gBlock Gene Fragments from Integrated DNA Technologies (IDT). All constructs were ordered with a non-cleavable N-terminal His-6 purification tag and an edge lysine (K25 or K27 depending on residue position) mutation which facilitates recombinant BMC protein expression and purification. The position and type of mutation are indicated in Table 1. An additional mutation near the central pore of the dimpled BMC face (E35G) was introduced to CmcB and CmcC in order to investigate the ramifications of neutralizing the charged glutamate residue at this position.

The various gBlocks were inserted into either a pET-22b(+) or pET-24a(+) expression vectors using the NdeI and HindIII restriction endonuclease sites (Table 1). We used the BL21(DE3) E. coli expression system (New England Biolabs) to express recombinant proteins. For recombinant protein expression, we used 6 mL 1mg/mL ampicillin or kanamycin-supplemented (Table 6.1) LB broth cultures that were grown overnight at 37˚C, shaking. These overnight cultures were used to inoculate 1L of TB medium or LB medium and expression that was also supplemented with ampicillin or kanamycin (Table 6.1). Cells expressing CmcB and CutN, were grown in kanamycin-supplemented TB and induced with 1mM IPTG at OD600 = 1.2 and subsequently expressed for two hours at 30˚ C. Cells harbouring CmcA, CmcB E35G, CmcC and CmcC E35G were grown in ampicillin-supplemented LB and induced with ZYP-5052 autoinduction sugars (Liu & Naismith, 2008) and grown at 20˚ overnight.

All cells were subsequently harvested by centrifugation at 5,000 x g for 15 minutes at 4˚C. Excess media was decanted and cells were resuspended in a lysis buffer containing 20 mM Tris, pH 8.0, 300 mM NaCl, protease inhibitor cocktail tablets (Pierce Protease Inhibitor Tables, EDTA-Free, Thermo Scientific), lysozyme (Sigma), DNase (Sigma) and RNase (Sigma). Cells were

vortexed to ensure uniform resuspension and subsequently lysed via sonication. Lysates were clarified by centrifugation at 15,000 x g for 30 minutes at 4˚C.

Recombinant proteins were subject to a two-step purification to ensure homogenous, high-purity samples. Clarified lysates were applied to a pre-equilibrated nickel immobilized metal affinity chromatography (IMAC) gravity column (HisPur Ni-NTA Resin, Thermo Scientific). Non-specific proteins were washed using five column volumes of a buffer containing 20 mM Tris, pH 8.0, 300 mM NaCl and 25 mM imidazole. The His-6 tagged recombinant proteins of interest were subsequently eluted and collected with a buffer containing 20 mM Tris, pH 8.0, 300 mM NaCl, 300 mM imidazole and 10% glycerol. Eluted recombinant proteins were concentrated and immediately subject to a second round of purification. Concentrated proteins were subsequently passed over a gel-filtration column (Superdex 200 10/300 GL, GE Healthcare) and buffer exchanged into a final buffer containing 50 mM Tris pH 8.0, 150 mM NaCl. All recombinant proteins were concentrated down using 10 KDa MWCO Amicon Ultra concentrators (Millipore) and syringe-filtered through 0.22 µm filters (Millipore). Recombinant proteins were either immediately used for crystallization experiments or flash frozen with liquid nitrogen and stored at -80˚C.

Crystallization

We were able to obtain diffraction quality crystals for all constructs using a TTP Labtech robotic mosquito to set up broad screens in 96-well hanging-drop vapor diffusion format. Based on commercially available screens, set up at 22˚ C, diffraction-quality crystals were obtained without the need for substantial optimization. While we observed several crystal forms of varied morphologies in numerous conditions (Fig. 6.2), the most prevalent forms were flat hexagonal disks (Fig. 6.2c). Final concentrations and crystallization conditions are summarized in Table 6.2.

Table 6.1 Macromolecule production

Construct	CmcA	CmcB	CmcB E35G	CmcC	CmcC E35G	CutN
PDB Code	7MGP	7MPW	7MN4	7MPV	7MPX	7MMX
Source organism	*Escherichia coli 536*	*Escherichia coli 536*	*Escherichia coli 536*	*Escherichia coli 536*	*Escherichia coli 536*	*Streptococcus intermedius*
DNA source	Synthesized gBlock	Synthesized gBlock	Synthesized gBlock	Synthesized gBlock	Synthesized gBlock	Synthesized gBlock
Cloning and expression vector	pET-22b(+)	pET-24a(+)	pET-22b(+)	pET-22b(+)	pET-22b(+)	pET-24a(+)
Expression host	*Escherichia coli BL21(DE3)*	*Escherichia coli BL21(DE3)*	*Escherichia coli BL21(DE3)*	*Escherichia coli BL21(DE3)*	*Escherichia coli BL21(DE3)*	*Escherichia coli BL21(DE3)*
Complete amino acid sequence of the construct produced	MHHHHHHMGDALG LIETKGLVACIAA ADAMCASANVELI GYENIGSGLVTVM VKGDVGAVKASVD SGLESAQHIGEVV TSLVIARPHNDIN KIVIKHKA	MHHHHHHMGDALG LIETKGLVACIEA ADAMCDAANVELI GYENVGSGLVTAM VKGDVGAVKAAVD SGVESAQRIGEVV TSLVIARPHNDIN KIVSHYKITD	MHHHHHHMGDALG LIETKGLVACIEA ADAMCAAANVELI GYGNVGSGLVTAM VKGDVGAVKAAVD SGVESAQRIGEVV TSLVIARPHNDIN KIVSHYKITD	MHHHHHHMREALG LIETKGLVACIEA ADAMCAAANVELI GYENVGSGLVTAM VKGDVGAVNAAVD SGVEAARRIGEVV TSRVIARPHNDIE KIASQHKA	MHHHHHHMREALG LIETKGLVACIEA ADAMCAAANVELI GYGNVGSGLVTAM VKGDVGAVNAAVD SGVEAARRIGEVV TSRVIARPHNDIE KIASQHKA	MHHHHHHMNSE ALGMIETKGLV GSIEAADAMVD AANVTLIGKEH VGGGLVTVLVR GDVGAVKAATD AGAAAAQRVGE LVSVHVIPRPH IEVETILPHSK
Mutations	K25A	K25D	K25A, E35G	K25A	K25A, E35G	K27D
Antibiotic	Ampicillin	Kanamycin	Ampicillin	Ampicillin	Ampicillin	Kanamycin

Data collection and processing

X-ray diffraction data sets were collected in-house and at the Advanced Photon Source in Chicago on beamlines 24-ID-C and 24-ID-E (NE-CAT). The home-source beam was a RIGAKU-FR-E+ and is equipped with a RIGAKU RAXIS HTC detector, while the NECAT-24IDC and NECAT-24IDE beamlines were equipped with a Pilatus 6M-F detector and a Dectris Eiger 16M detector, respectively. After data collection, intensities were indexed, integrated and scaled using XDS and XSCALE and converted to structure-factor amplitudes with XDSCONV (Kabsch, 2010). Complete collection and processing data are summarized in Table 6.3.

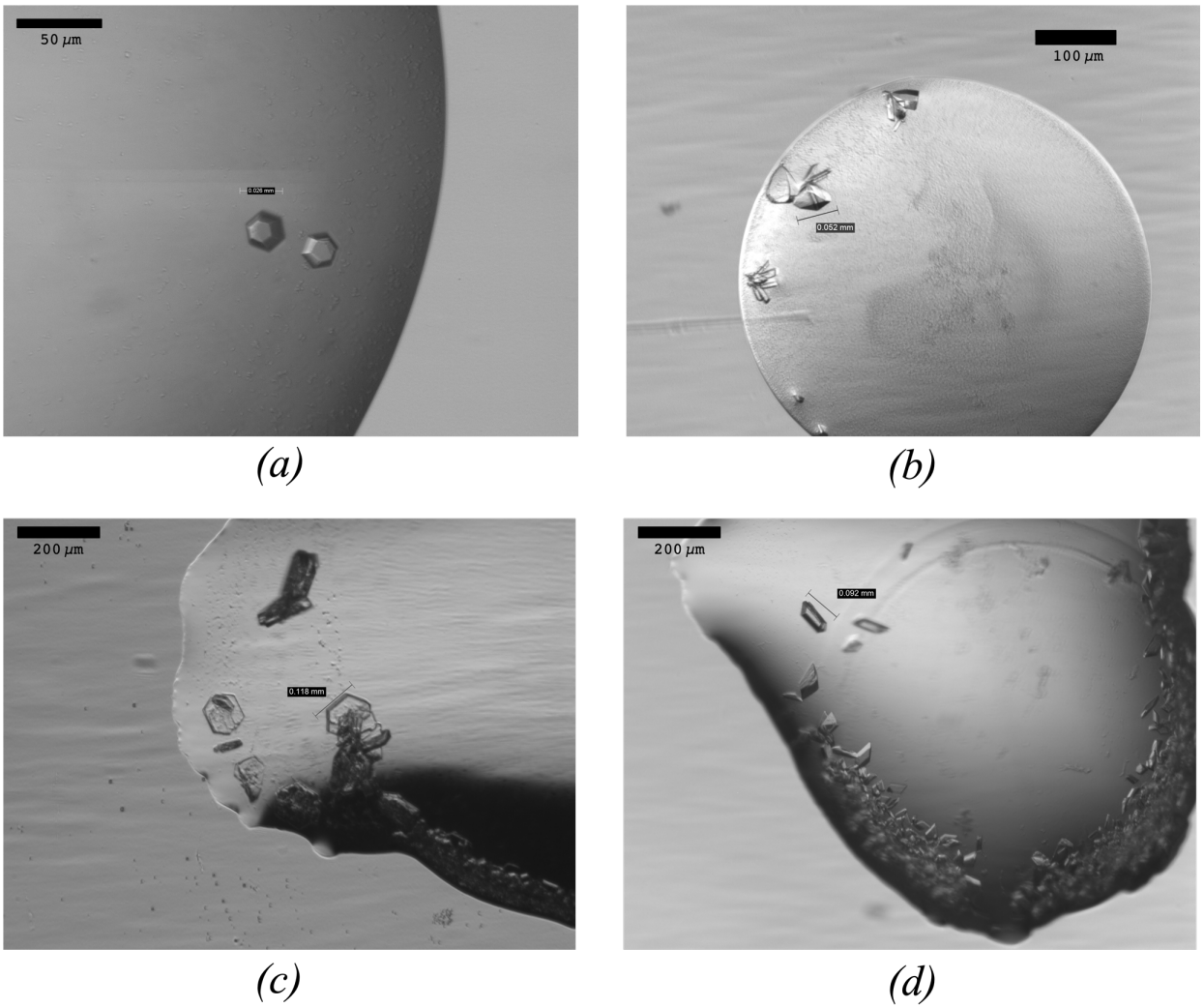


Figure 6.2

Crystal forms used for the determination of BMC structures. Photographs of (a) CmcA hexagonal disk-shaped crystals belonging to the P6 space group, (b) CmcB crystals belonging to the P1 space group and two CutN crystal forms (c,d). The most common morphologies across the proteins under study were flat hexagonal disks (c).

Table 6.2 Crystallization

Construct	CmcA	CmcB	CmcB E35G	CmcC	CmcC E35G	CutN
PDB Code	7MGP	7MPW	7MN4	7MPV	7MPX	7MMX
Method	Hanging-drop vapor diffusion	Hanging-drop vapor diffusion	Hanging-drop vapor diffusion	Hanging-drop vapor diffusion	Hanging-drop vapor diffusion	Hanging-drop vapor diffusion
Condition	Wizard D1	SS12 B5	JCSG D2	JCSG B4	PACT A3	Wizard F3
Temperature	294K	294K	294K	294K	294K	294K
Protein concentration	22.0 mg/mL	11.5 mg/mL	9.5 mg/mL	29.3 mg/mL	13.8 mg/mL	9.5 mg/mL
Buffer composition of protein solution	50 mM Tris, pH 8.0, 150 mM NaCl	50 mM Tris, pH 8.0, 150 mM NaCl	50 mM Tris, pH 8.0, 150 mM NaCl	50 mM Tris, pH 8.0, 150 mM NaCl	50 mM Tris, pH 8.0, 150 mM NaCl	50 mM Tris, pH 8.0, 150 mM NaCl
Composition of reservoir solution	2.5 M NaCl, 0.1 M Imidazole/ Hydrochloric acid pH 8.0	1.0 M Sodium acetate trihydrate, 0.1 M Imidazole pH 6.5	0.2 M $MgCl_2$, 0.1 M Sodium HEPES pH 7.5, 30% PEG 400	0.1 M HEPES pH 7.5, 10% PEG 8000, 8% Ethylene glycol	0.1 M SPG buffer pH 6.0, 25% PEG 1500	1.26 M $(NH)_4SO_4$, 0.1 M HEPES/ NaOH pH 7.5

Structure solution and refinement

We used molecular replacement to determine the structures of all four BMC shell proteins and their point mutants. PHASER (McCoy et al., 2007) was used to perform molecular replacement for all six structures. We used a BMC homolog, 4QIV, from a microcompartment of unknown function from Aeromonas hydrophila as models for CmcA and CmcB. We used 4AXJ, a BMC homolog from the ethanolamine utilization MCP (Pitts et al., 2012) as a model for CutN. Lastly, we used an early model of CmcB as a model for molecular replacement for CmcB E35G, CmcC and CmcC E35G. All models were built using Coot (Emsley & Cowtan, 2004) and refined using Phenix (Adams et al., 2002), Refmac (Winn et al., 2011) and Buster (Bricogne G., Blanc E., Brandl M., Flensburg C., Keller P., Paciorek W., & Roversi P, Sharff A., Smart O.S., Vonrhein C., Womack T.O., 2017). Because of the higher resolution data (1.8 Å) we obtained for the CmcB point mutant, we used CmcB E35G as an external reference structure and Buster in subsequent rounds of CmcB refinement. We utilized this strategy to apply external geometric restraints in

order to compensate for the moderate resolution (3.0 Å) of CmcB. Final refinement statistics are summarized in Table 6.4.

Table 6.3 Data collection and processing

Construct	CmcA	CmcB	CmcB E35G	CmcC	CmcC E35G	CutN
PDB Code	7MGP	7MPW	7MN4	7MPV	7MPX	7MMX
Diffraction source	APS-NECAT-24IDC	RIGAKU-FR-E+	APS-NECAT-24IDC	APS-NECAT-24IDC	APS-NECAT-24IDC	APS-NECAT-24IDE
Wavelength (Å)	0.9786	1.5418	0.9786	0.9786	0.9786	0.97918
Temperature	100K	100K	100K	100K	100K	100K
Detector	Pilatus 6M-F	RIGAKU RAXIS HTC	Pilatus 6M-F	Pilatus 6M-F	Pilatus 6M-F	Dectris Eiger 16M
Crystal-detector distance (mm)	250	170	350	400	350	300
Rotation range per image (°)	0.25	0.50	0.25	0.25	0.25	0.50
Total rotation range (°)	277	180	360	360	360	360
Exposure time per image (s)	0.247	300	0.247	0.247	0.247	0.25
Space group	*P6*	*P1*	$P2_1$	*P1*	$P2_1$	*C2*
a, *b*, *c* (Å)	68.86, 68.86, 27.78	36.22, 52.40, 61.16	67.39, 62.73, 67.39	50.60, 56.29, 92.03	66.57, 61.0, 66.61	107.50, 49.66, 92.41
α, β, γ (°)	90.00, 90.00, 120.00	107.43, 96.58, 101.96	90.00, 119.95 90.00	96.305, 90.083, 98.546	90.00, 119.92, 90.00	90.00, 93.834, 90.00
Mosaicity (°)	0.066	0.317		0.102		
Resolution range (Å)	59.6-1.65 (1.69-1.65)	57.3-3.00 (3.07-3.00)	58.4-1.80 (1.84-1.80)	91.5-2.3 (2.35-2.30)	57.7-2.1 (2.15-2.10)	92.2-1.90 (1.95-1.92)
Total No. of reflections	91175	31090	233088	152490	134627	249994
No. of unique reflections	9272	8055	44058	41777	26358	38026
Completeness (%)	100.0 (100.0)	97.1 (90.4)	96.4 (90.3)	93.2 (83.9)	96.1 (95.3)	98.5 (89.0)
Redundancy	9.83 (9.632)	3.85 (3.68)	5.29 (5.1)	3.65 (3.4)	5.11 (5.0)	6.57 (4.9)
$\langle I/\sigma(I)\rangle$	19.5 (3.1)	12.3 (2.32)	9.1 (2.1)	8.7 (1.71)	6.83 (0.87)	14.97 (2.25)
$R_{r.i.m.}$	0.069 (0.685)	0.108 (0.702)	0.093 (0.597)	0.091 (0.744)	0.118 (1.597)	0.078 (0.794)
Overall *B* factor (Å^2)	24.57	70.10	31.83	70.11	61.62	33.41

† The resolution at which $I/\sigma(I)$ falls below 2.0 are 2.29 Å (CmcC) and 2.27 Å (CmcC E35G), however the overall completeness and $CC_{1/2}$ suggests that resolutions of 2.3 Å (CmcC) and 2.1 Å (CmcC E35G) are appropriate for this data set. Overall $CC_{1/2}$ for CmcC and CmcC E35G are both 99.8%. $CC_{1/2}$ for the outer shell are 87.5% (CmcC) and 73.0% (CmcC E35G). Values for the outer shell are given in parentheses.

Table 6.4 Structure solution and refinement

Construct	CmcA	CmcB	CmcB E35G	CmcC	CmcC E35G	CutN
PDB Code	7MGP	7MPW	7MN4	7MPV	7MPX	7MMX
Resolution range (Å)	59.63-1.65 (1.66-1.65)	57.30-3.00 (3.03-3.00)	58.4-1.80 (1.82-1.80)	91.46-2.29 (2.31-2.29)	57.3-2.10 (2.14-2.10)	92.20 – 1.90 (1.92-1.90)
Completeness (%)	100.0 (100.0)	97.4 (90.4)	96.16 (83.0)	93.2 (83.9)	94.8 (91.0)	98.4 (83.4)
σ cutoff	0	0	0	0	0	0
No. of reflections, working set	9272 (181)	8055 (181)	43821 (1147)	41777 (753)	25881 (1204)	38026 (685)
No. of reflections, test set	928 (21)	806 (21)	4380 (128)	4178 (83)	2584 (135)	3803 (76)
Final R_{cryst}	0.212 (0.320)	0.221 (0.325)	0.200 (0.368)	0.233 (0.415)	0.237 (0.521)	0.212 (0.326)
Final R_{free}	0.233 (0.395)	0.257 (0.254)	0.240 (0.402)	0.269 (0.514)	0.283 (0.555)	0.236 (0.338)
Cruickshank DPI	0.112	0.473	0.334	0.43	0.212	0.165
No. of non-H atoms						
Protein	647	3320	3889	7267	3304	3607
Ion	0	0	0	0	5	20
Water	27	7	188	132	47	171
Total	674	3327	4077	7399	3356	3798
R.m.s. deviations						
Bonds (Å)	0.008	0.008	0.006	0.005	0.007	0.008
Angles (°)	1.02	0.990	0.769	0.830	0.864	0.950
Average *B* factors (Å²)						
Protein	29.9	70.2	31.53	70.3	61.6	33.02
Ion	n/a	n/a	n/a	n/a	97.9	49.7
Water	36.1	34.9	38.0	58.3	58.2	40.0
Ramachandran plot						
Favoured (%)	100	99.8	100	98.7	99.4	99.6
Allowed (%)	100	0.2	0	1.2	0.6	0.4
Outliers (%)	0	0	0	0.1	0	0

Values for the outer shell are given in parentheses

Results and discussion

Structural characterization

Here, we characterized a library of BMC shell proteins from the choline utilization MCP, types I and II. We determined the structures of four BMC shell proteins and two point mutants to final resolutions between 1.7 – 3.0 Å. This work reveals the structure of four BMC shell proteins that contain the characteristic secondary structure elements of the BMC domain. The BMC domain (Pfam 00936) is a roughly 100-residue domain that consists of a 4-stranded anti-parallel beta-sheet sandwiched between three alpha-helices (Fig. 6.3).

CmcA (PDB ID 7MGP) formed 26 μm crystals (Fig. 6.2a) in a solution containing 2.5 M sodium chloride, 0.1 M imidazole/hydrochloric acid at pH 8.0 (Table 6.2) and diffracted to a resolution of 1.7 Å (Table 6.3). These crystals belonged to the *P6* hexagonal space group and contained a single monomer in the asymmetric unit. The electron density map was of sufficient quality for modelling a portion of the N-terminal His-6 tag, residues 3-91 and twenty-seven ordered waters (Fig. 6.4a). Additional refinement statistics are summarized in Table 4.

CmcB (PDB ID 7MPW) formed up to 50 μm crystals (Fig. 6.2b) in a solution containing 1.0 M sodium acetate trihydrate and 0.1 M imidazole at pH 6.5 (Table 6.2) and diffracted to a resolution of 3.0 Å (Table 6.3). Crystals that gave rise to this structure belonged to the *P1* space group. CmcB contains one full hexamer (comprised of six monomers) in the asymmetric unit. Because of the slightly poorer resolution, we were only able to model residues 3-86, and observed density for seven ordered waters (Fig. 6.4b). By contrast, the E35G CmcB point mutant (PDB 7MN4) diffracted to markedly better resolution at 1.8 Å (Table 6.3). This point mutant crystalized in a solution containing 0.2 M magnesium chloride, 0.1 M sodium HEPES at pH 7.5 and 30% PEG 400 (Table 6.2) in the $P2_1$ space group, containing six monomers (one full BMC hexamer) in the asymmetric unit. We were able to model residues 3-93 and observed density for one hundred eighty-eight ordered waters (Fig. 6.4c).

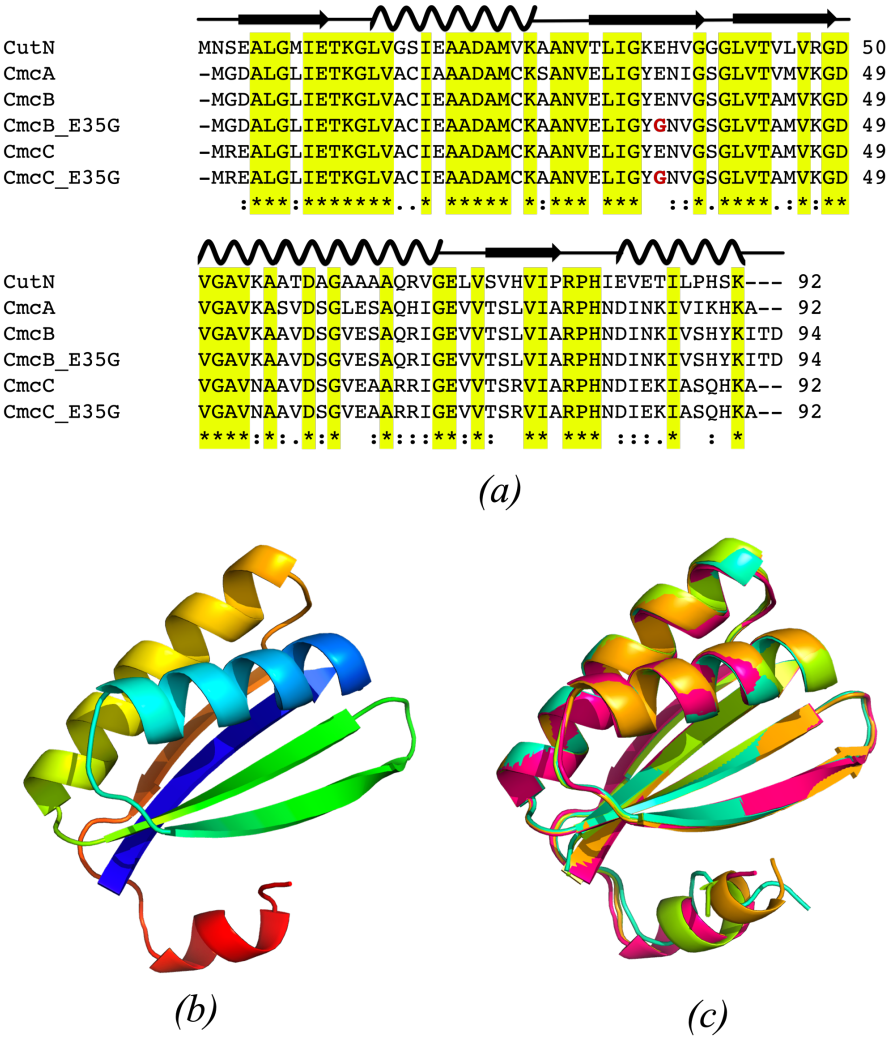


Figure 6.3

Sequence and secondary structure alignment of the four BMC homologs and associated mutants. *(a)* All four BMC shell proteins and their mutants were aligned using the *Clustal Omega* multiple sequence alignment. The E56G is highlighted in red and bold-face. Conserved residues are denoted by asterisks (*) and lime-green boxes. Conservative differences are denoted by one (.) or two (:) dots and non-conservative differences are denoted by blank spaces underneath the sequence. Point mutations are colored red. Secondary structure elements are indicated above the sequence with waves representing α-helices and solid arrows representing beta-sheets. *(b)* Color-ramped cartoon depiction of a single BMC domain. The N-terminus is blue while the C-terminus is red. *(c)* Structural superimposition - of CmcA (orange), CmcB (lime green), CmcC (teal) and CutN (fuchsia). All four structures were within 0.22 – 0.42 Å RMSD over backbone atoms.

CmcC (PDB7MPV) crystalized in a solution containing 0.1 M HEPES at pH 7.5, 10% PEG 8000 and 8% ethylene glycol and diffracted to 2.3 Å resolution. CmcC crystals belonged to the *P1* space group and contained twelve monomers in the asymmetric unit (two full BMC hexamers). We were able to model most of the protein (residues 1-89), one hundred thirty-two ordered waters and portions of the N-terminal His-6 tag (Fig. 6.4d). The E35G CmcC point mutant (PDB 7MPX) diffracted to slightly better resolution at 2.1 Å (Table 6.3). Diffraction quality crystals in the $P2_1$ space group formed in a solution containing 0.1 M SPG buffer at pH 6.0 and 25% PEG 1500. CmcC E35G contained six monomers in the asymmetric unit (one full BMC hexamer). We were able to model residues 2-82, one phosphate ion and forty-seven ordered waters (Fig. 6.4e).

Lastly, CutN (PDB 7MMX) from the type I Cut MCP, formed numerous crystal morphologies including hexagonal plates (Fig. 6.2c) and prisms (Fig. 6.2d). The final structure was determined from crystals that formed in 1.26 M ammonium sulfate and 0.1 M HEPES with sodium hydroxide at pH 7.5. These CutN crystals diffracted to a final resolution of 1.9 Å in the *C2* space group and had six monomers in the asymmetric unit. We found that these CutN crystals contained two half hexamers in the asymmetric unit. Applying a twofold symmetry operator to each monomer in the half-hexamer results in the complete biological assembly, a hexameric BMC. There was sufficient density to model residues 5-92, portions of the N-terminal His-6 tag, one hundred seventy-one ordered waters and four sulfate ions. In one half hexamer (chains A, B and C) there are two sulfates with 100% occupancy at a crystallographic interface, and one sulfate at the pore with 50% occupancy. In the second half-hexamer (chains D, E and F), there is one sulfate ion at the pore with 50% occupancy. Figure 4c represents the biological assembly generated from chains A, B and C. Additional refinement statistics are summarized in Table 4.

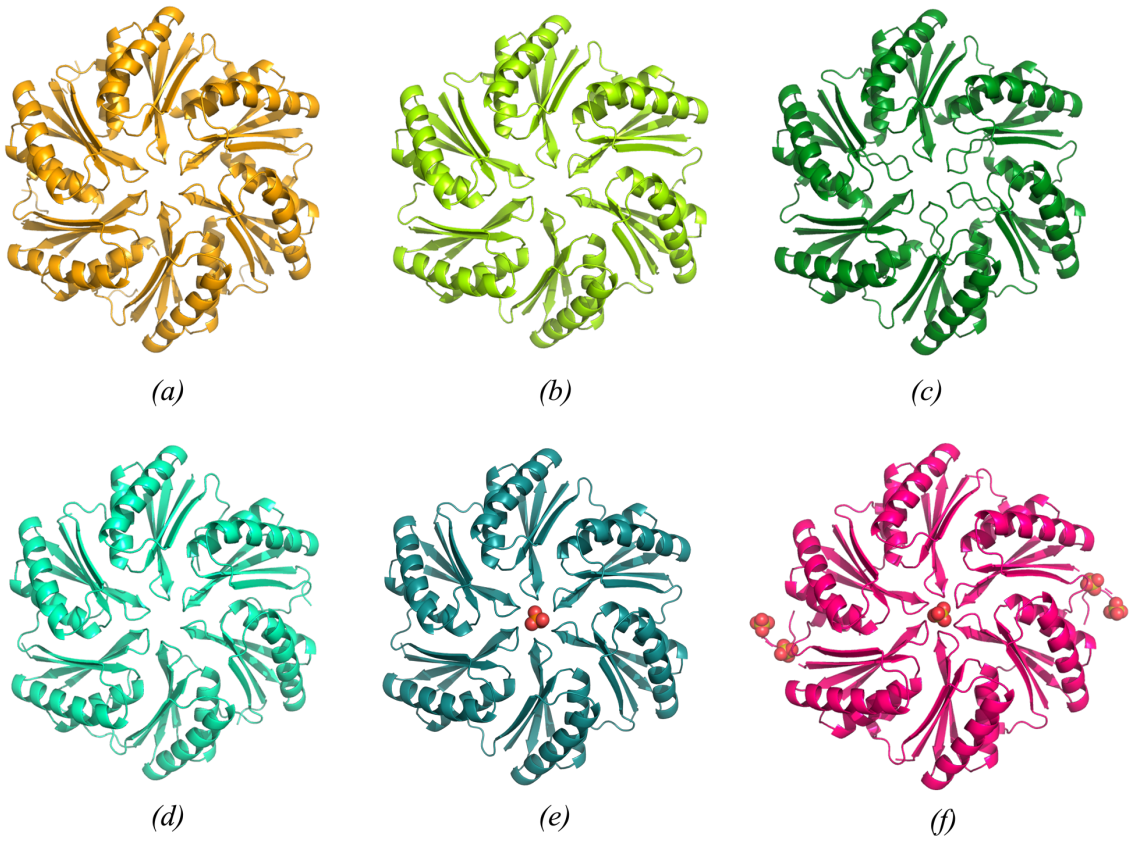


Figure 6.4

Overall structure of the six BMC shell proteins and point mutants in this study. These hexameric BMC shell proteins are similar in structure with RMSD values of a single monomer between 0.22 Å and 0.42 Å. We determined the structures of CmcA (a), CmcB (b), CmcB E35G (c), CmcC (d), CmcC E35G (e) and CutN (f).

Structural comparison

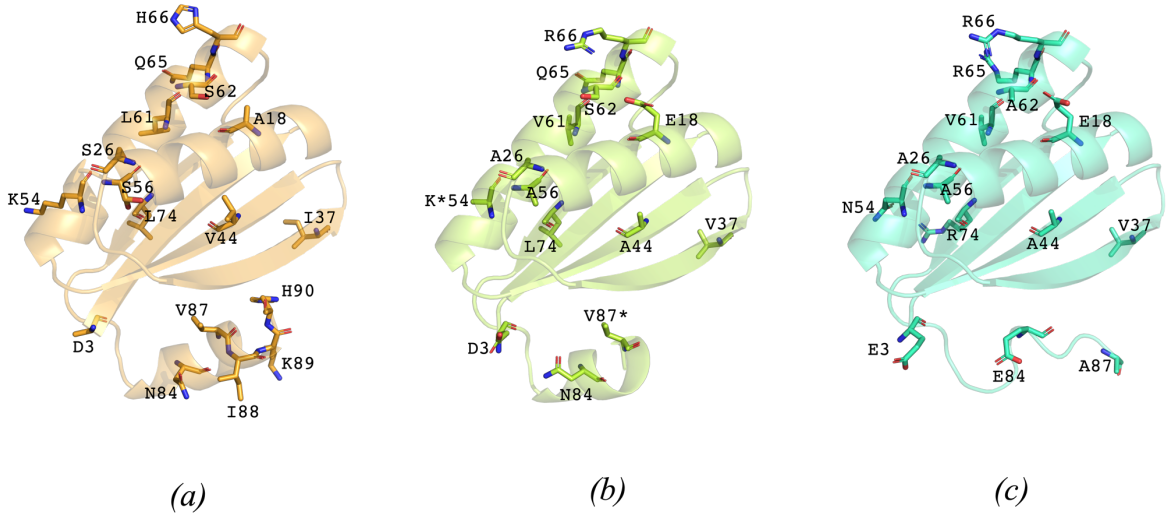
In this work we determined the structures of three paralogs encoded within the operon for the type II Cut MCP from *E.coli 536*; CmcA, CmcB and CmcC . These paralogs share greater than 78% sequence similarity with root mean square deviations (RMSD) between 0.22 – 0.28 Å for a single monomer (Fig. 6.3). The greatest amount of variability is located in primarily solvent-exposed regions of α-helices (Fig. 6.5). CmcA and CmcB both contain Gly2 and Leu74 while

CmcC has an arginine in equivalent positions (Fig. 6.5). By contrast, CmcB and CmcC both contain Glu18 and Ser88 while CmcA contains Ala18 and Ile88 in equivalent positions (Fig. 5).

By contrast, CutN shares approximately 52% sequence similarity with the type II Cut MCP shell proteins CmcA, CmcB and CmcC (Fig. 6.3a). Superimposition of a single monomer of CutN with any one of the type II BMC monomers yields an RMSD of 0.37 – 0.42 Å (Fig. 6.3c). Unsurprisingly, there are several key conserved residues that give rise to canonical BMC secondary structure elements, especially in the first fifty-four residues (Fig. 6.3a). In CutN Ala5-Gly, Ile7-Val15, Ile18, Ala29-Met24, Lys26, Ala28-Val30, Leu32-Gly34, Gly39, Gly41-Thr44, Val47 and Gly49-Val54 are conserved across the type I and type II shell proteins characterized in this work (Fig. 6.3a).

Hexameric BMC shell proteins have a distinct sidedness, with one surface being relatively flat and the other appearing more concave or dimpled owing to a central depression. We sought to investigate the differences in surface electrostatics between both faces of the various Cut MCP shell proteins. Using the Adaptive Poisson-Boltzman Solver (APBS) plugin (Baker *et al.*, 2001; Jurrus *et al.*, 2018) for PyMOL we generated models of the electrostatic surface potentials of all six structures (Fig. 6.6). APBS utilizes a continuum solvation method which first requires preparing the model. To do so, we used PDB2PQR (Dolinsky *et al.*, 2004, 2007) in order to add missing side-chain atoms, add hydrogens and assign charge and radius parameters. Electrostatic potentials maps were calculated using a 0.50 Å grid spacing and a solvent excluded surface with a range of +/- 5.0 kT/e. There is a distribution of positive and negative charge over the surface of the flat face on all four BMC types (CmcA, CmcB, CmcC and CutN). There are less pronounced regions of negative charges on the flat face of CmcC compared to the other three homologs (CmcA, CmcB and CutN) (Fig. 6). The N and C-terminal residues are located on the dimpled face of all four homologs. Because we were unable to model 100% of the residues for all structures, we only highlight the differences in surface electrostatics of the dimpled face

near the central pore, where we have complete structural data. The type II Cut MCP paralogs (CmcA, CmcB and CmcC) all contain a pocket of negative potential charge approximately 25 Å in diameter at the pore because of a conserved Glu35 (Fig. 6.3). This negative potential is not as prominent in the CutN homolog and is unsurprisingly partially neutralized in the E35G point mutants of CmcB and CmcC (Fig. 6.6). For all structures, there is a pocket of positive potential on the flat face, approximately 15 Å in diameter at the central pore (Fig. 6.6). This positive potential is the result of a conserved Lys11 (Fig. 6.3). The very center of the pore is neutral as it is lined by a conserved G-X-G motif at approximately residue 38 where X is either a serine or a glycine (Fig. 6.3).



(a) *(b)* *(c)*

Figure 6.5

Key residue differences between paralogs of the type II choline utilization microcompartment. Despite greater than 78% sequence identity between the three paralogs, key differences in residues exist in the primarily solvent exposed regions of α-helices in CmcA *(a)*, CmcB *(b)* and CmcC *(c)*. *V87 for CmcB *(b)* is modelled here for illustrative purposes but is not in the final model of CmcB due to insufficient density at the C-terminus.

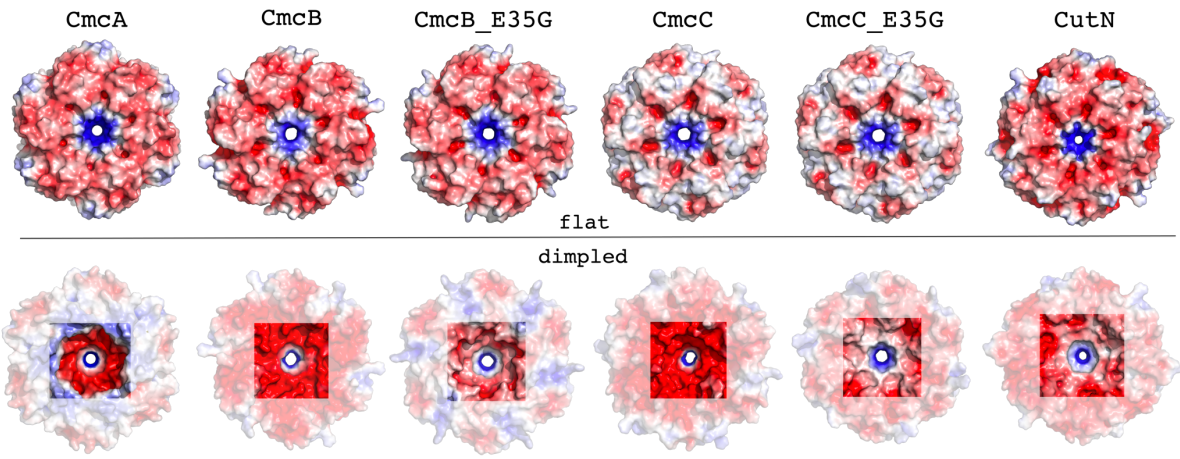


Figure 6.6

Electrostatic maps of the various BMC shell protein structures and their mutants. Blue is positive, red is negative with contour levels of +/- 5 kT/e. We have highlighted the pore-regions only on the dimpled faces of all six structures.

Conclusions

Previous work used cryo-electron microscopy to resolve the structure of a CmcC mutant from derived from Klebsiella pneumoniae to 3.3 Å (Kalnins et al., 2020). The present study provides a higher a resolution structure of CmcC from E. coli 536 (2.3 Å) in addition to two paralogs from the type II Cut MCP (CmcA and CmcB). The previously characterized mutant CmcC from Klebsiella pneumoniae and CmcC from E. coli 536 share 89% sequence identity and an RMSD of 0.39 Å upon superimposition of a single monomer. These new structures add to the growing library of BMC shell proteins (Ochoa et al., 2021) by resolving the structures of three type II Cut MCPs and their point mutants as well as a type I Cut MCP, CutN. We confirmed key structural features found across BMC shell proteins and began to probe differences in electrostatic surface potentials between paralogs and homologs of Cut MCPs. Complete structural data, in which all residues are modelled, could be helpful to fully interrogate both the flat and dimpled faces of Cut MCP BMC shell proteins.

BMC shell proteins present a unique and valuable system for studying selective transport through protein pores. A few computational studies have begun to explore this critical aspect of their function, evolution and diversity (Chowdhury et al., 2015; Slininger Lee et al., 2017; Faulkner et al., 2020) but important questions remain. Deriving from less-studied MCP types using choline as the transported substrate, the new structures reported here will provide important subjects for those ongoing studies. Understanding diffusive transport through BMC pores will ultimately provide an important counterpoint to decades of research on molecular transport across the pores of transmembrane protein channels.

Acknowledgements

We thank Prof. Tom Bobik for advice and insight on the choline utilization shell proteins. We also thank Michael Collazo, Duilio Cascio, Vy Nguyen, Mengxiao Nie, Mark Arbing, Sum Chan, Tiffany Cheng, Niloofar Mozhgani, Elizabeth Rojas and the UCLA Crystallization Facility. JMO is an HHMI Gilliam Fellow. We are grateful for the support of the NECAT staff.

References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). Acta Crystallogr. D Biol. Crystallogr. 58, 1948–1954.

Axen, S. D., Erbilgin, O. & Kerfeld, C. A. (2014). PLOS Computational Biology. 10, e1003898.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). PNAS. 98, 10037–10041.

Bobik, T. A., Lehman, B. P. & Yeates, T. O. (2015). Molecular Microbiology. 98, 193–207.

Bricogne G., Blanc E., Brandl M., Flensburg C., Keller P., Paciorek W., & Roversi P, Sharff A., Smart O.S., Vonrhein C., Womack T.O. (2017).

Campo, A. M., Bodea, S., Hamer, H. A., Marks, J. A., Haiser, H. J., Turnbaugh, P. J. & Balskus, E. P. (2015). MBio. 6,.

Cannon, G. C., Bradburne, C. E., Aldrich, H. C., Baker, S. H., Heinhorst, S. & Shively, J. M. (2001). Appl. Environ. Microbiol. 67, 5351–5361.

Chowdhury, C., Chun, S., Pang, A., Sawaya, M. R., Sinha, S., Yeates, T. O. & Bobik, T. A. (2015). PNAS. 112, 2990–2995.

Chowdhury, C., Sinha, S., Chun, S., Yeates, T. O. & Bobik, T. A. (2014). Microbiol. Mol. Biol. Rev. 78, 438–468.

Craciun, S. & Balskus, E. P. (2012). PNAS. 109, 21307–21312.

Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G. & Baker, N. A. (2007). Nucleic Acids Res. 35, W522-525.

Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. (2004). Nucleic Acids Res. 32, W665-667.

Emsley, P. & Cowtan, K. (2004). Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132.

Faulkner, M., Szabó, I., Weetman, S. L., Sicard, F., Huber, R. G., Bond, P. J., Rosta, E. & Liu, L.-N. (2020). Scientific Reports. 10, 17501.

Herring, T. I., Harris, T. N., Chowdhury, C., Mohanty, S. K. & Bobik, T. A. (2018). J. Bacteriol. 200, e00764-17.

Jorda, J., Lopez, D., Wheatley, N. M. & Yeates, T. O. (2013). Protein Science. 22, 179–195.

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G., Holst, M. J., McCammon, J. A. & Baker, N. A. (2018). Protein Sci. 27, 112–128.

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

Kalnins, G., Cesle, E.-E., Jansons, J., Liepins, J., Filimonenko, A. & Tars, K. (2020). Nature Communications. 11,.

Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F. & Sutter, M. (2018). Nature Reviews Microbiology.

Kerfeld, C. A. & Erbilgin, O. (2015). Trends in Microbiology. 23, 22–34.

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

Kerfeld, C. A., Sawaya, M. R., Tanaka, S., Nguyen, C. V., Phillips, M., Beeby, M. & Yeates, T. O. (2005). Science. 309, 936–938.

Liu, H. & Naismith, J. H. (2008). BMC Biotechnology. 8, 91.

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

Ochoa, J. M., Bair, K., Holton, T., Bobik, T. A. & Yeates, T. O. (2021). BioRxiv. 2021.01.09.426059.

Ochoa, J. M., Nguyen, V. N., Nie, M., Sawaya, M. R., Bobik, T. A. & Yeates, T. O. (2020). Protein Science. 29, 2201–2212.

Pitts, A. C., Tuck, L. R., Faulds-Pain, A., Lewis, R. J. & Marles-Wright, J. (2012). PLoS One. 7, e48360.

Slininger Lee, M. F., Jakobson, C. M. & Tullman-Ercek, D. (2017). ACS Synth. Biol. 6, 1880–1891.

Stewart, A. M., Stewart, K. L., Yeates, T. O. & Bobik, T. A. (2021). Trends in Biochemical Sciences.

Sutter, M., Greber, B., Aussignargues, C. & Kerfeld, C. A. (2017). Science. 356, 1293–1297.

Tanaka, S., Kerfeld, C. A., Sawaya, M. R., Cai, F., Heinhorst, S., Cannon, G. C. & Yeates, T. O. (2008). Science. 319, 1083–1086.

Tsai, Y., Sawaya, M. R., Cannon, G. C., Cai, F., Williams, E. B., Heinhorst, S., Kerfeld, C. A. & Yeates, T. O. (2007). PLoS Biol. 5,.

Wheatley, N. M., Gidaniyan, S. D., Liu, Y., Cascio, D. & Yeates, T. O. (2013). Protein Sci. 22, 660–665.

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

Yeates, T. O., Jorda, J. & Bobik, T. A. (2013). Journal of Molecular Microbiology and Biotechnology. 23, 290–299.

Yeates, T. O., Kerfeld, C. A., Heinhorst, S., Cannon, G. C. & Shively, J. M. (2008). Nature Reviews Microbiology. 6, 681–691.

Yeates, T. O., Thompson, M. C. & Bobik, T. A. (2011). Current Opinion in Structural Biology. 21, 223–231.

Chapter 7: Interrogating the topology of the 1,2-Propanediol Microcompartment

Abstract

Bacterial microcompartments, or MCPs, are organelle-like supramolecular structures found in nearly 20% of all bacteria that are capable of metabolizing a variety of substrates. Internal enzymes are encased within a roughly polyhedral proteinaceous shell comprised primarily of homologous tessellating hexameric proteins. These hexameric disks have distinctly flat and dimples faces. Previous cryo-electron tomography work on carboxysomes, the most regular of the MCPs, revealed a nearly icosahedral shell at modest resolution. The 1,2-propanediol (Pdu) MCP is an irregularly shaped MCP that exists in a variety of sizes and morphologies. To date, the no other MCP, including the Pdu MCP have been confirmed to obey icosahedral geometry (ie. a polyhedron with 20 faces and 12 vertices). Moreover, the native orientation of microcompartment shell proteins have yet to be confirmed within the context of an endogenous MCP. Here, we provide preliminary evidence that dissects the overall geometric and topological features of the Pdu microcompartment. We describe ongoing efforts to define their geometries and utilize subtomogram averaging and existing crystallographic data of individual shell proteins in order to resolve these shell proteins within the context of the Pdu MCP.

Introduction

In this work we sought to characterize the overall shape and morphology of the 1,2-propanediol (Pdu) microcompartment. Pdu MCPs were first characterized in *Salmonella enterica* serovar Typhimurium LT2 when it was discovered that *Salmonella* were capable of utilizing B_{12} and adenosylcobalamin to 1,2-propanediol (1,2-PD) [1,2]. Through a now extensively characterized pathway 1,2-PD is converted to propionaldehyde via coenzyme B_{12}-dependent diol dehydratase. The highly reactive propionaldehyde intermediate is subsequently converted to 1-propanol and propionic acid through a series of reactions involving propionaldehyde dehydrogenase, 1-propanol dehydrogenase, phosphotransacylase and propionate kinase in

order to provide the cell with ATP and propionyl-CoA [3–8]. The presence of a gene cluster, the *pdu* operon, was deemed essential for the formation of large polyhedral bodies of varying shape and morphology [1,2] with subsequent work identifying and characterized the various BMC shell proteins (PduA, PduB, PduB', PduJ, PduK, PduM, PduT and PduU) [9–15]. Crystallographic work has been used to determine the structures of PduA [14], PduJ [11] and PduU [12] at high resolution, providing important structural insight. As with other BMC shell proteins, PduA, PduJ and PduU have been confirmed to possess distinct sidedness; one surface has relatively flat face and the other appears more concave or dimpled possessing a central depression (Figure 7.1). More importantly, structural data has helped to reveal the location and positions of termini and residues of interest relative to their flat and dimpled faces. For PduA, PduJ and PduU, the C-terminus is located on the dimpled face (Figure 7.1). For PduA and PduJ, the N-terminus is also located on the dimpled face whereas the N-terminus is located on the flat face because of a circular permutation in the sequence that results in the reordering of the amino acid sequence (Figure 7.1) [12].

For several years, it was hypothesized that the dimpled face of these hexameric shell proteins was oriented inward, towards the internal lumen. Initial computational work on carboxysome shell proteins provided initial models for possible orientations of the hexameric and pentameric shell proteins within the context of a complete MCP shell [16,17]. Subsequent biophysical characterization found that luminal enzymes interact with the C-terminus of PduA, thus suggesting that the dimpled side faced the internal lumen [18]. Recent characterization of miniaturizes, enzyme-free recombinant MCP shells, however, revealed that the dimpled side was oriented outward, toward the cytosol [19–22] providing an alternative model to in long-held hypothesis surrounding MCP topology. While this remains an outstanding question, the overall architecture of less-regular MCPs, like the Pdu MCP remains equally elusive.

Previous transmission electron microscopy (TEM) revealed moderately resolved Pdu MCPs with irregular distribution in shape and morphology. Unlike carboxysomes, Pdu MCPs are less regular in shape. Carboxysomes assemble as nearly perfect icosahedrons, bearing twenty faces and twelve vertices [23,24]. Despite their more regular distribution in shape, carboxysomes remain poor candidates for single particle cryo-electron microscopy because of their heterogenous outer shell. Thus, determining the overall geometry of less-regular MCPs, including Pdu MCPs remained an elusive feat. Remarkably, recent work revealed exquisitely purified and TEM-imaged Pdu MCPs revealing, for the first time, that Pdu MCPs possessed more well-defined facets than was previously appreciated [25].

Here, we seek to structurally characterize the overall topology and architecture Pdu MCP structures. We utilize existing structural data and knowledge of endogenous metabolic systems to develop an assay that probes for lysine modifications by reaction intermediates. Finally, we utilize emerging advances in cryo-electron microscopy and cryo-electron tomography in order catalog the diversity in Pdu MCP morphology and propose methods for utilizing site-specific biotinylation and sub-tomogram averaging to determine their overall topology and long-range organization.

Materials and methods

Recombinant protein expression and purification

We purchased codon optimized gBlock® Gene Fragments of PduA and PduJ from *Salmonella enterica typhimurium* LT2 from Integrated DNA Technologies (IDT). Both recombinant proteins contain the fifteen-reside AviTag (GLNDIFEAQKIEWHE) and a non-cleavable C-terminal His-6 purification tag. We inserted the gBlock into the pET22b expression vector using *NdeI* and *HindIII* restriction endonuclease sites.

We expressed recombinant protein using the BL21(DE3) *E. coli* expression system (New England Biolabs). Briefly, we used 6mL 1 mg/L ampicillin-supplemented LB overnight cultures to inoculate 1 L of autoinduction media. We grew cells in TB medium supplemented with ampicillin and 5052 autoinduction sugars. Cultures were grown at 37˚C for six hours and then 18˚C overnight and subsequently harvested by centrifugation for 15 min at 5,000 x g.

We resuspended cells in a buffer containing 20 mM Tris pH 8.0, 300 mM NaCl, protease inhibitor cocktail tablets (Pierce™ Protease Inhibitor Tablets, EDTA-Free, Thermo Scientific), lysozyme (Sigma), DNase (Sigma) and RNase (Sigma). We lysed the resuspended pellets by sonication at 4˚C and clarified the lysate by centrifugation (15,000 x g for 30 minutes at 4˚C). The clarified lysate was applied over a pre-equilibrated Nickel IMAC gravity column (HisPur™ Ni-NTA Resin, Thermo Scientific) and the His-tagged recombinant protein was eluted using 20 mM Tris pH 8.0, 300 mM NaCl, 300 mM imidazole and 10% glycerol. We performed a second round of purification using gel-filtration chromatography (Superdex® 200 10/300 GL, GE Healthcare) and eluted samples with a buffer containing 50 mM Tris pH 8.0 and 150 mM NaCl. Fractions from peaks that corresponded to a molecular weight of approximately 82 kDa were pooled together and purify was verified using denaturing SDS-PAGE. Following the two-step purification, proteins were concentrated to approximately 10 mg/ml using 10 KDa MWCO Amicon Ultra concentrators (Millipore), flash-frozen with liquid nitrogen and stored at -80˚C until ready for use.

Chromosomal insertions

We constructed scarless chromosomal mutations in order to generate AviTag bearing Pdu MCP mutants. We appended the fifteen-residue AviTag (GLNDIFEAQKIEWHE) to the N-terminus of PduA, PduJ or PduU thus generating three *Salmonella enterica typhimurium* LT2 strains that harbored the AviTag on PduA (*N-Avi-PduA*), on PduJ (*N-Avi-PduJ*) or PduU (*N-Avi-PduU*). We used a modified recombineering as previously described [26–28]. Briefly, the *sacB-cat* cassette was subcloned into the pET-41a vector between the *SphI* and *XhoI* restriction sites.

The BE293 *Salmonella enterica typhimurium* LT2 strain containing the pKD46 plasmid (harboring lambda red recombinase) was transformed with the oligo corresponding the fifteen-residue AviTag via electroporation. We ordered and inserted oligos from Integrated DNA Technologies for *N-Avi-PduA*, for *N-Avi-PduJ* and for *N-Avi-PduU* (see supporting information). Following recombineering, we confirmed the presence of the N-terminal AviTag with PCR and genomic DNA sequencing. We proceeded to assess the integrity and morphology of these Pdu MCP mutant strains using transmission electron microscopy ad described below.

Pdu Expression and Purification

We utilized a previously described method (Haveman, Sinha) to optimize a robust and facile method for the expression and purification of endogenous Pdu MCPs from *Salmonella enterica typhimurium* LT2. Pdu MCPs were freshly purified prior to downstream analysis. Briefly, an LT2 -80˚C glycerol stock was used to inoculate 3mL of antibiotic-free LB from and grown overnight at 37˚C in a shaking incubator. 5µL of this overnight culture was used to inoculate a fresh 8mL LB culture and grown for six hours at 37˚C. After six hours, the full 8 mL culture was used to inoculate 400 mL of media that contained 1X NCE, 1 mM $MgSO_4$, 0.5% succinate and 0.6% 1,2-propanediol and grew at 37˚C for approximately 16-18 hours.

Cells were harvest by centrifugation 5,000 x g for 15 min at 4˚C. Excess media was decanted and cells were resuspended in 20mL of a Buffer A which contained 50 mM Tris-HCl (pH 8.0), 500 mM KCl, 12.5 mM $MgCl_2$ and 1.5% 1,2-propanediol in order to wash cells and remove residual media and subsequently re-pelleted by centrifugation at 5000 x g for 15 minutes at 4˚C. After decanting excess buffer, cells were resuspended in a lysis buffer containing a mixture of 40% Buffer A and 60% B-PER II (Thermo Scientific) supplemented with Pierce Protease Inhibitor tablets (Thermo Scientific), Lysozyme (Sigma-Aldrich) and DNase I (Sigma-Aldrich) at a ratio of 10 mL per 1 g of cell pellet. Cells were incubated and lysed for one hour at 30˚C with gentle

shaking. The lysate was clarified by 12,000 x g for 5 minutes at 4˚C in order to remove large cell debris.

The resulting supernatant, containing Pdu MCPs and other soluble proteins was transferred to a clean falcon tube and centrifuged at 20,000 x g at 4˚C for thirty-five minutes in order to pellet the MCPs. The excess supernatant, containing lower-molecular weight soluble proteins were disposed of. The resulting pelleted MCPs were translucent and difficult to see. We gently resuspended the MCP pellet in 400 µL of chilled buffer, with buffer composition defined by subsequent downstream experiments. For general transmission electron microscopy work, Pdu MCPs were resuspended in a buffer containing 50 mM Tris-HCl (pH 8.0), 50 mM KCl, 5 mM $MgCl_2$, and 1% 1,2-PD to a concentration of 0.5 mg/mL, referred to henceforth as Buffer B. For all other work, Pdu MCPs were resuspended in an amine-free buffer (KIT buffer) containing 100 mM triethyleammonium bicarbonate buffer pH 8.5 (Sigma-Aldrich), 50 mM KCl, 1% 1,2-propanediol and 10 mM $MgCl_2$.

Lysine modification

In order to determine the sidedness and orientation of BMC shell proteins, we performed a series of assays in order to covalently modify lumen and cytosol-facing lysines. Purified Pdu MCPs derived from the same parent biological sample were divided into two aliquots and used for modifying interior (lumen) and exterior (cytosol) facing lysines as described below.

For interior MCP lumen-facing lysines, we modified the previously described Diol dehydratase (DDH) activity assay [28,29] using 1 mL reactions, referred to henceforth as *DDH**. Purified Pdu MCPs in KIT buffer were diluted to 10 mg/mL. DDH* reactions contained the following; 50 µL of 10 mg/mL purified Pdu MCPs, 20 µL 60% 1,2-propanediol (Sigma-Aldrich), 20 µL sodium cyanoborohydride ($NaCNHB_3$) (Sigma-Aldrich), 133 µL of 10 mM Ado B_{12} (Sigma-Aldrich) and 777 µL of KIT buffer. Reactions were incubated at 37˚C for thirty minutes protected from light. Ado B_{12} was added last in order to initiate the reaction. Following the thirty-minute

incubation, samples were stored at -20˚C until ready for mass spectrometry analysis. We used three biological replicates and performed reactions in duplicate.

In order to modify external, cytosol-facing lysines, we added free propionaldehyde to purified Pdu MCPs. As with DDH*, we performed 1 mL reactions. Propionaldhyde reactions contained the following; 50 µL of 10 mg/mL purified Pdu MCPs (in KIT buffer), 20 µL of 99% propionaldehyde (Sigma-Alridch), 20 µL of $NaCNHB_3$ (Sigma-Aldrich) and 910 µL of KIT buffer. Reactions were incubated for thirty minutes at 37˚C and subsequently stored at -20˚C until ready for mass spectrometry analysis. We used three biological replicates.

Mass Spectrometry

Samples were prepared for liquid chromatography and tandem mass spectrometry. Briefly, samples were buffer exchanged to 100 mM ammonium bicarbonate butter, pH 8.0, denatured with 8 M urea, reduced with 20 mM DTT at 60˚ C and subsequently alkylated with iodoacetamide, protected from light. Samples were then digested with Trypsin and endoproteinase GluC (New England Bioolabs) to ensure sufficient cleavage of lysine-modified peptides. Peptides were dried and desalted using 3M Empore C18 solid phase extraction disks.

100 ng of sample were injected and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) using a nano-LC system (Ultimate 3000 RSL Cnano, Dionex) coupled to a Quadrupole Orbitrap mass spectrometer (Q-Exactive Plus, Thermo Scientific). The mass-spectrometer was operated in a data-dependent acquisition mode. A full MS scan was performed in the mass range of 300 – 1800 m/z. MS/MS scans were acquired with a dynamic mass range and precursor ions were isolated using a quadrupole isolation window of 2 m/z and fragmented in the HCD trap with a normalized collision energy.

Q Exactive *.RAW files were analyzed using ProteomeDiscoverer (version 1.4) employing Matrix Science's Mascot search algorithm (version 2.5). Peak lists were searched against common contaminants and the *Salmonella enterica* proteome. Critical search parameters include

lysine propylation. The relative abundance of lysine-modified sites were calculated by dividing the total peak area for all modified peptides containing the lysine-modification by the total peptide peak for the shell protein of interest (PduA, PduJ or PduU).

Biotinylation and Streptavidin conjugation

In order to probe the sidedness of the PduA, PduJ and PduU BMC shell proteins, we developed a biotinylation and gold-conjugated streptavidin conjugation assay. AviTag-harboring Pdu MCP mutants were purified as previously described resuspended in Buffer B. We performed site-specific biotinylation using biotin ligase (BirA) using a modified protocol as previously described [30]. For each mutant, we prepared biotinylation reactions by mixing 400 µL of mutant Pdu MCPs at 6.5 mg/mL, 5 µL of 1M $MgCl_2$, 20 µM of 100 mM ATP (Sigma-Aldrich), 50 µL of BirA and 3 µL of 50 mM Biotin (Fischer Scientific). Reactions were incubated for one hour at room temperature. An additional 5 µL of 1M $MgCl_2$, 20 µM of 100 mM ATP (Sigma-Aldrich), 50 µL of BirA and 3 µL of 50 mM Biotin were added to the reaction and incubated for one more hour. Reactions were subsequently dialyzed into Buffer B to remove excess biotin.

Following dialysis, samples were incubated for fifteen minutes with Buffer B-equilibrated nickel resin (HisPur™ Ni-NTA Resin, Thermo Scientific) in order to capture and remove BirA. Samples were centrifuged for 12,000 x g at 4˚C for five minutes to pellet the nickel, and the resulting supernatant containing biotinylated mutant Pdu MCPs were carefully collected and transferred to a separate Eppendorf tube. Biotinylated Pdu MCPs were mixed with 20nm gold-conjugated streptavidin (Ted Pella) for one hour and imaged with transmission electron microscopy as described below.

Transmission electron microscopy

For negative stain transmission electron microscopy, Formvar/Carbon 300 mesh copper grids (Ted Pella) were glow discharged for forty-five seconds. Purified MCPs were diluted to approximately 0.5 mg/mL and 5 µL were applied to the grid and allowed to incubate for sixty

seconds. Excess sample was blotted away with filter paper, washed twice with 5 µL of 0.1% 1,2-propanediol and stained with 5 µL 2% uranyl acetate for thirty seconds. Excess uranyl acetate was wicked away and allowed to dry. Grids were imaged using an FEI Tecnai T12 transmission electron microscope.

For Cryo-electron microscopy (Cryo-EM) we diluted Pdu MCPs to a final concentration of 5 mg/mL. 2.5 µL of Pdu MCPs were pipetted onto Quantifoil R 1.2/2.3 200 mesh grids (Electron Microscropy Sciences). Grids were blotted and frozen in liquid ethane using a Vitrobot Mark IV (FEI) and stored for under liquid nitrogen until ready for data collection. Pdu MCPs were screened to ice thickness, stability and distribution using an FEI T20 iCorr microscope equipped with a Gatan US4000 CCD Camera or using an FEI TF20 microscope equipped with a bottom mount TVIPS F416, 4K x 4K CMOS camera.

We prepared samples for Cryo-electron tomography (Cryo-ET) by premixing purified Pdu MCPs with gold fiducial markers. MCPs were mixed with 10 nm BSA gold conjugate (Ted Pella) or 20 nm streptavidin gold conjugate (Ted Pella) and stored on ice. 2.5 µL of the MCP-fiducial mixture were pipetted onto Quantifoil R 1.2/2.3 200 mesh grids (Electron Microscropy Sciences). Grids were blotted and frozen in liquid ethane using a Vitrobot Mark IV (FEI) and stored for under liquid nitrogen until ready for data collection.

Image acquisition and processing

Preliminary tilt-series were collected on a 300 keV Tecnai TF-30 equipped with a single-axis tilt holder and a 4k x 4k CMOS TVIPS camera. We collected unidirectional projections at 29,000x magnification using 2 x 2 binning which corresponded to a pixel size of 0.59 nm. All data sets had a tilt range of +/- 40˚ in 0.3˚ increments with 1 second exposures with a target total dose of ~80 e per $Å^2$. Following data collection, we used the IMOD software package [31] to generate preliminary tomographic reconstructions.

Preliminary results

The endogenous Pdu metabolic pathway can be hijacked and upregulated in order to probe propionaldehyde-based lysine modification

We modified the Diol dehydratase (DDH) activity assay [28,29] in order to exploit the native Pdu metabolic pathway and overproduce the propionaldehyde intermediate. The propanediol dehydratase activity can be quantified by indirectly measuring the formation of propionaldehyde through monitoring the conversion of NADH to NAD^+. This assay requires a fixed concentration of purified Pdu MCPs and 1,2-propanediol and NADH, and also requires an excess of Ado B_{12} and *Saccharomyces cerevisiae* alcohol dehydrogenase. Under these conditions, the 1,2-propanediol in solution is readily converted into its final products propionyl-PO_4^{2-} and 1-propanol. In this work, we excluded NADH and alcohol dehydrogenase which led to the overproduction and accumulation of propionaldehyde within the Pdu MCP lumen. The propionaldehyde intermediate is reactive and capable of modifying lysines via propyl modifications (Schiff base) and hemi-thioacetal modifications. We used tandem mass-spectrometry to search for these propionaldehyde-based lysine modifications and identify 1) which lysines of PduA, PduJ and PduU were modified and 2) determine the location of these lysines relative to the flat or concave face using existing structural data.

Previous crystallographic work on PduA (PDB 3NGK [14]), PduJ (PDB 5D6V [11]) and PduU (PDB 3CGI [12]) has revealed the location of lysines relative to their flat and dimpled faces (Table 7.1). PduA has three lysines located on its flat face (K12, K55 and K72) and three lysines on its dimpled face (K37, K86, K90). PduJ has three lysines located on its flat face (K11, K54 and K71) and two lysines on its dimpled face (K36, K89). Lastly, PduU has two lysines located on its flat face (K18, K96) and three lysines located on its dimpled face (K31, K35, K36).

Upon adding excess propionaldehyde to purified MCPs or performing the DDH* reaction and overproducing the propionaldehyde intermediate, we were able to successfully identify propyl modifications of lysines using tandem mass-spectrometry (LC-MS/MS). We used recombinantly purified N-Avi-PduA, N-Avi-PduJ or N-Avi-PduU incubated with excess propionaldehyde (20 µL of 99% propionaldehyde to 50 µL of 10 mg/mL recombinantly purified PduA, J or U) as controls in order to assess the relative reactivity of the various BMC lysines. Despite successfully identifying lysine-harboring peptides with LC-MS/MS, we only observed modifications of flat-face-located lysines. Because lysine propyl modifications are reversible (Schiff base), we decided to repeat these experiments with $NaCNHB_3$ in order to reduce the reaction and generate covalent modifications. This work resulted in detectable lysine modifications on both the flat and dimpled faces of recombinantly purified PduA, PduJ and PduU. Moreover, for purified intact Pdu MCPs, we observed no baseline propyl modification in the absence of $NaCNHB_3$. In the presence of $NaCNHB_3$, we observed some propyl-based lysine modifications in purified intact Pdu MCPs, however these modifications were several orders of magnitude lower than conditions in which we added or overproduced (DDH*) propionaldehyde.

Preliminary mass spectrometry data suggests that when exogenous propionaldehyde is added to purified Pdu MCPs, there are significant differences in proportions of propyl-modified lysines (Figure 7.2). We performed a two-sided Mann-Whitney U test and found that a greater proportion ($p<0.5$) of flat-face located lysines (K55 and K12 for PduA and K11 for PduJ) are propylated upon addition of excess propionldehyde. For K55 of PduA, approximately 46% of K55-harboring peptides possess a propyl modification when propionaldehyde is added to Pdu MCPs (Figure 7.2a). For the same lysine, only 40% of K55-harbring peptides possessed a propyl modification when propionaldehyde is overproduced using the afore-mentioned modified diol dehydratase assay (Figure 7.2a). The same trend is observed for K12 of PduA and K11 of PduJ (Figure 7.2).

1,2-Propanediol microcompartments are robust, flexible and dynamic

We found that the Pdu MCP tolerates the addition of the fifteen-residue AviTag to the N-terminus of three BMC shell proteins; PduA, PduJ and PduU. We successfully inserted the AviTag to the N-terminus of PduA, PduJ and PduU, resulting in three separate classes of mutant MCPs. During expression and purification, *N-Avi-PduA*, *N-Avi-PduJ* and *N-Avi-PduU* mutants behaved like their wild-type counterparts. We were able to obtain comparable yields for all mutants. Moreover, these Pdu MCPs had no obvious differences in shape, distribution and morphology compared to their wild-type counterparts (Figure 7.3). Under negative stain TEM, we found that *N-Avi-PduA-MCP*, *N-Avi-PduJ-MCP* and *N-Avi-PduU-MCP* possessed characteristic features including flat, well-defined facets and prominent vertices. We were also able to observe internal cargo and enzymes.

Pdu MCPs tolerate a variety of additives with no obvious effects on overall shape or morphology. Our lysine modification-based experiments required the addition of several reagents including $NaCNHB_3$, propionaldehyde, Ado B_{12} and an excess of 1,2-propadediol (components of DDH*). In order to ensure the integrity of our Pdu MCPs, we incubated purified Pdu MCPs with comparable concentrations of these key reagents in excess of one hour. We proceeded to image these samples with negative stain TEM and observed no aberrant MCPs. Under these conditions, Pdu MCPs maintained their shape and integrity with no obvious signs of degradation (Figure 7.4).

The Pdu MCP exists in a variety of shapes and morphologies. Using negative stain TEM we were able to clearly see the boundaries of Pdu MCPs (Figure 7.5a). Internal cargo and enzymes are readily observed, appearing as prominent dark spots within the MCP. While some Pdu MCPs appear amorphous with irregular bounds, others have sharp, well-defined facets and vertices. Their native shape and internal cargo are better maintained under cryogenic conditions (Figure 7.5b). Unlike their negative stain counterparts, nearly all Pdu MCPs retain sharp, clearly defined bounds with cryo-ET. In several instances, these two-dimensional cross-sections reveal

MCPs with five sides and MCPs with six sides (Figure 7.5, Figure 7.6). Moreover, we observed minority subpopulations of rounded, spherical MCPs as well as elongated rods. The elongated rods are roughly 75 nm wide and between 500 – 800 nm long. While we could clearly resolve internal enzymes, not all Pdu MCPs possessed obviously visible cargo (Figure 7.5, Figure 7.6).

Pdu MCPs do not require internal enzymes for assembly. In numerous instances, we observed cargo-less Pdu MCPs that appeared to be devoid of internal enzymes (Figure 7.7). MCPs lacking their internal enzymes did not have obvious differences in overall shape or morphology compared to other MCPs. In some instances, Pdu MCPs had darker regions within the bounds of the MCP (Figure 7.7a). These regions are not readily discernable with cryo-EM.

Discussion

Preliminary data suggests that BMC shell proteins assemble with their concave faces oriented inward, towards the lumen. We found interesting evidence that appears to suggest a preference for modification of flat-face located lysines upon addition of excess propionaldehyde. Intriguingly, we also observed a trend that a greater proportion of K89 on the dimpled face of PduA and K36 on the dimpled face of PduJ are found to have propyl modifications when propionaldehyde is overproduced (*DDH) than when propionaldehyde is exogenously added to solution (Figure 7.2). This ongoing work will utilize additional technical and biological replicates in order to confirm our preliminary data.

We observed several instances of apparently empty MCPs, or rather, MCPs with no obvious internal features or enzymes (Figure 7.7). This observation may challenge current hypothesis on assembly principles [32]. Current theories propose two alternative modes of MCP assembly. The first suggests a core-first assembly that involves the aggregation of internal enzymes and their subsequent encapsulation by the MCP shell [32]. Alternatively, the concomitant assembly hypothesis suggests that enzymatic core and external shell proteins

assemble in parallel, thus requiring the presence of internal enzymes [32]. The observance of empty MCPs contradicts the theory that internal enzymes and cargo are required for assembly. Previous work also demonstrated [cite paper where the operon is added *to e. coli* and expressed, no obvious changes in morphology or expression levels. This work supports previous work by demonstrating that endogenously produced Pdu MCPs are indeed capable of formation absent of internal cargo.

Pdu MCPs are more regular than previously observed. We attribute the well-defined facets and bounds of our Pdu MCP samples to the minimal processing and handling of our samples. Previous work includes using glutaraldehyde as a fixative [33] which may have contributed to additional deformation. We found that we were able to efficiently image our samples without additional manipulation prior to staining with uranyl acetate. Large concentrations of negative stain, as seen by dark spots on the MCPs have been attributed to pools of negative stain that sit above collapsed and deflated MCPs [33]. We hypothesize that these prominent dark spots are the result of large areas and concentrations of internal enzymes (Figure 7.6a).

Despite the observation that a majority of TEM-imaged Pdu MCPs had will defined vertices and facets, we found several aberrant MCPs with diversity in shape and morphology. As a minor population, elongated Pdu MCP rods ranging from approximately 500 – 800 nm in all purifications. This elongated rod morphology was previously observed . As an imperfect biological system that contains numerous BMC paralogs with essential and redundant roles [11,34], their variety in shape, size and morphology is unsurprising.

While recent advances have propelled this field forward, MCPs remain poorly understood from a functional, structural organizational level. Mechanisms of transport across a protein shell have only been partially revealed. Despite being closed systems that prevent the efflux of toxic intermediates, our work demonstrates that Pdu MCPs appear dynamic and subject to bi-

directional passive diffusion of propionaldehyde. Moreover, it is unclear how the numerous hexameric shell paralogs of MCPs are organized and distributed across the outer shell. Our future work will seek to elucidate the long-range organization of the Pdu MCP shell by utilizing subtomogram averaging and existing crystallographic data on Pdu BMC shell proteins. Additionally, we will utilize AviTag-harboring MCP mutants in order to perform site-specific biotinylation and optimize assays that seek to use gold-conjugated streptavidin in order to label external-facing AviTags and probe the orientation and topology of BMC shell proteins. This ongoing work will enhance our foundational knowledge of bacterial microcompartments confirming and perhaps revealing yet unknown principles of their assembly, functions and structure.

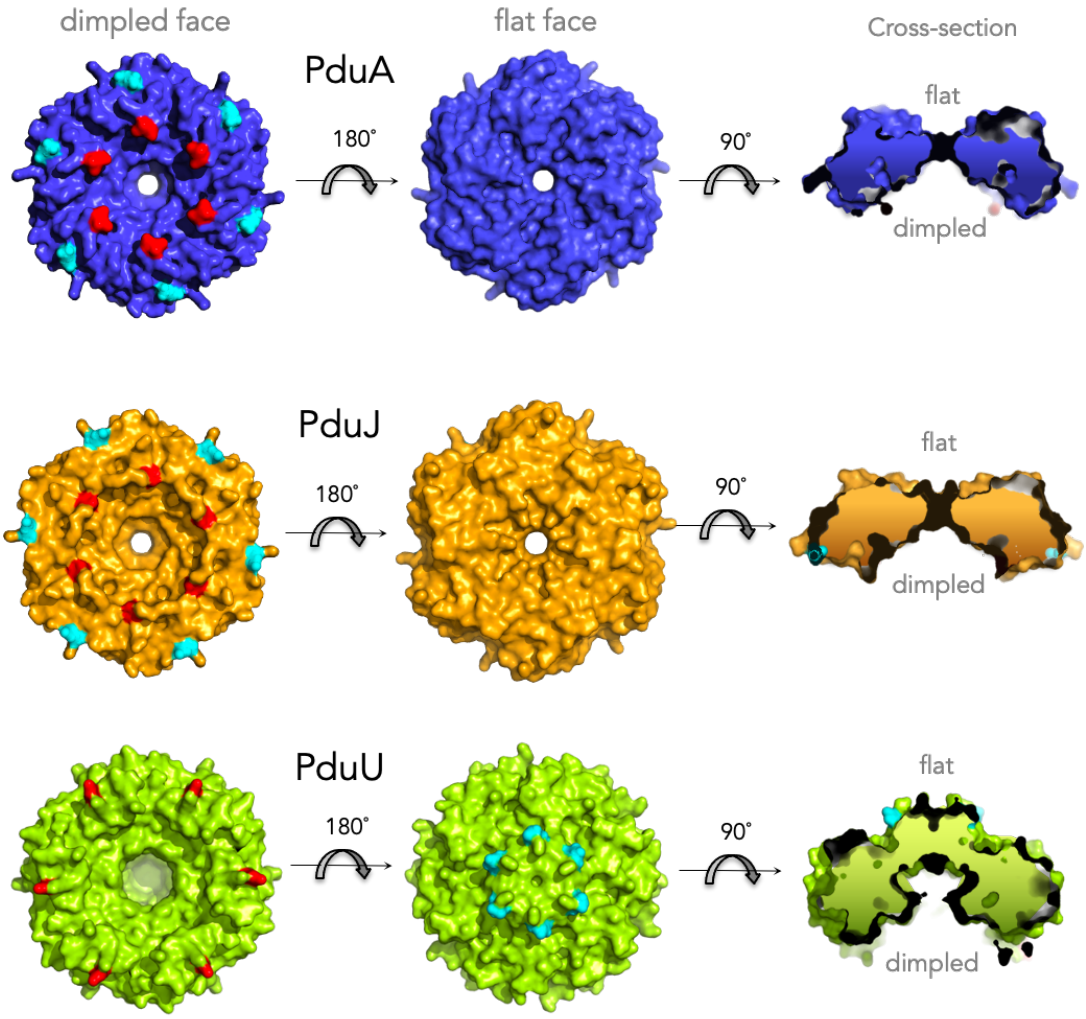


Figure 7.1

Structure of Pdu BMC shell proteins. PduA (cobalt), PduJ (orange) and PduU (lime green) are hexameric shell proteins from the 1,2-propanediol MCP. All three possess a flat and dimpled face. We identify the location of their termini highlighting their location relative to their flat and dimpled faces (N-terminus in cyan, C-terminus in red).

	Flat	Dimpled
PduA	K12 K55 K72	K37 K86 K90
PduJ	K11 K54 K71	K36 K89
PduU	K18 K96	K31 K35 K36

Table 7.1

Location of lysines relative to the flat and dimpled faces. We highlight the locaion of flat and dimpled-face located lysines for PduA, PduJ and PduU.

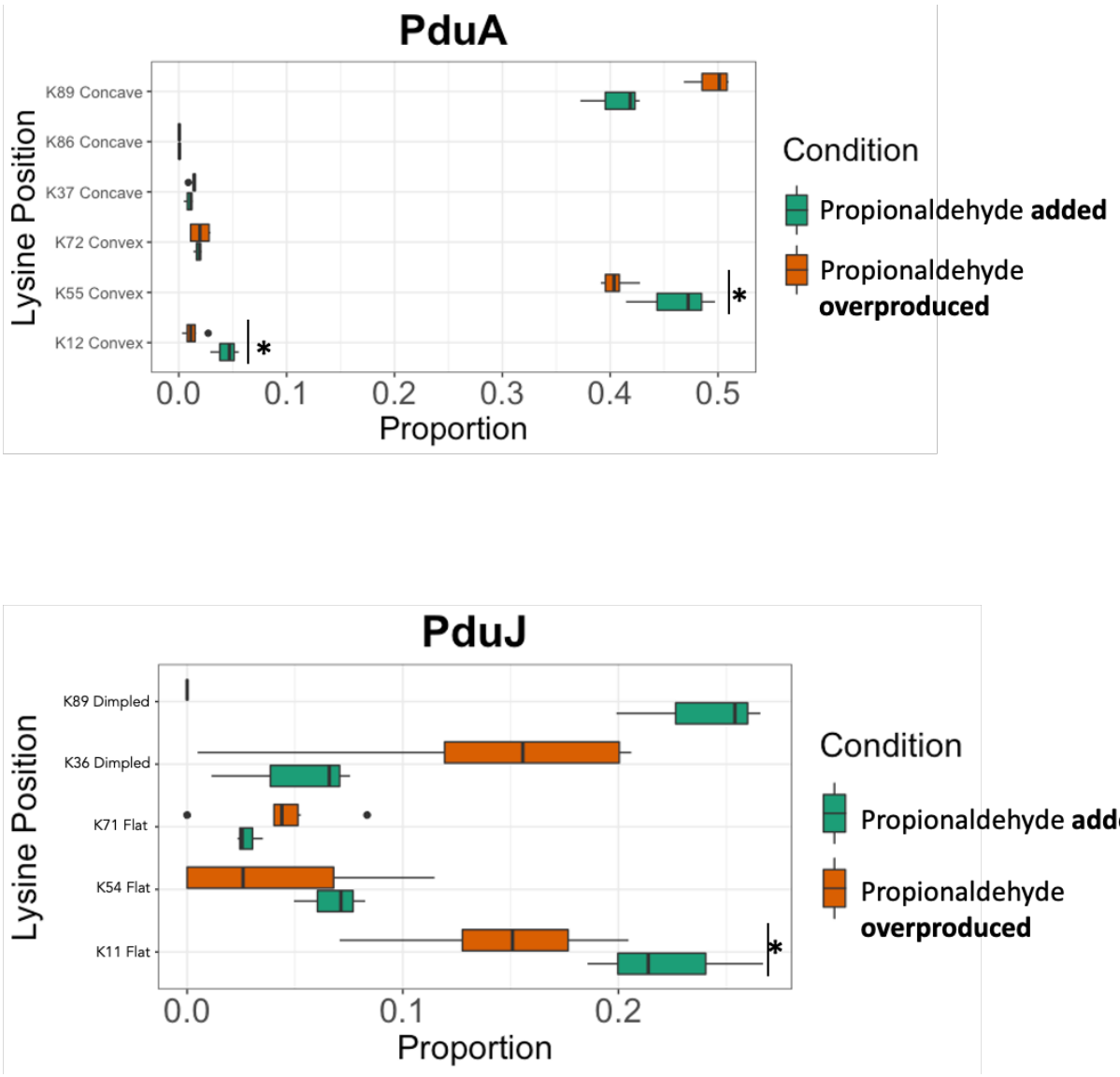


Figure 7.2

Proportion of modified lysines on PduA and PduJ. We compare the proportion of modified lysines on PduA and PduJ when exogenous propionaldehyde is added compared to when endogenous propionaldehyde is overproduced. Significance ($p<0.05$) is denoted with an asterisk (*).

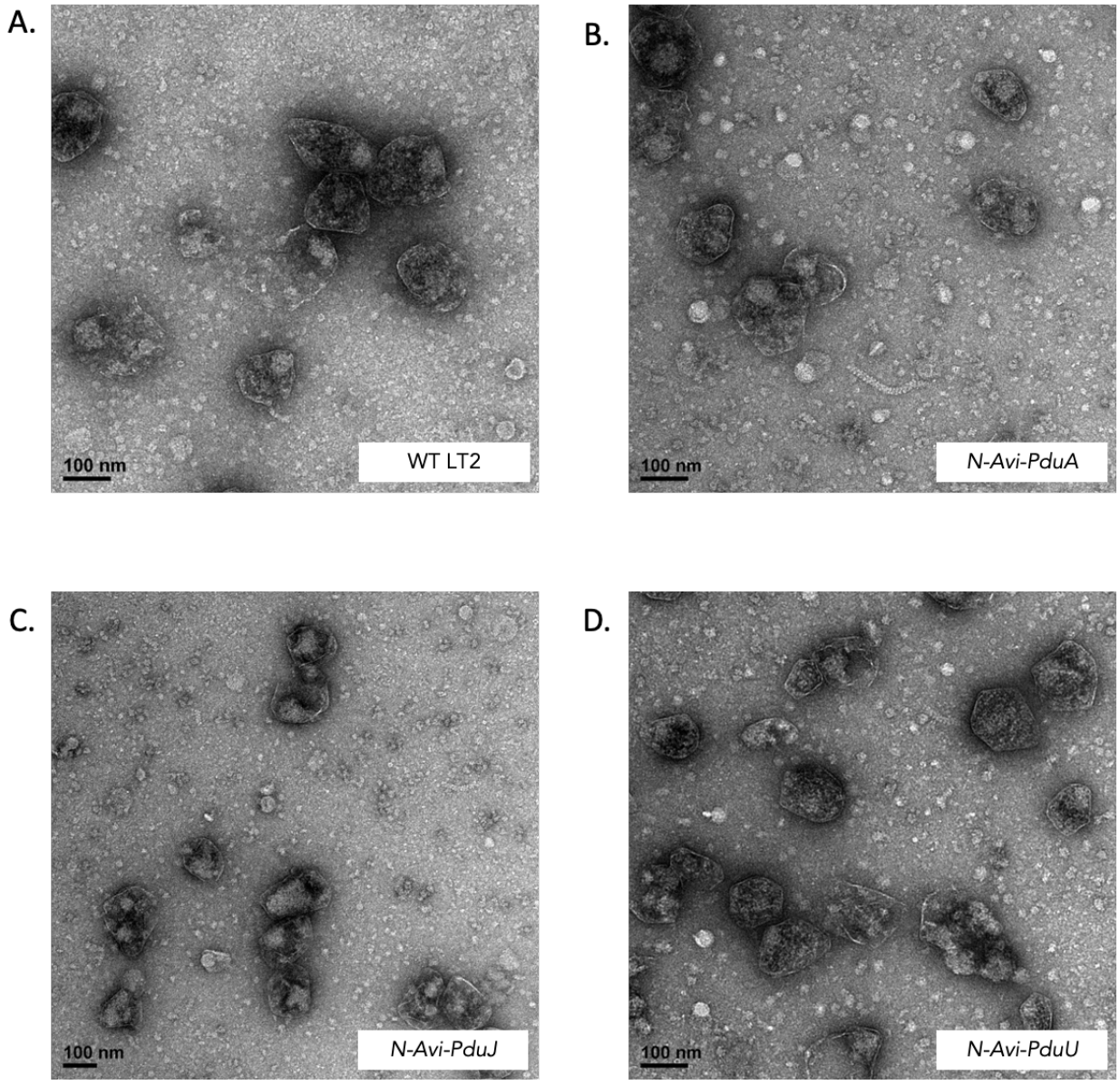


Figure 7.3

AviTag-bearing mutant Pdu MCPs. WT Pdu MCPs (A) compared to mutant MCPs harboring the fifteen-residue AviTag to the N-terminus of PduA (B), PduJ (C) and PduU (D) does not appear to alter the overall morphology of Pdu MCPs.

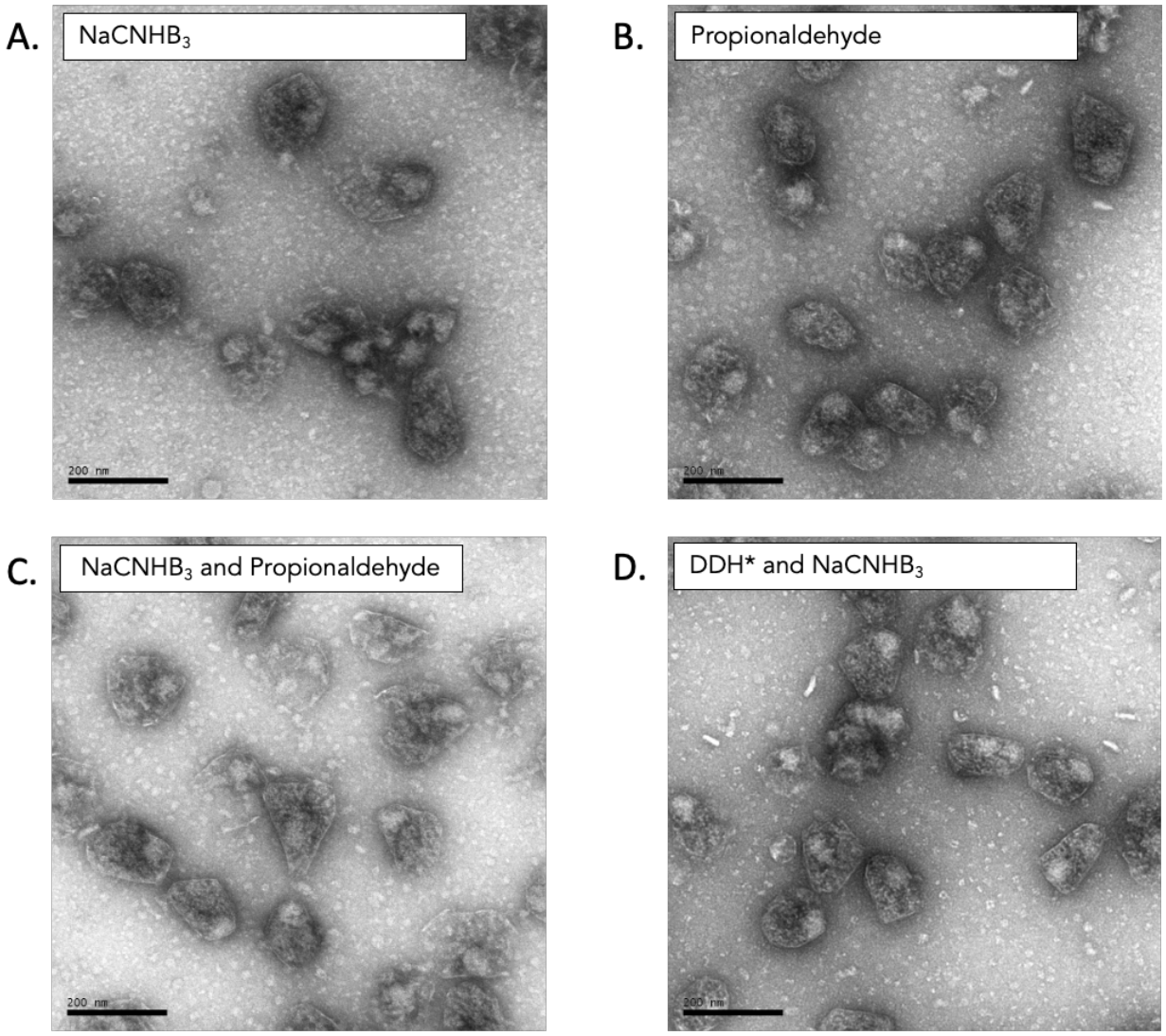


Figure 7.4

Pdu MCPs incubated with numerous reagents. The following reagents are conditions were used for probing lysine modification in subsequent work; (A) $NaCNHB_3$, (B) propionaldehyde, (C) $NaCNHB_3$ and propionaldehyde, and (D) $NaCNHB_3$ and DDH* reagents (Ado B_{12}, and 1,2-propanediol).

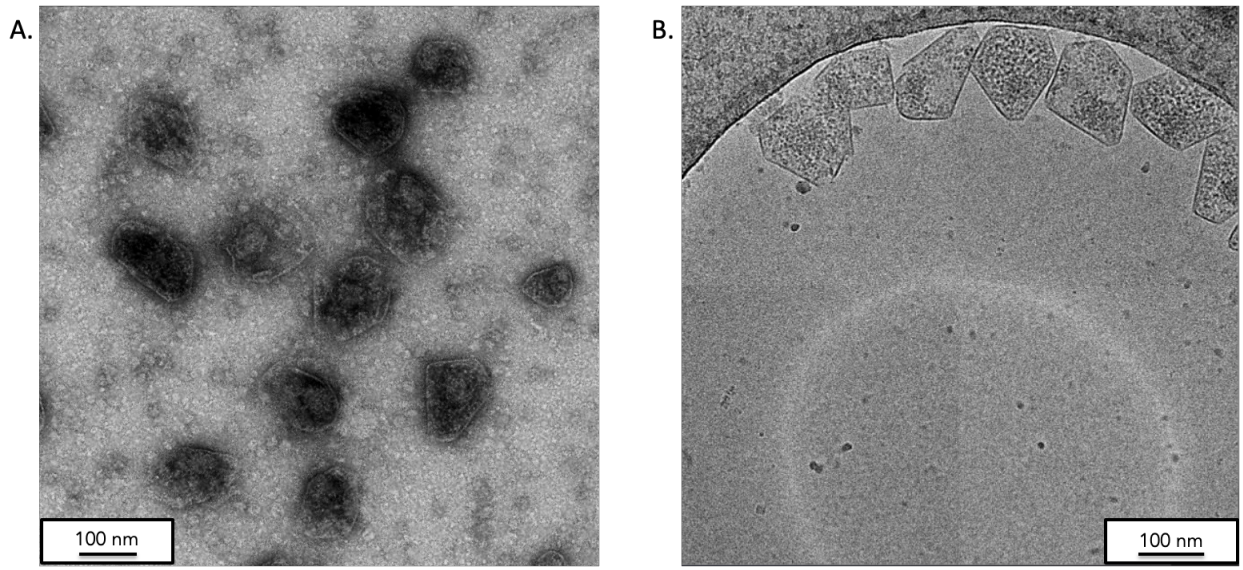


Figure 7.5

TEM of Pdu MCPs. Negative stain (A) and cryo-EM (B) of Pdu MCPs. Scale bar is 100nm.

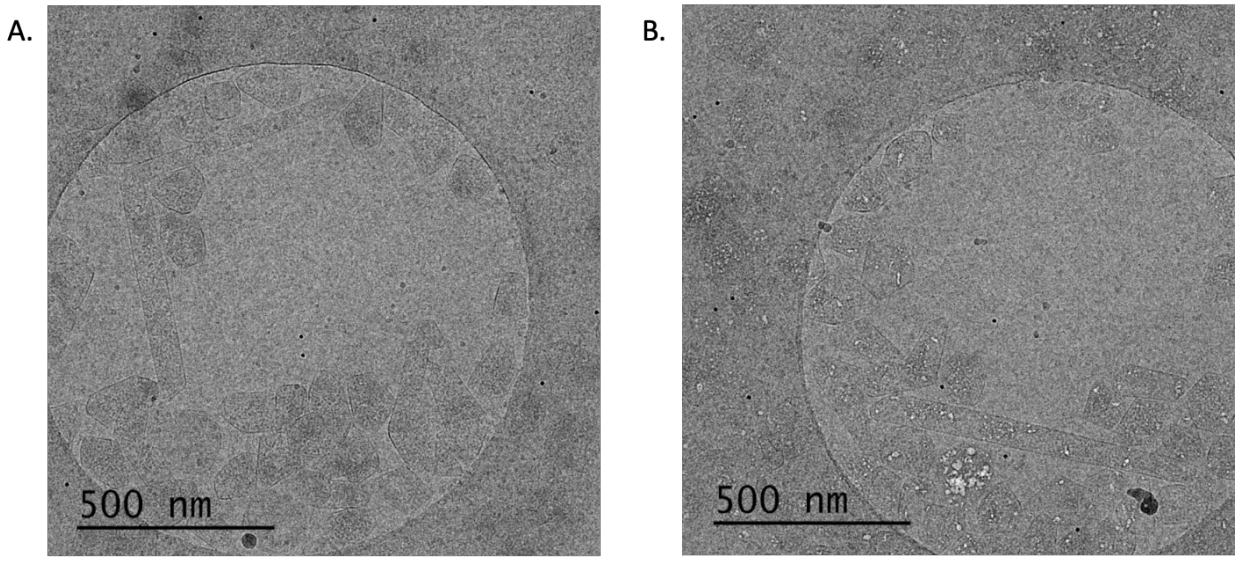


Figure 7.6

Cryo-EM of Pdu MCPs. Two micrographs revealing variety in shape size and morphology, especially highlighting elongated rods. Scale bar = 500 nm.

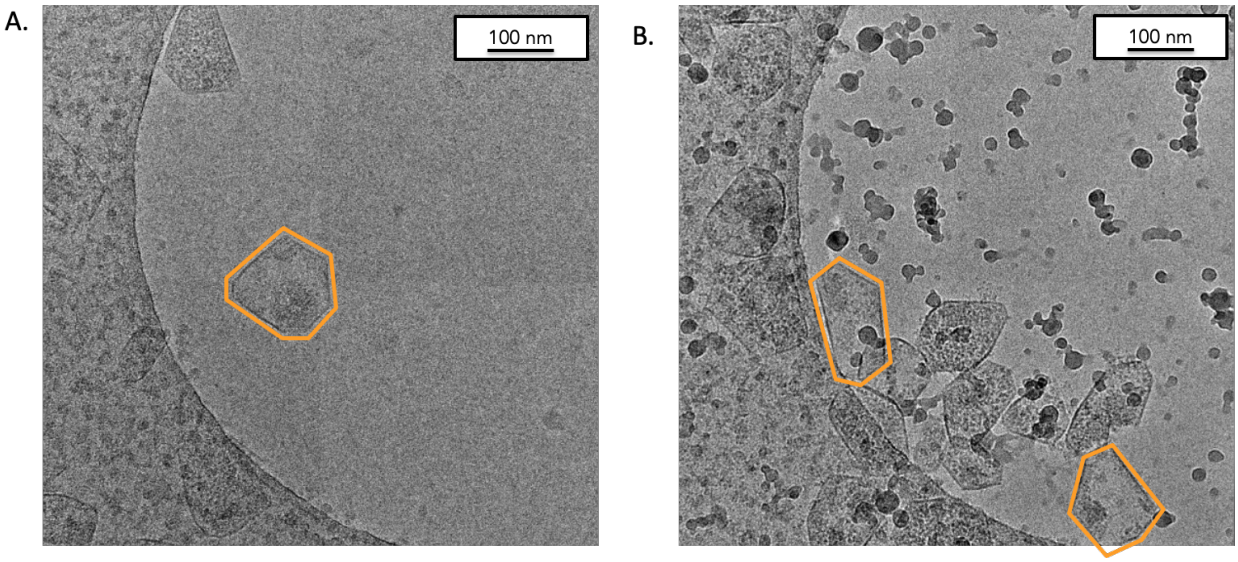


Figure 7.7

Cryo-EM of empty Pdu MCPs. Two micrographs highlighting empty MCPs, devoid of internal cargo and enzymes. Scale bar = 100nm.

Supporting Information

Oligo sequences

N-Avi-PduA-MCP:

CTGATGCTCAACAGCAAGTCAGCCATTCTTTTTCCTCGCATCTTCTTATAGTCCCAACTA

TCGGAACACTCCATGCGAGGTCTTTATGTCAGACGGCTTGAACGACATTTTTGAAGCCC

AAAAGATTGAATGGCATGAGGGCTCTATGCAACAAGAAGCACTAGGAATGGTAGAAACC

AAAGGCTTAACCGCAGCCATAGAGGCCGCTGATGCAATGGTTAAGTCAGCCAATGTGAT

GTTAGTGGGCTATGAAAAGATTGGCTCCGGGCTGGTAACCGTCATCGTGCGCGGCGAT

GTTGGCGCGGTCAAAGCGGCCACCGATGCAGGTGC

N-Avi-PduJ-MCP:

CGCATCGGCGCCGCTTTTTACGCTGATGCATCATCAGGACAGTCAGGCCCAACGCAAC

ACCGGTAATAACGCGGCACGGCTGGTCAAAGGGATCCCCTTTCGGGATCTCCATGCTT

AATCACAGGAGAACGGCAGTATGTCAGACGGCTTGAACGACATTTTTGAAGCCCAAAAG

ATTGAATGGCATGAGGGCTCTATGAATAACGCACTGGGACTGGTTGAAACAAAAGGGCT

GGTCGGCGCCATTGAAGCCGCCGATGCAATGGTTAAATCCGCCAACGTACAGCTGGTG

GGCTACGAAAAAATTGGTTCTGGCCTGGTGACCGTCAT

N-Avi-PduU-MCP:

CGTTTCTGACGTCAATAACGCCGTGACGGTTGCCAGCGAAAGCGCGGGCGAGAAAGGG

TTGTTGGTTTACCGTTCGGTGATCCCACGCCCGCATGAAGCCATGTGGCGACAGATGGT

GGAGGGGTAATGACTGAGGGACTGAACGATATTTTCGAAGCACAGAAAATCGAGTGGC

ACGAGGGATCTATGGAAAGACAACCGACAACGGATCGCATGATTCAGGAATACGTGCC

GGGGAAACAGGTCACTCTCGCGCACCTGATTGCTAATCCAGGGAAAGATCTCTTTAAGA

AGCTGGGCCTGCAGGATGCAGTGTCCGCCATTGGCATC

References

1. Bobik TA, Havemann GD, Busch RJ, Williams DS, Aldrich HC: **The Propanediol Utilization (pdu) Operon ofSalmonella enterica Serovar Typhimurium LT2 Includes Genes Necessary for Formation of Polyhedral Organelles Involved in Coenzyme B12-Dependent 1,2-Propanediol Degradation**. *J Bacteriol* 1999, **181**:5967–5975.

2. Bobik TA, Xu Y, Jeter RM, Otto KE, Roth JR: **Propanediol utilization genes (pdu) of Salmonella typhimurium: three genes for the propanediol dehydratase.** *J Bacteriol* 1997, **179**:6633–6639.

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

4. 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**. *Journal of Bacteriology* 1979, **139**:39–47.

5. Leal NA, Havemann GD, Bobik TA: **PduP is a coenzyme-a-acylating propionaldehyde dehydrogenase associated with the polyhedral bodies involved in B12-dependent 1,2-propanediol degradation by Salmonella enterica serovar Typhimurium LT2**. *Arch Microbiol* 2003, **180**:353–361.

6. Liu Y, Leal NA, Sampson EM, Johnson CLV, Havemann GD, Bobik TA: **PduL Is an Evolutionarily Distinct Phosphotransacylase Involved in B12-Dependent 1,2-Propanediol Degradation by Salmonella enterica Serovar Typhimurium LT2**. *J Bacteriol* 2007, **189**:1589–1596.

7. Cheng S, Bobik TA: **Characterization of the PduS Cobalamin Reductase of Salmonella enterica and Its Role in the Pdu Microcompartment**. *J Bacteriol* 2010, **192**:5071–5080.

8. Cheng S, Fan C, Sinha S, Bobik TA: **The PduQ Enzyme Is an Alcohol Dehydrogenase Used to Recycle NAD+ Internally within the Pdu Microcompartment of Salmonella enterica**. *PLOS ONE* 2012, **7**:e47144.

9. Havemann GD, Sampson EM, Bobik TA: **PduA Is a Shell Protein of Polyhedral Organelles Involved in Coenzyme B12-Dependent Degradation of 1,2-Propanediol in Salmonella enterica Serovar Typhimurium LT2**. *J Bacteriol* 2002, **184**:1253–1261.

10. Pang A, Frank S, Brown I, Warren MJ, Pickersgill RW: **Structural Insights into Higher Order Assembly and Function of the Bacterial Microcompartment Protein PduA**. *J Biol Chem* 2014, **289**:22377–22384.

11. Chowdhury C, Chun S, Sawaya MR, Yeates TO, Bobik TA: **The function of the PduJ microcompartment shell protein is determined by the genomic position of its encoding gene**. *Molecular Microbiology* 2016, **101**:770–783.

12. Crowley CS, Sawaya MR, Bobik TA, Yeates TO: **Structure of the PduU Shell Protein from the Pdu Microcompartment of Salmonella**. *Structure* 2008, **16**:1324–1332.

13. Pang A, Warren MJ, Pickersgill RW: **Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe–4S cluster-binding site**. *Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr* 2011, **67**:91–96.

14. Crowley CS, Cascio D, Sawaya MR, Kopstein JS, Bobik TA, Yeates TO: **Structural Insight into the Mechanisms of Transport across the Salmonella enterica Pdu Microcompartment Shell**. *J Biol Chem* 2010, **285**:37838–37846.

15. Sinha S, Cheng S, Fan C, Bobik TA: **The PduM protein is a structural component of the microcompartments involved in coenzyme B12-dependent 1,2-propanediol degradation by Salmonella.** *J Bacteriol* 2012, doi:10.1128/JB.06529-11.

16. Tanaka S, Kerfeld CA, Sawaya MR, Cai F, Heinhorst S, Cannon GC, Yeates TO: **Atomic-Level Models of the Bacterial Carboxysome Shell**. *Science* 2008, **319**:1083–1086.

17. Tsai Y, Sawaya MR, Cannon GC, Cai F, Williams EB, Heinhorst S, Kerfeld CA, Yeates TO: **Structural Analysis of CsoS1A and the Protein Shell of the Halothiobacillus neapolitanus Carboxysome**. *PLoS Biol* 2007, **5**.

18. Fan C, Cheng S, Sinha S, Bobik TA: **Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments**. *PNAS* 2012, **109**:14995–15000.

19. Jorda J, Leibly DJ, Thompson MC, Yeates TO: **Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein**. *Chem Commun* 2016, **52**:5041–5044.

20. Sutter M, Greber B, Aussignargues C, Kerfeld CA: **Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell**. *Science* 2017, **356**:1293–1297.

21. Sutter M, Laughlin TG, Sloan NB, Serwas D, Davies KM, Kerfeld CA: **Structure of a Synthetic β-Carboxysome Shell**. *Plant Physiology* 2019, **181**:1050–1058.

22. Kalnins G, Cesle E-E, Jansons J, Liepins J, Filimonenko A, Tars K: **Encapsulation mechanisms and structural studies of GRM2 bacterial microcompartment particles**. *Nature Communications* 2020, **11**.

23. Schmid MF, Paredes AM, Khant HA, Soyer F, Aldrich HC, Chiu W, Shively JM: **Structure of Halothiobacillus neapolitanus Carboxysomes by Cryo-electron Tomography**. *Journal of Molecular Biology* 2006, **364**:526–535.

24. Iancu CV, Ding HJ, Morris DM, Dias DP, Gonzales AD, Martino A, Jensen GJ: **The Structure of Isolated Synechococcus Strain WH8102 Carboxysomes as Revealed by Electron Cryotomography**. *Journal of Molecular Biology* 2007, **372**:764–773.

25. Ochoa JM, Nguyen VN, Nie M, Sawaya MR, Bobik TA, Yeates TO: **Symmetry breaking and structural polymorphism in a bacterial microcompartment shell protein for choline utilization**. *Protein Science* 2020, **29**:2201–2212.

26. Datsenko KA, Wanner BL: **One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products**. *PNAS* 2000, **97**:6640–6645.

27. Sinha S, Cheng S, Sung YW, McNamara DE, Sawaya MR, Yeates TO, Bobik TA: **Alanine Scanning Mutagenesis Identifies an Asparagine–Arginine–Lysine Triad Essential to Assembly of the Shell of the Pdu Microcompartment**. *Journal of Molecular Biology* 2014, **426**:2328–2345.

28. Chowdhury C, Chun S, Pang A, Sawaya MR, Sinha S, Yeates TO, Bobik TA: **Selective molecular transport through the protein shell of a bacterial microcompartment organelle**. *PNAS* 2015, **112**:2990–2995.

29. Fan C, Bobik TA: **The N-Terminal Region of the Medium Subunit (PduD) Packages Adenosylcobalamin-Dependent Diol Dehydratase (PduCDE) into the Pdu Microcompartment**. *Journal of Bacteriology* 2011, **193**:5623–5628.

30. Fairhead M, Howarth M: **Site-specific biotinylation of purified proteins using BirA**. *Methods Mol Biol* 2015, **1266**:171–184.

31. Mastronarde DN, Held SR: **Automated tilt series alignment and tomographic reconstruction in IMOD**. *Journal of Structural Biology* 2017, **197**:102–113.

32. Kerfeld CA, Aussignargues C, Zarzycki J, Cai F, Sutter M: **Bacterial microcompartments**. *Nature Reviews Microbiology* 2018, doi:10.1038/nrmicro.2018.10.

33. Kennedy NW, Hershewe JM, Nichols TM, Roth EW, Wilke CD, Mills CE, Jewett MC, Tullman-Ercek D: **Apparent size and morphology of bacterial microcompartments varies with technique**. *PLOS ONE* 2020, **15**:e0226395.

34. Kennedy NW, Ikonomova SP, Slininger Lee M, Raeder HW, Tullman-Ercek D: **Self-assembling Shell Proteins PduA and PduJ have Essential and Redundant Roles in Bacterial Microcompartment Assembly**. *Journal of Molecular Biology* 2021, **433**:166721.