ABSTRACT

Title of Dissertation:	THE ASSEMBLY STATE OF A BACTERIAL MICROCOMPARTMENT AND ITS RELATED STRUCTURES
	Daniel Stephen Trettel, Doctor of Philosophy, Biochemistry, 2022
Dissertation directed by:	Professor Wade Winkler, Department of Cell Biology and Molecular Genetics

Bacterial microcompartments (BMCs) are polyhedral, protein-based organelles present in a wide range of bacteria. This mode of compartmentalization is highly modular and can accommodate a wide range of chemistries within them, including carbon fixation. These aspects make them a promising target to serve as bioplatforms for commodity chemical synthesis and enhanced carbon fixation. However, it is challenging to investigate the structure and function of BMCs using classical methods. As such, the native structure of BMCs remains largely enigmatic, hampering their synthetic adaption. This dissertation addresses these concerns by describing the assembly state of the model 1,2-propanediol (Pdu) BMC using a variety of approaches. Chemical probing reveals the Pdu BMC is surprisingly permeable to and permissive of derivatization. This insight enabled application of crosslinking mass spectrometry to describe its protein interactome. The interactome map reveals that small domains called encapsulation peptides dominate interior interactions while reporting on the

organization of the outer protein shell. Laser scanning confocal approaches were developed to study the solution behavior of BMCs. These experiments heavily suggest that the Pdu BMC is a dynamic entity that exchanges protein elements; a result we primarily attribute to the protein shell. These confocal microscopy approaches were further used to study the super-structures formed by individual shell proteins and to describe their interactions with one another. Together, the results from this project give important insight on the assembly state of the model Pdu BMC including its biogenesis, organization, and behavior. These data answer some of the open questions concerning the assembly of BMC structures, which will help innovate the next generation of BMC-based biotechnology tools.

THE ASSEMBLY STATE OF A BACTERIAL MICROCOMPARTMENT AND ITS RELATED STRUCTURES

by

Daniel Stephen Trettel

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Biochemistry
2022

Advisory Committee:
Professor Wade Winkler, Chair
Associate Professor Peter Nemes
Associate Professor Jiqiang Ling
Associate Professor Jason Kahn
Associate Professor Ed Eisenstein, Dean's Representative

Acknowledgements

I would like to express my sincere gratitude towards my advisor, Dr. Wade Winkler. Wade gave me a rare treat as a graduate researcher – agency. The freedom to explore and govern myself as I saw fit was the key to my development into the researcher I am now. Wade continually encourages intellectual freedom, curiosity, and creativity and for that I am thankful.

Thank you to my dissertation committee composed of Dr.'s Jason Kahn, Peter Nemes, and Lanny Ling as well as my Dean's Representative Dr. Ed Eisenstein. Specifically, thank you Dr. Kahn for continuing to foster my fledgling interest in synthetic biology as well as Dr. Nemes for sharing his expertise in mass spectrometry.

I would like to thank the Georgetown Running Club for fostering my non-academic half during this journey. Athletics is a form of expression for me, and the Club not only gave me that outlet at a competitive level but also lifelong friends and memories.

Thank you to my lab-mates, particularly Tanner, Madi, and Conor. While we all work on very different projects, our goals are the same – to get Wade to order things in a timely manner from Quartzy.

Thank you to my parents, Mom and Dad, as well as my older brothers, Andrew and Ben. My brothers trailblazed the path while my parents gave me the freedom to explore it.

Finally, I am so grateful for my partner through this journey, Emily. Thank you for listening to my conspiracy-style rants about hexagons fitting together.

Table of Contents

List of Figures

List of Abbreviations

AFM	Atomic force microscopy
AlcDH	Alcohol Dehydrogenase
AldDH	Aldehyde dehydrogenase
BMC	Bacterial microcompartment
BMC-H	Hexameric shell protein
BMC-P	Pentameric shell protein
BMC-T	Trimeric shell protein
BS(PEG)5	PEGylated bis(sulfosuccinimidyl)suberate
DSBU	Disuccinimidyl dibutyric urea
DSSO	Disuccinimidyl sulfoxide
Eut	Ethanolamine
EP	Encapsulation peptide
GFP	Green fluorescent protein
LLPS	Liquid-liquid phase separation
Pdu	1,2-propanediol
RT	Room temperature
RubisCO	Ribulose-1,5-biphosphate carboxylase-oxygenase
Sulfo-EGS	Sulfo-ethylene glycol bis(sulfosuccinimidyl succinate)
TEM	Transmission electron microscopy

Articles Generated

The work present in this dissertation led to two primary manuscripts. Data from (1) is presented in Chapter 2, 3, and 4 with some in Chapter 5 as well. Manuscript (2) contains data primarily from Chapter 6 but also some from Chapter 5 too. Manuscript (3) contains data outside of this dissertation. A review article is also currently being prepared for submission, extending from a portion of Chapter 1.

1. Trettel, D.S., Resager, W., Jenkins, C.C., Ueberheide, B.M., Winkler, W.C. (2022) 'Chemical probing provides insight into the native assembly state of a bacterial microcompartment' *Structure*. (Accepted for publication)
2. Trettel, D.S. and Winkler, W.C. (2022) 'Bacterial-microcompartment-related structures exhibit dynamic character' *Small*. (Submitted)
3. Goodson J.R., Zhang, C., Trettel, D., Ailinger, H.E., Lee, P.E., Spirito, C.M., Winkler, W.C. (2020) 'An autoinhibitory mechanism controls RNA-binding activity of the nitrate-sensing protein NasR. *Molecular Microbiology* 114:348-360.

Chapter 1: Introduction

Section 1.1: The BMC is an Archetype for Prokaryotic Organization

Subsection 1.1.1: Prokaryotic Compartmentalization

All life straddles a thin line. On one side, organisms need to embrace macromolecular and metabolic complexity to respond to their environment. On the other side is chaos; the potential for runaway metabolism as the inherent complexity swallows its own tail with competing, and sometimes damaging, reactions. Life, however, has balanced these two competing themes by organizing and segregating its internal parts. Intracellular compartmentalization, therefore, is key to life as we know it.

Eukaryotic cells have long been understood to house numerous membrane-bound organelles. The presence of these organelles, typified by plastids and mitochondria (Ris and Singh, 1961; Keeling and Archibald, 2008), are thought to have emerged through the symbiogenesis (Zimorski *et al.*, 2014) of formerly prokaryotic brethren resulting in the eukaryote-prokaryote split over 1500 million years ago (Martin *et al.*, 2015; Gruber, 2019; Parfrey *et al.*, 2011). In contrast, prokaryotes have been regarded as being simplistic, lacking organelles and other subcellular structures (Murat *et al.*, 2010).

At least, that is what I was taught growing up – that the assumption that bacteria have no organelles is a hallmark of what makes bacteria, bacteria. Bacteria, however,

are also being found to be increasingly skilled in compartmentalization. Far from being devoid of structure, bacteria use many of the same tricks as eukaryotes to make meaning from their internal chaos from extensive cytoskeletal systems (Shih and Rothfield, 2006) to organelles (Murat *et al.*, 2010). While the term 'organelle' typically refers to a membrane-bound compartment, I will use a broader definition wherein an organelle is any subcellular compartment with a defined lumen and macromolecular boundary (Greening and Lithgow, 2020). Casting this wider net means that bacteria can be considered as having four classes of organelles; those that are (1) lipid-monolayer bound, (2) membrane-bound, (3) protein-bound, and (4) those that are phase separated (Saier Jr. and Bogdanov, 2013; Greening and Lithgow, 2020).

Whatever the method of compartmentalization, bacterial organelles act as either a storage hub or as catalytic compartments. Storage compartments include the lipid-monolayer bound chlorosomes, which contain light-harvesting bacteriochlorophyll molecules (Bryant and Frigaard, 2006) as well as the protein-bound magnetosomes, which store iron intracellularly to aid in magnetotaxis (Shively, 1974). In contrast to storage organelles are the catalytic, membrane-bound anammoxosomes, which oxidize ammonium to produce nitrogen gas (van Niftrik *et al.,* 2004) and the protein-bound encapsulins (Nichols *et al.,* 2021; Gabashvili *et al.,* 2020). This dissertation will principally focus on yet another class of protein-bound, catalytic partitions called bacterial microcompartments (BMCs) (Yeates *et al.*, 2008; Yeates *et al.*, 2010; Kerfeld, 2018).

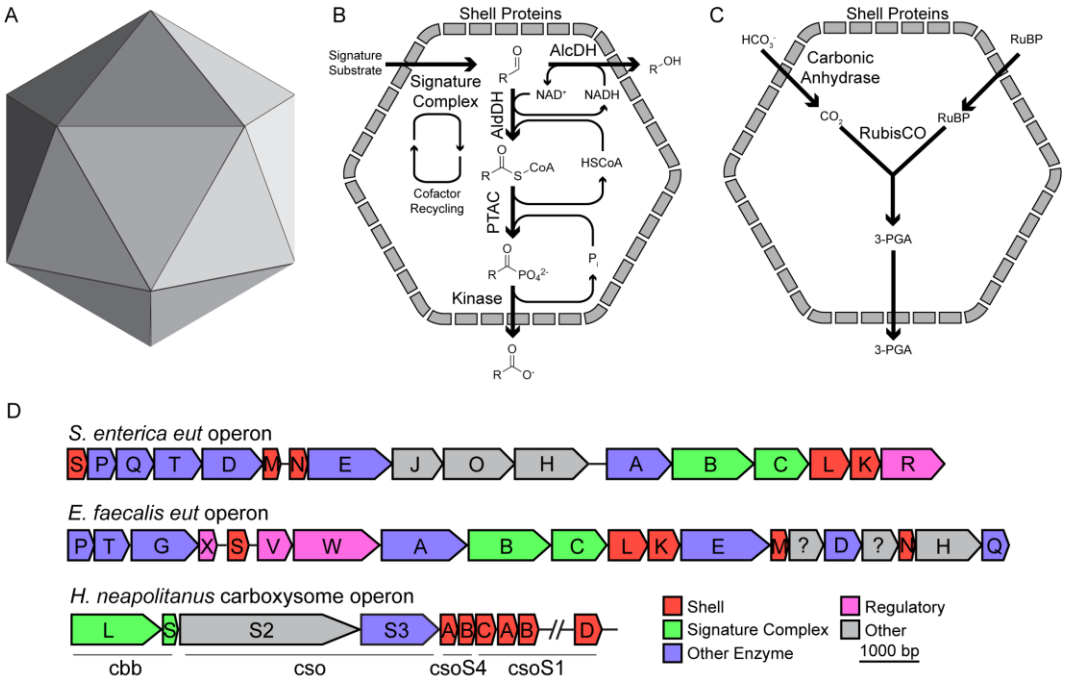


Figure 1.1: Schematics of bacterial microcompartments. (A) BMCs form icosahedral structures with 20 faces, 30 sides, and 12 vertices. (B) A general schematic of a catabolic metabolosome. Here, the signature substrate is converted into an aldehyde by the signature complex before being acted on by either an aldehyde (AldDH) or alcohol dehydrogenase (AlcDH). (C) A schematic of the anabolic carboxysome. Bicarbonate is converted to carbon dioxide and combined with ribulose 1,5-biphosphate (RuBP) to form 3-phosphoglycerate (3-PGA). (D) The layout of several BMC operons. Some operons for the catabolism of similar substrates can differ greatly in their organization. Other operons, such as the one encoding for a carboxysome in *H. neapolitanus*, can encode shell proteins (*csoS1D*) in distant satellite loci.

Subsection 1.1.2: Bacterial Microcompartments

Catalytic compartments in prokaryotes are exemplified by BMCs (Yeates *et al.*, 2008; Yeates *et al.*, 2010; Kerfeld *et al.*, 2018). BMCs have an entirely protein composition - there are no known lipid or nucleic acid components (Kerfeld *et al.*, 2018). While these supramolecular entities range in size from 40-500 nm in diameter, they all consist of an icosahedral outer protein shell (Yeates *et al.*, 2011) and the inner enzymatic cargo encased within (Figure 1.1a). Overall, BMCs are composed of thousands of copies of numerous different proteins. BMCs can have two modes of catalytic activity, being either anabolic or catabolic in nature. Anabolic BMCs refer

strictly to the carbon-fixing carboxysomes while catabolic BMCs are comprised of a more diverse group generally called metabolosomes (Figure 1.1b, 1.1c (Chowdhury *et al.*, 2014; Stewart *et al.*, 2021)). I will use the term BMC to refer generally to all BMCs but will specify between carboxysomes and metabolosomes when important distinctions arise.

Section 1.2: BMCs Are an Omnipresent, Diverse Class of Metabolic Modules

Subsection 1.2.1: The Presence of Structural Components Defines BMC Loci

BMCs have as large of a genomic footprint as they do an intracellular footprint. BMCs can be encoded in single operons typically between 10-20 kilobases in length or are scattered over several independent loci within a genome deemed satellite loci (Figure 1.1d (Axen *et al.*, 2014; Sutter *et al.*, 2021)). A case in point is the well-characterized 1,2-propanediol utilization (Pdu) BMC from *Salmonella enterica*, which is expressed from a 17 kb operon that encodes 21 genes (Bobik *et al.*, 1999; Havemann *et al.*, 2002; Havemann and Bobik, 2003). The Pdu BMC in *S. enterica* was the first described metabolosome, thanks to the identification of "polyhedral body" genes that shared homology to the shell proteins of previously established carboxysome models (Shively *et al.*, 1974). As such, it has become commonplace to identify a putative BMC locus not by their enzymatic components, which can vary greatly, but by identifying the shell proteins that coincide with them (Axen *et al.*, 2014; Ravcheev *et al.*, 2019; Asija *et al.*, 2021; Kirst and Kerfeld, 2021; Sutter *et al.*, 2021). Further, the shell proteins are highly conserved in both sequence and structure between different organisms (Yeates *et al.*, 2011) and are useful in building phylogenetic trees (Sutter *et*

al., 2021). Shell proteins, therefore, effectively act as a fingerprint for BMCs and can differentiate BMC loci from catalytically related loci that lack the ability to express shells (Lundgren *et al.*, 2016; Kaval and Garsin, 2018).

Subsection 1.2.2: Classes of BMC are Defined by Their Signature Substrate

While the enzymes encoded in BMC operons do not define it as a BMC, they do help with sub-classification. This stems from the fact that BMC shell proteins are too conserved to serve as the basis of functional identification. Therefore, the logic behind BMC classification comes from the anticipated substrate that a BMC acts on, also called the signature substrate, by classifying ancillary genes (Kirst and Kerfeld; 2021). Again, using the Pdu BMC as an example, the Pdu BMC operon encodes three genes for a B_{12}-dependent propanediol dehydratase (Abeles and Lee, 1961; Bobik *et al.*, 1997). Therefore, this particular BMC operon is responsible for propanediol utilization and is dubbed the Pdu BMC.

The widespread availability of genomic sequencing data and development of bioinformatic tools have enabled researchers to identify new classes of BMC beyond the carboxysome and Pdu BMC. In these workflows, BMC loci are identified by the presence of shell proteins, classified by their enzymatic components, then further sub-classified based on their overall loci organization and the ancillary genes. These analyses have found that BMC loci exist in 45 bacterial phyla and approximately 20% of sequenced bacterial genomes, with many genomes encoding multiple different BMCs (Sutter *et al.*, 2021). BMCs have likewise been linked to pathogenicity (Jakobson and Tullman-Ercek, 2016; Prentice, 2021) in the human gut microbiota (Ravecheev *et al.*, 2019). Further, BMCs that share the same substrate but are from

different organisms can organize very differently (Figure 1.1d). For example, *S. enterica* encodes for both a Pdu and ethanolamine utilization (Eut) BMC but the *eut* operon from *S. enterica* is structured differently than the same operon from *Enterococcus faecalis* (Figure 1.1d).

Thanks to bioinformatic inferences, the known range of metabolosome functionality has expanded beyond just propanediol utilization to include ethanolamine, aromatic amino acids, ethanol, sugar phosphate, choline (Herring *et al.*, 2018; Craciun and Balskus, 2012) and aminoacetone substrates (Sutter *et al.*, 2021; Mallette and Kimber, 2018; Petit *et al.*, 2013) with both B_{12}-dependent and independent modes (Ferlez *et al.*, 2019; Lundin *et al.*, 2020; Zarzycki *et al.*, 2017) of catalysis. Most of these, however, remain to be biochemically and structurally confirmed. Other BMC classes need to be further annotated or are difficult to assess due to shell components being decentralized among several satellite loci (Figure 1.1d, bottom (Sutter *et al.*, 2021)). The range of internal chemistries that BMCs, specifically metabolosomes, can adopt upon a structurally conserved chassis is being increasingly appreciated for synthetic goals (expanded on more in Section 1.4) (Chessher *et al.*, 2015; Gonzalez-Esquer *et al.*, 2016; Planamente and Frank, 2019; Kirst and Kerfeld, 2019; Stewart *et al.*, 2021).

Subsection 1.2.3: BMC Shell Proteins Have Bacterial Origins

The outer protein shells of BMCs are generally regarded as having an icosahedral geometry (Figure 1.1a). This fact, coupled with their proteinaceous composition and size, can tempt one to wonder if these structural components can trace their origins back to viral capsid proteins. Many apply the same rules of assembly

between the two, such as triangulation numbers (Capsar and Klug; 1962). Indeed, the encapsulin class of protein compartments in bacteria share structural homology to that of the bacteriophage HK97 major capsid protein (Krupovic and Koonin, 2017; Nichols *et al.*, 2017). BMC shell proteins, in contrast, share no such structural homology (Krupovic and Koonin, 2017). Recent structural and sequence comparisons have found that BMC shell proteins were exapted from the PII signaling protein (Ninfa and Jiang, 2005) and the OB-fold domain-containing protein (Krupovic and Koonin, 2017), leading to two different classes of shell proteins critical for shell assembly (for more detail, see Chapter 1.3.1). These two ancestral shell proteins commonly co-occur in bacterial genomes and may have allowed their parallel co-evolution (Krupovic and Koonin, 2017).

Section 1.3: The Organization of Bacterial Microcompartments

Subsection 1.3.1: All BMCs Rely on a Structurally Conserved Shell

It was previously established in Chapter 1.2.1 that BMC shell proteins are highly conserved and that conservation allows for defining and identifying BMCs. This is true regardless of the signature substrate; all BMCs are composed of three types of shell proteins: hexamers (BMC-H), trimers (BMC-T), and pentamers (BMC-P) (Figure 2a (Yeates *et al.*, 2011; Kerfeld *et al.*, 2018; Ochoa and Yeates, 2021)). All three structurally present as flat tiles that are cyclically arranged about a central pore (Figure 1.2a). These pores accommodate metabolite influx/efflux through the protein shell (Chowdhury *et al.*, 2015; Mahinthichaichan *et al.*, 2018; Faulkner *et al.*, 2020) and may even "open" in response to metabolites (Heldt *et al.*, 2009; Pang *et al.*, 2012). Some

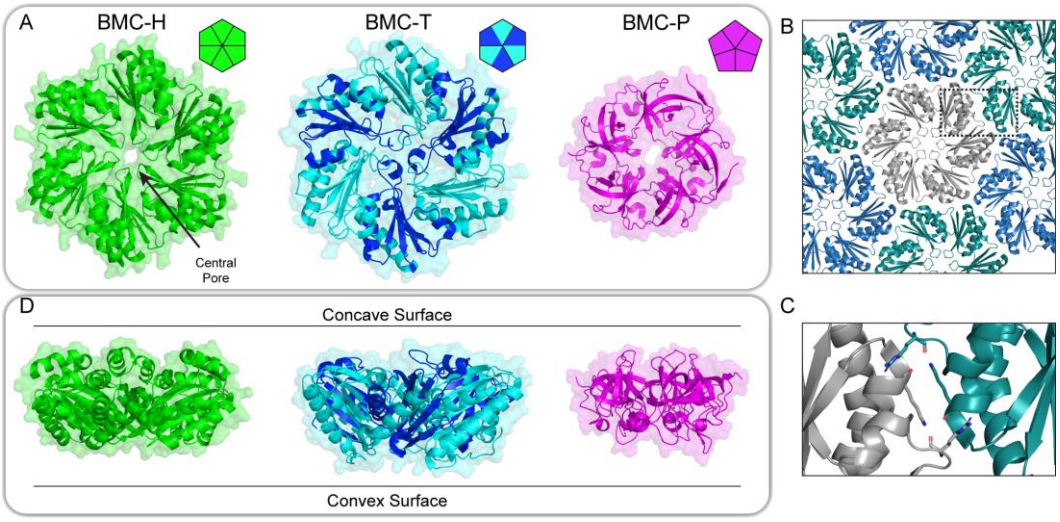


Figure 1.2: The shells of BMCs are composed of three distinct shell proteins. (A) Crystal structures of shell protein hexamers (BMC-H), trimers (BMC-T), and pentamers (BMC-P). Trimers contain a fused tandem repeat of the BMC domain which are represented by alternating colors. BMC-H is CcmK4 (2A18, Kerfeld *et al.*, 2005), BMC-T is PduB (4FAY, Pang *et al.*, 2014), and BMC-P is CsoS4B (6J5Y. Zhao *et al.*, 2019). (B) BMC-H and BMC-T proteins tessellate together to form the facets of the shell while the pentamers cap the vertices. Shown is the BMC-H protein PduA (Crowley *et al.*, 2010). The selected region from panel (B) shows the basis of facet formation (C) being an antiparallel pairing of lysine residues and shape complementarity from arginines docking into adjacent hexamers. (D) Side view of shell proteins. All shell proteins have a characteristic concave and convex surface.

pores even accommodate electron transport as opposed to metabolite trafficking. An example of this comes from the BMC-T PduT which hosts an iron-sulfur cluster at its pore (Pang *et al.*, 2011). The shell is thought to be otherwise impregnable (Dou *et al.*, 2008). The BMC-H and BMC-T classes tessellate in a honeycomb-like fashion to build the single-layer facets of the polyhedral structure with the pentamers capping the 12 vertices (Figure 1.2b (Tanaka *et al.*, 2008)). This tessellation is dependent upon shape complementarity and the pairing of lysine residues along the edge of each tile to specifically accommodate antiparallel hydrogen bonding between backbone carbonyl oxygens and side chain amines (Figure 1.2c (Sutter *et al.*, 2017)). Further, each shell protein has a characteristic concave and convex sidedness to them (Figure 1.2d). The convex face is known to present luminally in empty, synthetic BMC derivatives (Sutter *et al.*, 2017; Jorda *et al.*, 2016; Tan *et al.*, 2021; Sutter *et al.*, 2019a; Kalnins *et al.*,

2020) but the orientation in a native context remains to be identified (Tanaka *et al.*, 2008).

BMC-H are made from six copies of monomers, each containing a single copy of the BMC domain (pfam00936), oligomerizing around a central pore (Kerfeld *et al.*, 2005). A rarer sub-class of circularly permuted hexamers (BMC-H^P) also naturally exist (Sutter *et al.*, 2021; Crowley *et al.*, 2008). Some synthetic permutations have been noted to change the oligomer state to a pentamer (Jorda *et al.*, 2016). Regardless, all known BMC contain at least one copy of a BMC-H protein (Sutter *et al.*, 2021). Accordingly, these proteins are the most abundant by mass and copy number in BMC systems (Yang *et al.*, 2020; Mayer *et al.*, 2016; Havemann and Bobik, 2003). This observation has earned the BMC-H class of shell proteins the reputation of being the fundamental building block of the BMC shell. This fact may also be reflected by the ability of BMC-H proteins to form a variety of higher-order structures when overexpressed *in vivo* and overproduced *in vitro,* such an nanotubes and flat sheets (Uddin *et al.*, 2018; Noël *et al.*, 2015; Hagen *et al.*, 2018b; Crowley *et al.*, 2010; Pang *et al.*, 2014).

BMC-T proteins consist of a tandem repeat of the BMC domain, effectively making them a tandem BMC-H protein (Pang *et al.*, 2011). Correspondingly, it is thought that BMC-T proteins arose from a gene duplication of a BMC-H gene. This aspect has allowed researchers to create synthetic BMC-T proteins by direct duplication and/or fusion of BMC-H proteins (Sutter *et al.*, 2019b). Far from being just a genetic quirk, the fusion of BMC-H genes into a sole BMC-T protein entails functional ramifications. For instance, it has been proposed that the duplication can allow each

BMC domain to evolve independently of one another to accommodate different structural roles along each edge of the pseudo-hexamer. This suspicion is supported by a high-resolution structure of a small, proteobacterial BMC wherein the BMC-T was found to accommodate different curvature angles on different edges, reinforcing the notion that BMC-T proteins may help control overall shell curvature (Sutter *et al.*, 2017). There are three sub-classes of BMC-T proteins: the standard trimer (BMC-T^{S}), the permuted trimer (BMC-T^{SP}), and the permuted, dimerized trimer (BMC-T^{DP}) (Sutter *et al.*, 2021). BMC-T^{DP} class can dimerize along their concave faces to form an inner chamber (Klein *et al.*, 2009; Cai *et al.*, 2013). The exact roles and differences in the context of shell assembly among these sub-classes is not yet fully understood although they appear to each be equivalently abundant across BMC loci (Sutter *et al.*, 2021). Some suggest that BMC-T make up the edges of the BMC 10olyhedral (Liu *et al.*, 2021; Yang *et al.*, 2020, Mallette and Kimber 2017), but other structures place them in the facets (Sutter *et al.*, 2017).

Lastly, BMC-P proteins are unrelated to the hexamer and trimer varieties, instead consisting of a single copy of a pfam00319 domain (Tanaka *et al.*, 2008; Wheatley *et al.*, 2013). This is likely a direct result of separate evolutionary histories, where BMC-P came from OB-fold domain-containing proteins and BMC-H/T came from PII signal transduction proteins (Krupovic and Koonin, 2017). Despite this, BMC-P still topologically arranges like BMC domain-containing proteins with a cyclical arrangement and convex/concave sides (Figures 1.2a, d). The pentameric structure serves a critical purpose in capping the vertices of the polyhedron to provide stability (Tanaka *et al.*, 2008). Some synthetic systems have been reported to exist without them

(Hagen *et al.*, 2018a). Regardless, all naturally occuring BMC loci to date are thought to encode at least one BMC-P gene (Sutter *et al.*, 2021).

Subsection 1.3.2: The Luminal Composition of BMCs

The luminal space of BMCs remains to be defined; it has not been as clearly studied as the outer protein shell. This is particularly true for metabolosomes where the inner cargo composition can change depending on the class of BMC and, consequently, also the way it arranges and organizes itself. For example, most cargo components in the *S. enterica* Pdu BMC are presumed to be encased within the protein shell (Jakobson *et al.*, 2015; Fan *et al.*, 2010), but a model of the aminoacetone BMC from *M. smegmatis* proposed that some components might be adhered to the outside of the protein shell (Mallette and Kimber, 2018).

What is more certain are inferences drawn from cyanobacterial α- and β-carboxysomes which have been intensely scrutinized since they were first purified in 1973 (Shivley *et al.*, 1973). While both types of carboxysomes package and use D-ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO) to fix CO_2, they both use different forms of RubisCO and are thought to package it differently within the protein shell, among other caveats (Rae *et al.*, 2013). In the context of RubisCO packaging, electron micrographs of α-carboxysomes have supported the notion that the luminal space is largely disordered (Rae *et al.*, 2013), while, interestingly, β-carboxysomes are described as having a paracrystalline arrangement that extends throughout the lumen (Cameron *et al.*, 2013; Faulkner *et al.*, 2017; Kerfeld and Melnicki, 2016). Phylogenetic analysis of BMC-P proteins has shown that β-carboxysomes are more closely related to metabolosomes than α-carboxysomes (Kerfeld and Melnicki, 2016),

suggesting metabolosomes may have an organized lumen. However, similar paracrystalline observations have yet to be found to apply to metabolosomes. This may be a result of the luminal diversity inherent in metabolosomes since they appear to package a greater diversity of cargo than carboxysomes.

Subsection 1.3.3: The Role of Encapsulation Peptides in Shell-Cargo Interactions

β-carboxysomes share another critical feature with metabolosomes, they make use of small peptides fused to cargo components to guide their assembly inside the protein shell (Kerfeld and Melnicki, 2016). Since these peptide sequences guide cargo encapsulation, they are fittingly known as encapsulation peptides (EPs). EPs are 15-17 amino acid α-helical extensions on either the N- or C-terminus of some cargo proteins (Lawrence *et al.*, 2014; Aussignargues *et al.*, 2015). They are fused to cargo proteins by an intrinsically disordered linker that can vary greatly in length from 4-91 residues (Aussignargues *et al.*, 2015). However, not all cargo proteins in any given operon contain an EP. These proteins are presumably packaged through 'piggybacking' interactions with cargo that do have EPs (Kerfeld *et al.*, 2018).

EPs were first discovered in 2010 when Fan *et al.* took note that the cargo in operons that also encoded shell proteins contained N-terminal extensions that are not present in homologous enzymes that exist outside of a BMC context (Fan *et al.*, 2010). This initial investigation found that deleting this extension (the EP) of the AldDH PduP led to lessened cargo packaging and that adding this EP to heterologous cargo, like GFP, led to internalization of the heterologous cargo (Fan *et al.*, 2010). A later study revealed that EPs allow cargo proteins to interact directly with shell proteins, specifically the EP of PduP was proposed to interact with the C-terminus of the shell

protein PduA, which lies on the concave face (Fan *et al.*, 2012). Molecular docking analyses also supported this conclusion (Jorda *et al.*, 2015). It is worth nothing that this would entail the concave side of the shell facing luminally which is in stark contrast with later structural data of synthetic BMC shell models (Sutter *et al.*, 2017). Some work has also found that the EP of PduP also interacts with the shell protein PduK with a k_d of 331 ± 64 nM, although the binding site was not established (Lawrence *et al.*, 2014). Regardless, researchers have established that EPs utilize a conserved hydrophobic motif where hydrophobic residues lay predominantly on one face of the helix and are critical for their function (Aussignargues *et al.*, 2015; Jakobson *et al.*, 2015). This conservation is functionally reliable enough that EPs between different classes of BMC can be swapped and result in cargo packaging (Jakobson *et al.*, 2015). These findings have enabled others to design flow cytometry-based assays to design new EPs (Kim and Tullman-Ercek, 2014).

Despite only being appreciated for their biotechnological applications in the past decade, EPs initially established themselves as a nuisance in early purification efforts. For instance, early high-resolution X-ray structures of the signature complex PduCDE from *Klebsiella oxytoca* do not include the N-termini of PduD and PduE as they hampered purification and crystallization (Masuda *et al.*, 2000). Today, it is understood that these deletions corresponded to removing the EPs of these two proteins, although the EP of PduE is not required for PduCDE packaging (Fan and Bobik, 2011). As such, EPs have also been appreciated to generally lead to aggregation (Juodeikis *et al.*, 2020) and have been used to create active shell-free enzyme aggregates (Lee *et al.*, 2016).

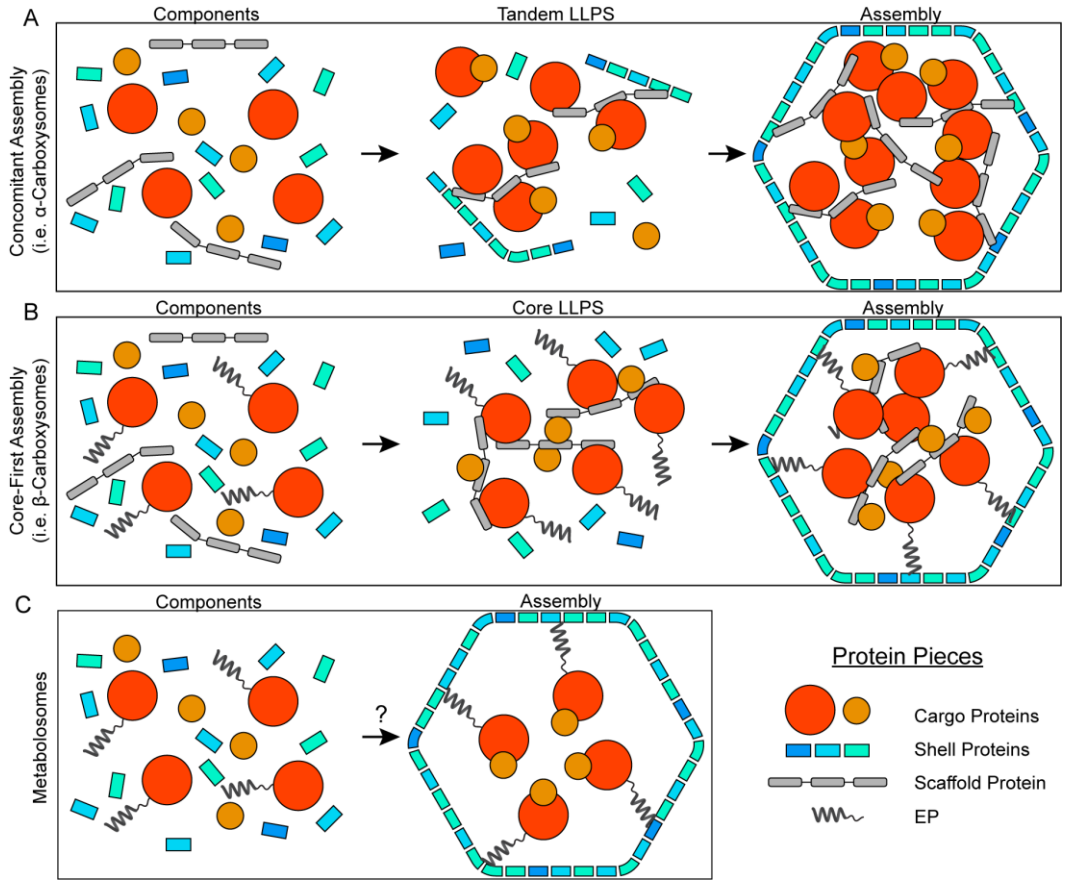


Figure 1.3: Current biogenesis models for BMCs. (A) Concomitant assembly requires the tandem phase separation of cargo components with shell proteins. This process is facilitated by multivalent scaffold proteins that bind both cargo and shell proteins. The α-carboxysome is thought to be assembled in this way. (B) Core-first assembly instead entails the phase separation of cargo components in the presence of a protein scaffold prior to shell envelopment. EPs help with shell recruitment. The β-carboxysome is thought to assemble in this manner. (C) The biogenesis pathway for metabolosomes has not been defined. Metabolosomes are more closely related to β-carboxysomes, having EPs, but lack dedicated scaffold proteins known to initiate cargo aggregation.

Subsection 1.3.4: Biogenesis Models

The important discovery of EPs allowed researchers to propose two biogenesis models for BMCs. The first, called core-first assembly, stipulates that core-assembly proteins (*i.e.*, intrinsically disordered scaffolds) aggregate cargo proteins prior to shell envelopment governed by shell-EP interactions. The second, called concomitant assembly, is where the inner core and outer shell assemble together simultaneously (Kerfeld *et al.*, 2018). Core-first assembly is the favored biogenesis model for β-

carboxysomes (Kerfeld and Melnicki, 2016; Kerfeld *et al.*, 2018). In this instance, a pro-carboxysome aggregation of RubisCO is triggered by complexation with the intrinsically disordered, multivalent scaffold protein CcmM followed by shell protein envelopment and subsequent budding of mature carboxysomes from the pro-carboxysome (Figure 1.3b (Cameron *et al.*, 2013; Chen *et al.*, 2013)).

The process of the pro-carboxysome formation has more recently been tied to liquid-liquid phase separation (LLPS) governed by the multivalent, disordered scaffold protein CcmM (Zang *et al.*, 2021). This finding is, perhaps, unsurprising given that multivalency and disorder tend to be inherent to phase separated protein systems (Banani *et al.*, 2017). α-carboxysomes have likewise been recently tied to phase separation in the initial stages of biogenesis (Oltrogge *et al.*, 2020) thanks to multivalent interactions between RubisCO and the IDP scaffold CsoS2, although electron micrographs (Iancu *et al.*, 2010) and computational modeling (Mohajerani et al, 2021) suggest that the shell is involved in this process as well (Figure 1.3a). Why shell recruitment is tandem to core LLPS in α- and not β-carboxysomes is not yet understood. Whether metabolosomes follow core-first or concomitant assembly has yet to be established (Figure 1.3c). The parallels between metabolosomes and β-carboxysomes might suggest the latter; however, metabolosomes do not typically encode for a specific, intrinsically disordered scaffold protein, suggesting their assembly pathway may differ. Given their functional diversity, it is also unclear if all metabolosomes would fall neatly into one biogenesis model or another.

Section 1.4: BMCs in the World of Synthetic Biology

Subsection 1.4.1: BMCs Have a Range of Desirable Properties for Bioengineering

BMCs are of considerable biotechnological interest due to their intrinsic ability to self-assemble and encapsulate a wide range of functionally diverse cargo (Frank *et al.*, 2013; Panamente and Frank; 2019; Chessher *et al.*, 2015; Kerfeld, 2018; Polka *et al.*, 2016; Papapostolou and Howorka, 2009; Pokhrel *et al.*, 2021; Lee *et al.*, 2019). Their benefits for synthetic applications can be understood as those being intrinsic to compartmentalization as a whole or as the modular aspects specific to BMCs (Figure 1.4).

The intrinsic benefits to BMC-based partitions stem from concentrating and recycling enzymes/cofactors to increase reaction rates and limit metabolic crosstalk between the BMC lumen and cytosol. For example, BMCs have been known to retain and recycle a private pool of cofactors within their inner lumen (Huseby and Roth, 2013). This recycling can be understood as a necessity within a closed system and to aid in increasing reaction rates in conjunction with the presumably high concentration of enzymatic components compared to shell-free systems. Further, BMCs, like other compartments, limit the influx and efflux of metabolites. This has become understood as a necessity given that the central metabolite in most metabolosomes is a reactive aldehyde intermediate (Figure 1.1b), which can damage cytosol macromolecules if it leaks out (Sampson and Bobik, 2008). These aldehyde intermediates are also volatile, thus retaining them not only prevents cellular damage but also increases metabolic flux (Jakobson *et al.*, 2017). In carboxysomes, shell disruption has been tied to interruption of RubisCO due to leakage of molecular oxygen into the lumen (Berry *et al.*, 2005). As

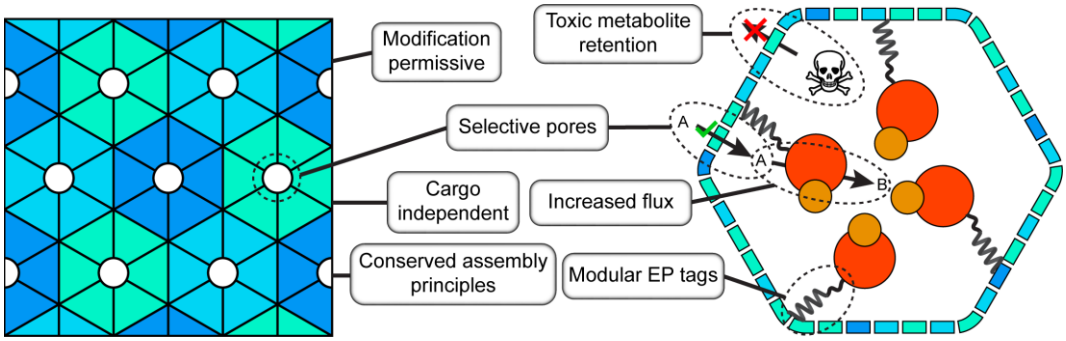


Figure 1.4: The benefits of BMC-based synthetic biology platforms. The protein shell of all BMCs use the same, predictable principles. The shell assembly is also largely agnostic to the inner cargo composition, allows for modifications, and gates the influx/efflux of metabolites, including toxic pathway intermediates. Novel cargo can be encapsulated using EP domains and benefit from being concentrated within the lumen, increasing metabolic flux through designer pathways.

such, BMC-based metabolite platforms would be useful if one wished to encapsulate pathways that are prone to inactivation or that produced toxic intermediates.

The above benefits are not intrinsic to just BMCs. BMCs do, however, contain several unique modular, designable properties that have caught the attention of researchers. Chief among these are characteristics of the protein shell where the selectively of the pores can be altered by simple mutagenesis (Chowdhury *et al.*, 2015; Aussignargues *et al.*, 2016). The protein shell and its assembly is also extremely modular, fitting together like Lego pieces, and can accommodate protein fusions for added functionality (Lee *et al.*, 2018; Hagen *et al.*, 2018a; Parsons *et al.*, 2010). Some fusions, like the SUMO tag, have been used to prevent *in vitro* synthetic shell formation until activated by cleavage (Hagen *et al.*, 2018b). EPs themselves are also modular and can be fused to novel cargo to direct their encapsulation (Fan *et al.*, 2010; Jakobson *et al.*, 2015) but newer, non-EP methods of encapsulation have also been developed, such as the SpyTag/SpyCatcher system (Zakeri *et al.*, 2012; Hagen *et al.*, 2018a) and the use of coiled-coil domains fused to both cargo and shell proteins alike (Lee *et al.*, 2018).

Even the genetic structure of BMC operons is modular. Several BMC operons have been cloned and successfully expressed in a range of heterologous hosts (Bonacci *et al.*, 2012; Graf *et al.*, 2018; Wilson, 2021). For example, the *pdu* operon from the thermophile *Parageobacillus thermoglucosidasius* was transferred to *Bacillus subtilis* for the IPTG-inducible expression of Pdu BMCs (Wade *et al.*, 2019). In another instance, researchers were able to successfully express *Synechococcus* β-carboxysomes in *E. coli* by designing a synthetic operon that combined the carboxysome genes scattered across several different loci (Fang *et al.*, 2018). This ability highlights the ubiquity and utility of BMCs to be used as a broadly applicable *in vivo* modular platform.

Subsection 1.4.2: Current BMC-Based Synthetic Applications

Numerous published studies have speculated on the benefits of BMC-based synthetic biology. These studies can be classified as either focusing entirely on synthetic shell biotechnology or those attempting to redesign BMCs for novel chemistries (Panamente and Frank, 2019). With regards to the latter class, the Kerfeld group is pioneering work on engineered synthetic BMC shells. This was first initiated by their discovery of a BMC operon from *Haliangium. ochraceum* which, when heterologously expressed from a synthetic operon encoding just shell proteins, led to the expression of purifiable, uniform BMC shells with a diameter of 40 nm (Lassila *et al.*, 2014). This finding later led to a high resolution cryo-EM structure of the shell (Sutter *et al.*, 2017), now dubbed the Hoch BMC - a first of its kind that led to unique insights on shell construction. Since then, the Hoch BMC has served as the structural model for the field and has been covalently modified with heme (Huang *et al.*, 2019),

enabled site-specific cargo loading (Ferlez *et al.*, 2019; Hagen *et al.*, 2018a), accommodated shell fusions to hexahistidine and Strep tags to facilitate purification (Hagen *et al.*, 2018a), and enabled the design of simplified shells requiring only two proteins (Hagen *et al.*, 2018b; Sutter *et al.*, 2019b). Other small BMC shells have been designed and their structures elucidated, ranging from 13-40 nm in diameter (Jorda *et al.*, 2016; Sutter *et al.*, 2019a; Kalnins *et al.*, 2020; Tan *et al.*, 2021). There are instances of larger BMC shells being purified, such as the Pdu BMC shell from *Citrobacter. freudnii* (Mayer *et al.*, 2016), but these are more irregularly shaped and require more components than their smaller cousins.

Groups have also redesigned BMCs for novel chemistries. In a recent instance, Li *et al.* redesigned a β-carboxysome shell to be expressed in *E. coli* and encapsulate hydrogenases to produce molecular hydrogen (Li *et al.*, 2020). *In vivo* expression of these synthetic BMCs produced 4-fold more H_2 than the hydrogenase expressed without a shell under anaerobic conditions and were far more resistant to oxygen exposure *in vitro*. Similar improvements were found concerning ethanol production using an AlcDH and pyruvate decarboxylase fused to EPs and packed into an otherwise empty Pdu shell (Lawrence *et al.*, 2014). Another study again used EPs to package β-galactosidase into Pdu shells and found that the packaging protected the luminal enzymes from pH, but not temperature or related stress (Wagner *et al.*, 2017). In a wholly different direction, synthetic BMC shells have also helped produce, package, and purify bacteriophage lysis proteins (Yung *et al.*, 2017).

The modularity of BMC systems has also inspired bioengineering non-native BMC-based platforms. In one instance, researchers used BMC-H derived nanotubes *in*

vivo to enhance ethanol production. Here, an AlcDH and pyruvate dehydrogenase were fused to synthetic coiled-coil targeting domains and colocalized to the cytoskeletal-like BMC-H filaments (Lee *et al.*, 2018). The aggregative tendencies of EPs have also been used to generate shell-free enzyme inclusion bodies to produce 1,2-propanediol (Lee *et al.*, 2016). Interestingly, this study found that the shell-free system outcompeted a similar system co-expressed with shell proteins *in vivo*.

Subsection 1.4.3: Issues with Current Design Strategies

The initial bioengineering work described above are proof-in-principle of the potential held by BMCs. BMCs biotechnology is coming of age; change happens quickly in this field. Despite 'polyhedral bodies' being described over 60 years ago (Drews and Niklowitz; 1956), the past decade alone has seen the field go from not understanding the mechanisms of encapsulation to having atomic-level resolution of empty BMC shells. According to PubMed (January, 2022), there have been 305 articles published containing the term "bacterial microcompartment" since 1995 with over half of those published since 2017. Clearly, there is interest in these enigmatic entities. However, current bioengineering efforts have a few hurdles to overcome before the full potential of BMC biotechnology can be realized.

First, most synthetic empty shells that have been structurally characterized (Sutter *et al.*, 2019; Tan *et al.*, 2021, Kalnins *et al.*, 2020) are small - 40 nm in diameter at the largest (Sutter *et al.*, 2017). Native BMCs, in contrast, are well over 100 nm in diameter. Knowledge of the structure of these large, natural BMCs would be an immense help to the field. It would help researchers understand why different BMCs use different combinations of shell components and how to leverage that for their own

goals. Such knowledge could help construct designer BMC shells of various sizes and properties. Work that attempts to replicate larger BMC shells are often dismayed to realize how relatively heterogeneous their structures are (Kennedy *et al.*, 2020) and, in some cases, that they localize to the poles of cells *in vivo* (Lee *et al.*, 2018) and may be conflated with being inclusion bodies instead of the desired polyhedra. Discrepancies also exist between the reported "sidedness" of the protein shell in synthetic versus natural systems (Tanaka *et al.*, 2008). Therefore, there is much to learn when it comes to understanding the native structure of BMCs.

Synthetic BMCs, however, that replicate the morphology of naturally occurring BMCs do exist. Examples of this are the hydrogen producing BMC from Li *et al.* (Li *et al.*, 2020) and the ethanol bioreactor from Lawrence *et al.* (Lawrence *et al.*, 2014). These are largely proof-in-principle applications that resulted in modest production gains over shell-free systems. While a promising start, they all suffer from only modest packaging of their desired enzymatic cargo despite using natural EP localization signals. It may be that there is more to EP-based encapsulation. This may be unsurprising given the scant data concerning the mechanics of how EPs operate. Knowledge on how EPs act in natural systems is therefore crucial to leveraging them in synthetic systems. Perhaps this explains why the 1,2-propanediol producing shell-free enzyme aggregate of Lee *et al.* (Lee *et al.*, 2018) was able to outcompete a similar shell-based system.

Lastly, the use of the higher-order structures presented by BMC-H nanotubes can be greatly expanded on. As of this writing, the literature on these structures is limited to only a handful of publications. What we do know is that they can act as a

proxy cytoskeleton *in vivo* to scaffold enzymes but they also may cause damage to the host cell as they inhibit septation (Lee *et al.*, 2018). Delineating the higher-order structures of a wider range of BMC shell proteins could greatly expand the non-native BMC toolkit presented by shell proteins. Knowing the structures created by a diverse range of shell proteins can also help establish the individual contributions of shell proteins to native BMC structure, furthering design efforts beyond just hoping one's BMC system of interest behaves well.

Section 1.5: Uncovering the Structural Features of BMC-Related Entities

BMCs hold considerable promise as platforms for synthetic biology efforts. Their application, however, would greatly benefit from a proper understanding of how native BMCs are assembled. Uncovering the blueprint of a complete BMC would allow researchers to better mimic native methods of assembly. Structural assessment of native BMCs, specifically metabolosomes, has stagnated. Many techniques commonly used to assess structure, like X-ray crystallography and cryo-EM, would fail from their heterogeneity and complexity. Using NMR to study these objects would likewise fail from their sheer size alone. TEM and light microscopy techniques may have carried the field for decades, but new approaches must be developed to better understand these gargantuan protein structures.

This dissertation project focused on uncovering new structural insights into these complex objects, specifically using the established *S. enterica* Pdu BMC as a model system. First, I present evidence that fluorescent chemical probes can be used to demarcate surface features of these BMCs and contrast them with the interior portions. Pdu BMCs are permissive of these chemical modifications, which allows for native

high-resolution confocal imaging of Pdu BMC foci, as well as the development of structural proteomics workflows based on crosslinking. Crosslinking mass spectrometry, meanwhile, gives incredible insight into not just the shell structure of a native Pdu BMC, but also the oft neglected lumen. Confocal imaging of fluorescent Pdu BMCs enabled us to better understand BMC structures *in vitro* as potentially more dynamic than previously appreciated. Lastly, I use several fluorescently labeled shell proteins in conjunction with TEM and confocal approaches to understand their individual structures, their combinatorial effects on one another, and their influence with perturbing native Pdu BMCs. These results give a blueprint and allow us to speculate on the native assembly state and biogenesis pathway of the Pdu BMC that may apply broadly towards other metabolosomes. These results also prescribe remedies that can be easily translated into the next generation of synthetic BMCs.

Chapter 2: Chemical Modifications of Pdu BMCs

Section 2.1: Introduction

Subsection 2.1.1: BMCs Act as a Selective, Stable Barrier

BMC shells are composed of hexagonal, flat proteins that contain a central pore (Figure 1.2a). The combination of shell proteins altogether acts as a barrier while the pores in the individual shell units allow for selective trafficking of central metabolites (Kerfeld *et al.*, 2005; Yeates *et al.*, 2008; Crowley et al, 2010; Cai *et al.*, 2009; Mahinthichaichan *et al.*, 2018; Chowdhury *et al.*, 2015). The pores of shell proteins are small, at under 10 Å in diameter (Kerfeld *et al.*, 2005; Tsai *et al.*, 2007; Dryden *et al.*, 2009). For the major Pdu BMC shell protein PduA, the pore is lined by serine residues and is just under 9 Å in diameter - not taking into account atomic radii (Crowley *et al.*, 2010) (Figure 2.1a). This makes the pore of PduA too small to allow passage of cofactors such as NADH and B_{12} used in the Pdu BMC (Figure 2.1b) but large enough and of the right chemical properties to allow the influx of the intended 1,2-propanediol substrate (Crowley *et al.*, 2010). As a result, it is thought that NADH and B_{12} become packaged within the Pdu BMCs lumen during assembly and are internally recycled (Cheng *et al.*, 2012; Liu *et al.*, 2015). Some shell proteins, like the BMC-T PduB, may also participate in actively gating metabolites since its pore has been observed in both open and closed conformations (Thompson *et al.*, 2014; Takenoya *et al.*, 2010; Pang *et al.*, 2012). The other Pdu BMC-T, PduT, is known to coordinate a 4Fe-4S cluster in its

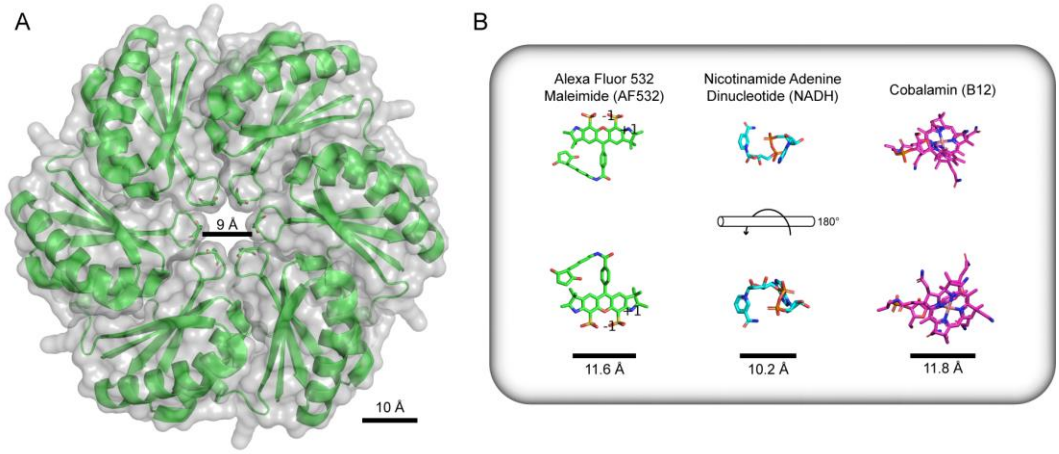


Figure 2.1: The pores of shell proteins allow passage of small metabolites. (A) The major Pdu BMC shell protein PduA (3NGK, Crowley *et al.*, 2010) has a serine-line pore with a diameter of 9 Å. This pore is thought to allow influx of 1,2-propanediol and block efflux of the more apolar propanaldehyde (Chowdhury *et al.*, 2015). (B) The shell protein pores are smaller than internal cofactors and chemical modifiers used in this research. AF532, the primary fluorophore described here, is larger than the pores and of very different properties than natural substrates. NADH and B_{12} are also larger than the pore and are thought to be recycled internally.

pore (Pang *et al.*, 2011). The Pdu BMC shell may encode numerous different shell proteins not just for structural roles, but to fulfill distinct metabolite trafficking functions.

Subsection 2.1.2 Using Tags to Demarcate Surface Features of the Pdu BMC

We initially sought to leverage the ability of the Pdu BMC shell to act as a barrier in order to demarcate the surface features. We believed that the protein shell would limit the diffusion of chemical tags to the inner cargo thereby causing preferential treatment of the shell exterior. Purified Pdu BMCs have been observed to withstand temperatures as high as 60°C, pHs between 6-8, and storage at room temperature for up to 9 days (Kim *et al.*, 2014). Due to this, we believed Pdu BMCs could also withstand chemical modifications. We began with cysteine-reactive fluorescent probes (Figure 2.1b) in hopes that these probes would lead to less severe systemic perturbations while being larger than the pores in the shell (Figure 2.1),

limiting their diffusion into the lumen. The fluorescent probes should therefore allow for qualitative assessments of surface accessibility among the Pdu BMC when separated with SDS-PAGE. We then altered this initial protocol to introduce a second probe for direct contrasting of interior versus exterior elements when resolved by 2D-PAGE. This work was also expanded to include other probes with different functionalities as well as protein crosslinkers.

Section 2.2 Results and Discussion

Subsection 2.2.1: Preparation of Fluorescent Probes for Protein Tagging

Protein conjugation most commonly relies on cysteine-reactive maleimides or amine-reactive NHS esters (Toseland, 2013; Spicer and Davis, 2014; Hoyt *et al.*, 2019). These reactive handles, particularly NHS esters, are prone to hydrolysis and inactivation (Mattson *et al.*, 1993). Vendors sell a suite of fluorophores with these functionalities, but they are expensive (>$300 per milligram) and sold in inconvenient dry aliquot sizes. Due to this, I prepared single-use aliquots of fluorescent probes from vendor stocks to make experiments more convenient and cost-friendly.

The two fluorophores that were most used in this work are Alexa Fluor 532 maleimide (AF532) and Alexa Fluor 680 maleimide (AF680). These fluorophores are both sold in one milligram dry aliquots. These aliquots were resuspended in anhydrous DMSO, as DMSO will not react with the maleimide handle, to a concentration of 8 mM and further aliquoted in 1.5 mL microcentrifuge tubes. Dye aliquots were then stored under desiccation at -20°C. Probes prepared in this manner maintain their

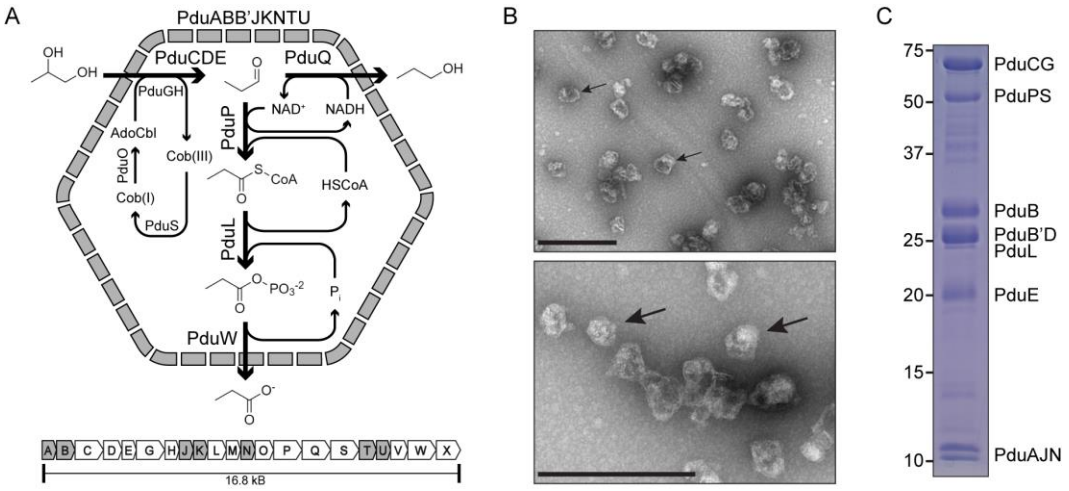


Figure 2.2: The model Pdu BMC used in this dissertation. (A) A schematic of the Pdu BMC from *S. enterica* Typhimurium LT2. Shell proteins are shaded gray in the BMC and operon models. PduB' is a truncated translation product of the *pduB* gene lacking the N-terminal 30 amino acids. (B) Negative stain transmission electron micrographs of purified Pdu BMCs. BMCs appear as protein globules and are denoted by arrows. Scale bar, 500 nm. (D) SDS-PAGE of *S. enterica* Pdu BMCs expressed and purified from a heterologous host.

reactivity for at least six months. A typical preparation from a 1 milligram aliquot of dye would yield at least a dozen smaller aliquots which drastically reduced costs.

Subsection 2.2.2: Purification of Pdu BMCs

The *S. enterica* Pdu BMCs (Figure 2.2a) used in this project were expressed and purified from a heterologous *E. coli* host. This *E. coli* strain carries a bacterial artificial chromosome containing the entire *pdu* locus as well the adjacent *cob/cbi* locus responsible for cobalamin biosynthesis (Graf *et al.*, 2018). This artificial chromosome encoded these 38 kilobases cloned from *S. enterica* through FRT-Capture (Graf *et al.*, 2018; Santiago *et al.*, 2011). The fact that a heterologous host was used allowed us to culture cells in rich media instead of the minimal media typically used in BMC preparation from *S. enterica* (Sinha *et al.*, 2012). The *pdu* operon in either case is induced by addition of 1,2-propanediol to the growth medium thanks to the transcriptional activator, PocR (Bobik, 1992; Cheng *et al.*, 2008).

Purification of BMCs followed similar protocols described elsewhere that rely on a series of centrifugation steps to selectively pellet Pdu BMCs away from other cellular matter (Graf *et al.*, 2018; Sinha *et al.*, 2012). Our preparations of purified Pdu BMCs appeared similar to the published images of purified Pdu BMCs when imaged with TEM. Briefly, purified Pdu BMCs appeared as discreet, albeit deflated, polyhedra with some clear edges (Figure 2.2b). Sample warping is not atypical for metabolosomes like the Pdu BMC and has been noted to occur during TEM sample preparation from dehydration and stain deposition (Kennedy *et al.*, 2020). Purified samples resolved with SDS-PAGE confirm the presence of many Pdu BMC components such as the major shell proteins PduABB'J and cargo proteins PduCDELP (Figure 2.2c).

Subsection 2.2.3: Fluorescent Labeling is Biased for Shell Proteins

Fluorescent labeling of Pdu BMCs first needed to be optimized. Accordingly, a series of AF532 concentrations were assayed for their ability to sufficiently label Pdu BMCs over time and then assayed by SDS-PAGE. Appropriate labeling conditions would have to yield sufficient signal over background for major BMC proteins while also allowing for differentiation of shell and cargo proteins. For instance, the major shell targets are PduBB' which encode a cysteine on each side of the trimer and are therefore expected to demonstrate intense fluorescence. Note, the major shell protein PduA does not contain a cysteine and will not label through this method. The method is also dependent on the morphology of the compartments - *i.e.*, intact compartments would be expected to label differently than those that have been denatured. AF532 concentrations of 1, 5, 31.25, and 125 μM were assayed such that [DSMO] and [BMC] concentration in each labeling pot was equal (Figure 2.3a).

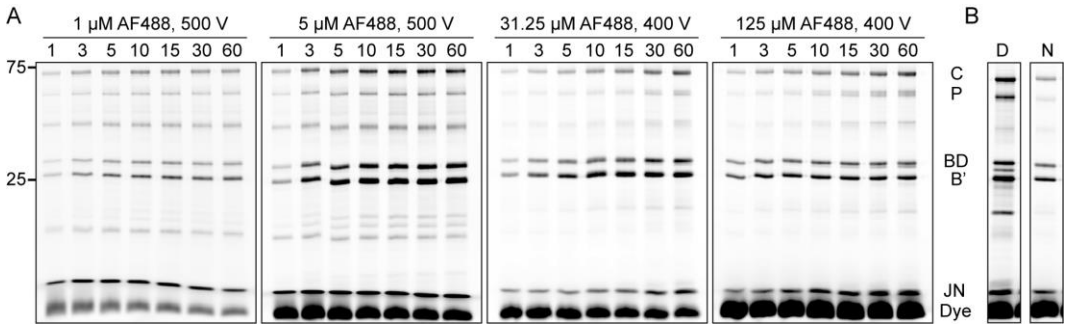


Figure 2.3: Pdu BMC labeling with AF532 were optimized for dye concentration. Conditions were screened for good signal over background, ability to label BMCs over time, and for the ability to differentiate between shell and cargo proteins. (A) A series of AF532 concentrations were assayed over an hour. The 1 and 5 µM gels were scanned with a more sensitive detector setting (500 V) compared to the 31.25 and 125 µM gels. (B) The labeling profile of Pdu BMCs is morphology dependent. Denatured BMCs (D) exhibits more bands that are brighter than native (N) BMCs when labeled under similar conditions (125 µM AF532 for 5 minutes at RT). Distinct protein bands are labeled

Optimization results indicated that lower concentrations, such as the 1 and 5 µM conditions, become limiting for reactive dye, such that labeling effectively ceases after 10 minutes. This contrasts with the 31.25 and 125 µM conditions where some proteins, such as PduC, continue to label over time (Figure 2.3a) indicating that AF488 is not a limiting reagent. Further, the higher concentrations appeared to bias the labeling towards the major shell proteins PduJBB' at earlier time points than for cargo (PduCP). We anticipate this may be due to the higher [AF532] resulting in higher reaction rates thereby derivatizing the BMC exterior quicker and more robustly than the diffusion of the dye into the lumen. Note, some proteins such as PduC are in high abundance (Yang *et al.*, 2020) and contain 7 cysteines, perhaps allowing a biased labeling profile. Lower concentrations of dye, however, may be able to capture labeling intermediates such as the triplet of bands found in the 1 and 5 µM time courses around 20 kD (Figure 2.3a) although this was not explored further. Lastly, it was found that BMCs denatured with SDS and heat exhibit quicker labeling and a different profile compared to intact BMCs (Figure 2.3b). In this sample, all potential bands appear at early time points with longer

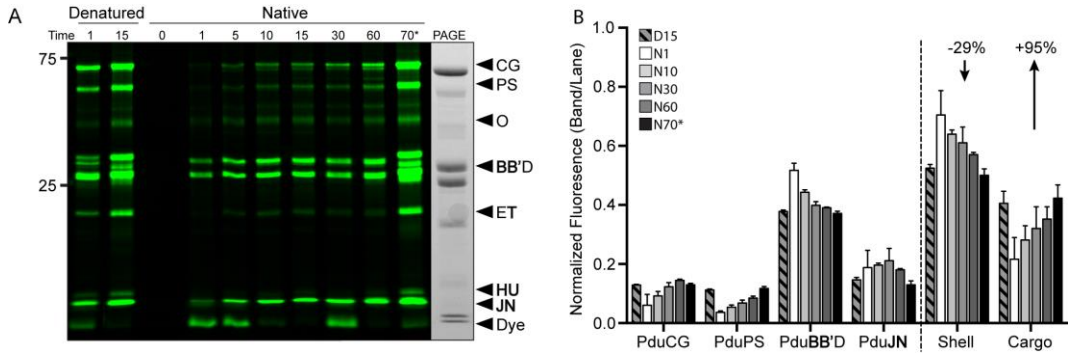


Figure 2.4: Chemical probing of Pdu BMCs. (A) Time course labeling of BMCs with the cysteine reactive fluorophore AF532. Here, native BMCs are reacted with AF532 *in vitro* before being quenched at the indicated timepoints. 70* refers to the 60-minute timepoint being denatured and allowed to label for another 10 minutes. Shell proteins are bolded. (B) Quantification of time course labeling. Bands are quantified as a fraction of their fluorescent contribution per timepoint (band/lane). Error bars are S.D. of three experimental replicates.

incubations only leading to more intense labeling as opposed to exposing new bands. This shows that the labeling profile observed in Figure 2.3a is due to a different, presumably intact, morphology compared to the denatured control. Overall, we concluded that the 125 μM AF488 condition was a suitable concentration for further testing.

We applied these conditions to label BMCs in triplicate to quantify if shell proteins exhibited a different labeling profile than cargo (Figure 2.4). We included two new caveats; (1) an intervening dialysis step between labeling/quenching and SDS-PAGE separation to remove excess dye and (2) the addition of a new time point (70*) where samples labeled for 60 minutes were then denatured and allowed to denature for another 10 minutes (Figure 2.4a). The 70* time point again reflects that excess dye is available and further that some proteins, like PduEP, only become highly labeled once exposed. Visually, the doublet of bands corresponding to shell proteins PduBB' label quickly and eventually saturate while cargo like PduCP label in a linear fashion (Figure 2.4a). Note, AF532 has a mass close to 1 kD so multiple derivatization events can cause mass shifts in the unlabeled versus labeled Pdu BMC gels (Figure 2.4a).

Bands were quantified as their percent contribution of the fluorescence in that entire time point (Figure 2.4b). For example, PduBB'D contributed ~50% of the total fluorescence after Pdu BMCs were labeled for 1 minute despite only being 16.5% of the Pdu BMC by copy number only (Yang *et al.*, 2020). The major cargo protein PduC, in contrast, only contributed ~5%. Over time the contribution of all shell proteins decreased by 29% and those for all cargo proteins increased by 95% (Figure 2.4b, right). This indicates that shell proteins label quickly and become saturated while cargo proteins are protected from labeling until dye continually diffuses into the BMC lumen. These results are consistent with the notion that BMC shells, while not impregnable, do protect the inner lumen. The nature of this unexpected permeability is speculated on more in Chapters 5 and 6. This method is also able to distinguish between shell and cargo proteins but suffers from low resolution inherent to one-dimensional SDS-PAGE from some Pdu BMC proteins having similar masses.

Subsection 2.2.4: A 2-Probe Protocol Contrasts Interior Versus Exterior Features

Our initial results indicated that some proteins only become labeled once Pdu BMCs are denatured in the presence of dye. We reasoned that a second probe could be used to label these proteins/remaining thiols after an initial labeling step saturated exposed proteins. In this scheme, Pdu BMCs were labeled at first with AF532 before the reaction was quenched and BMCs denatured. A second probe, AF680, was then added to label remaining cysteine residues (Figure 2.5a). We then turned to 2D-PAGE to better resolve the overlapping protein bands instead as individual spots (Figure 2.5b). Here, the sample is separated based on pI before being fully resolved by their differing masses.

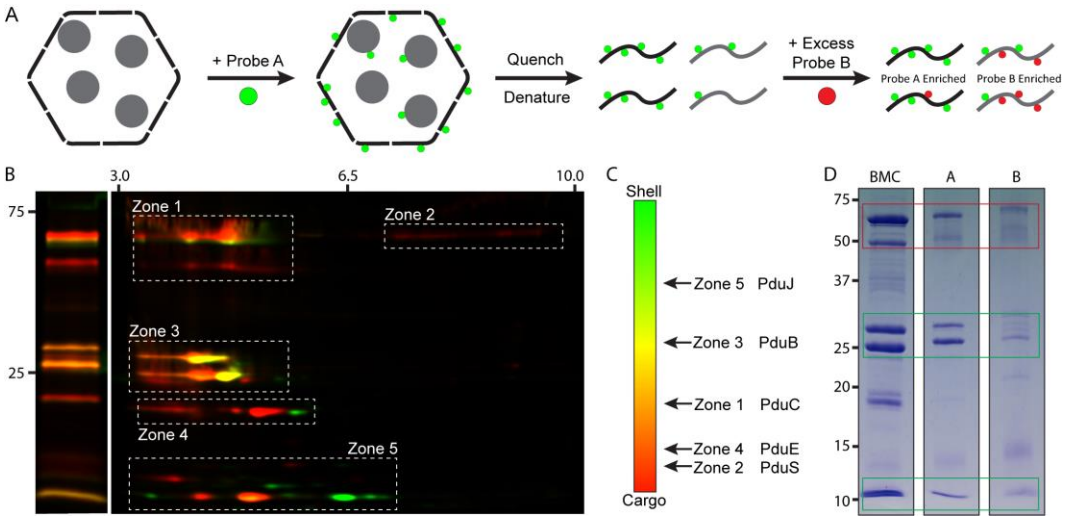


Figure 2.5: A modified labeling workflow that uses two probes. (A) A schematic of a dual-labeling workflow where BMCs are reacted with a first probe in a native confirmation before being denatured and a second probe is added to contrast structural features of the BMC. (B) 1D and 2D-PAGE of a dual-labeled BMC sample. 2D-PAGE allows for increased resolution of individual spot contributions. (C) Quantification of fluorescent intensity of specific zones from panel D. Zones are clusters of either mainly shell or cargo components. Red and green fluorescence is quantified separately, and the ratio of their intensities is plotted. A representative protein identified by peptide mass fingerprinting per zone is listed. (D) SDS-PAGE of dual-probing relying on biotinylation as the basis for separation. Lane A, elution of samples labeled with biotin then maleimide; Lane B, elution of samples labeled with maleimide then biotin. Green boxes correspond to shell protein bands, red boxes correspond to prominent cargo proteins.

Pdu BMCs treated in this way do result in separately colored foci when separated by 2D-PAGE and yield higher resolution than a corresponding 1D-PAGE (Figure 2.5b). This indicates that differential labeling occurred. The spot pattern in our 2D-PAGE results resemble previously reported 2D-PAGE for the *S. enterica* Pdu BMC (Havemann *et al.*, 2003) with a general shift towards acidic pI, an effect we attribute to the influence of the sulfate-containing Alexa Fluor dyes. Individual foci are more abundant with lower mass proteins, like PduJ (Figure 2.5b, zone 5) which only have one or a few cysteines. Interpretations become more complicated with higher mass proteins like PduC (Figure 2.5b, zone 1) with 7 cysteines. To simplify analysis, we separated 2D-PAGEs into distinct zones that can be characterized as being either majority shell or cargo proteins. Peptide mass fingerprinting was applied in conjunction

with fluorescence quantification of individual spots to confirm specific proteins and their bias for either green (exterior) or red (interior) labeling (Figure 2.5c). This confirmed that zones 3 and 5, containing the shell proteins PduB and PduJ respectively, are more biased for green labeling compared to zones 1, 2, and 5 which contain the cargo proteins PduC, PduS, and PduE, respectively. We conclude that using 2-probe protocols can be used to assay and characterize BMCs. This leads to a largely qualitatively separation of shell versus cargo components while also reporting on shell permeability.

Subsection 2.2.5: 2-Probe Contrasting Can Include Diverse Handles

Two-probe approaches need not be limited to fluorescent dyes. As proof-of-principle, our fluorescent probing scheme (Figure 2.5a) was reworked to replace AF488 and AF680 with plain maleimide and biotin-maleimide. This combination of probes would enable selective streptavidin pulldown of either shell or cargo components. Shell components may be preferentially selected by reacting Pdu BMCs with biotin-maleimide first while cargo proteins could be preferentially enriched by reacting with maleimide first to block accessible thiols. This approach worked well when biotin-maleimide was used first (Figure 2.5c, lane A) to pull down shell proteins. The major products from the pulldown were PduAJBB' (Figure 2.5c, green boxes) with minor contributions from PduCP (red boxes). In contrast, cargo selectivity was not observed with maleimide applied first as a blocking reagent (Figure 2.5d, lane B). This may be due to the drastically smaller size of maleimide compared to biotin-maleimide (or the Alexa Fluors) allowing its rapid diffusion into the BMC and blocking of most thiols. These results indicate that a diverse array of functionally selective handles may be used

to select for cargo or shell proteins but that considerations like probe size and reaction times need to be optimized.

Subsection 2.2.6: Chemical Modifications Do Not Impact BMC Morphology

The work described here assumes that the chemical modifications applied do not significantly impact the morphology of the Pdu BMC. Labeling results establish that native Pdu BMCs are very different from those that are denatured in the presence of SDS and heat (Figure 2.3b). The reagents used are also delivered to reaction aliquots in DMSO which itself may perturb protein structures (Tjernberg *et al.*, 2006; Arakawa *et al.*, 2007; Jackson and Mantsch, 1991). To alleviate these concerns, we directly assess the morphology of Pdu BMCs in 10% (v/v) DMSO (Figure 2.6a) - well above the maximum of 2% used in our assays. These BMC still largely appear as distinct structures (Figure 2.6a, top) similar to wild-type Pdu BMCs (Figure 2.2b). However, BMC clumping was also observed as well as tube-like structures adjoining some Pdu BMCs (Figure 2.6a, bottom). It is unclear if these filaments are a result of BMC denaturation or an artifact of grid dissolution due to the relatively high concentration of DMSO used. Regardless, these results indicate that Pdu BMCs can largely withstand DMSO as a buffer additive.

BMCs were also analyzed by DLS to evaluate if chemical modifications or buffer conditions perturb BMC size. Specifically, native BMCs were compared to those prepared in 2% DMSO, those labeled with 125 μM AF532, and Pdu BMCs crosslinked with the common amine crosslinker BS3 (5 mM) using DLS. DLS can report on the size of particles in solution and their polydispersity (a measure of homogeneity whose values range from 0, indicating homogenous sample, to 1) in a way that is free of

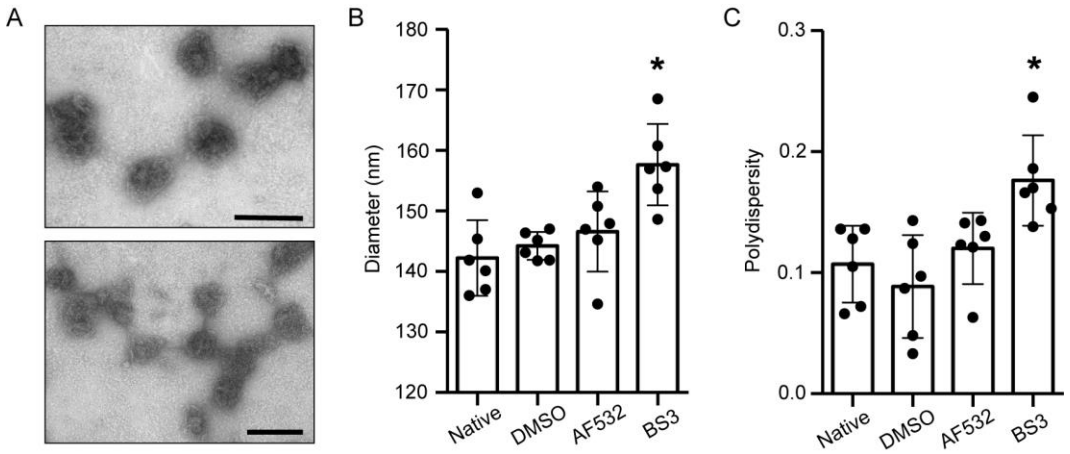


Figure 2.6: Pdu BMCs are agnostic to chemical modifications. (A) BMCs were incubated with 10% (v/v) DMSO and examined with TEM. BMCs form largely similar structures to native BMCs with some clumping. Scale bar, 200 nm. (B) Dynamic light scattering measurements of purified BMCs under different experimental conditions. BMC samples examined were either native or incubated with DMSO (2%) as a buffer control, cysteine reactive AF532 (125 μM), and lysine reactive BS3 (5 mM). (C) Polydispersity from DLS measurements. Asterisk indicates samples that were significantly different ($p > 0.05$).

warping artifacts common in TEM (Kennedy *et al.*, 2020). DLS measurements were obtained with two scans of biological triplicate Pdu BMCs for a total of 6 measurements. Native BMCs were found to have an average diameter of 142.2 ± 6.3 nm (instrumental error, not indicative of size ranges) which did not significantly differ from those treated with 2% DMSO (144.2 ± 2.3 nm) and AF532 (146.6 ± 6.6 nm) (Figure 2.6b). Those crosslinked with BS3 (157.7 ± 6.7 nm) did differ significantly. We reasoned that the AF488-treated and BS3-treated samples were reported as being larger than native Pdu BMCs because the chemical modifications add mass, thereby affecting their diffusion and calculated size. This effect is increased in BS3-crosslinked samples, perhaps due to a relatively high concentration of crosslinker but also because of the much higher prevalence of lysine residues compared to cysteines. BS3-treated BMCs again have the highest polydispersity (Figure 2.6c), which may result from other protein impurities becoming crosslinked to the BMCs or dead-end crosslinks existing on the shell exterior resulting in higher drag through the solution. The remaining

samples have polydispersities close to 0.1 which indicates being fairly monodisperse. Overall, these results allow us to conclude that Pdu BMCs can withstand a wide range of chemical modifications and buffer additives.

Section 2.3: Conclusions

These results describe the first instance of chemically modifying any class of native BMC *in vitro*. TEM and light scattering analysis (Figure 2.6) confirm that BMCs are largely agnostic to these modifications. While we specifically used the *S. enterica* Pdu BMC, we believe that any purifiable BMC can be chemically derivatized. Derivatizing Pdu BMCs *in vitro* is a facile, quick, and cheap (with proper preparation) process that requires no difficult genetic manipulations to add new functions. These methods showcase just the tip of the iceberg; it is easy to imagine future studies using click-chemistry (Moses and Moorehouse, 2007) or functional handles beyond fluorescence and biotinylation (Figure 2.5d). Bioconjugation holds the potential to fuel new insight into the structure/function of native BMCs or arm synthetic BMC shells with otherwise non-encodable functionalities for new applications.

The finding that Pdu BMCs are permissive to bioconjugation allowed us to develop fluorescent labeling methods. These methods, while largely qualitative, allowed us to assess the relative positioning of Pdu BMC proteins within the polyhedra. Specifically, shell proteins were found to label quicker than cargo proteins (Figure 2.4) when one probe was used. Introducing a second probe allowed for colored contrasting of the BMC interior and exterior (Figure 2.5).

Time course labeling showed that the exterior shell proteins PduBB'J were principally responsible for early labeling events and maximized their labeling potential

in as little as 15 minutes (Figure 2.4b). Interestingly, the Pdu BMC shell was not able to fully block probe access to the inner lumen. Luminal proteins, exemplified by PduC, did get fluorescently labeled albeit more slowly and never reached saturation under the conditions tested. It may be that the protein shell of the Pdu BMC is more porous than previously described. It would be interesting to compare labeling profiles of the Pdu BMC to the more regularly shaped carboxysome (Iancu *et al.*, 2010) or even other metabolosomes. It may be that carboxysomes, forming much more regular polyhedra, would have their cargo less prone to labeling than the results described here for the Pdu BMC. It is not known structurally why carboxysomes present as more regular than their metabolosome counterparts. Functionally, carboxysomes may have evolved to be more rigid to prevent detrimental O_2 exposure to RubisCO (Parry *et al.*, 2008; Cameron *et al.*, 2013). Chemical probes, akin those used in this study, will undoubtedly feature prominently in these future efforts to connect these structure-function relationships.

Section 2.4: Methods and Materials

Subsection 2.4.1: Probe Preparation

Fluorescent probes were prepared as smaller single-use aliquots to extend their utility and cut down on costs. AF532 and AF680 were resuspended to 8 mM in 100% anhydrous, degassed DMSO and portioned into 1.5 mL microcentrifuge tubes. Aliquots were then capped, sealed in parafilm, and stored under desiccation at -20°C under in the dark. Aliquots were equilibrated to room temperature just prior to use to prevent condensation. AF532 and AF680 were found to be highly reactive >6 months after being prepared as described.

Subsection 2.4.2: Pdu BMC Purification from a Heterologous Host

BMCs were purified in a similar way to previously described methods (Graf *et al.*, 2018). Briefly, A single colony of R995 + PduST was used to inoculate 100 mL of 2xYT supplemented with kanamycin and 0.5% 1,2-propanediol and grown overnight at 37°C for ~16 hours. The next day, the culture was pelleted and resuspend in 20 mL of 50 mM Tris-HCl pH 8.0, 200 mM KCl, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol (βME), 0.5 mM EDTA, 0.5 mg/mL lysozyme, 0.5 mM PMSF and 50% B-PER (Thermo) (v/v) and gently rocked at RT for 25 minutes and then placed on ice for 5 minutes. gDNA was then cleared by 2x1s pulses in a sonicator. Insoluble matter was pelleted for 20 minutes at 13000xg, 4°C. BMCs were pelleted at 17500xg, 4°C for 40 minutes. Pelleted BMCs were then resuspended in labeling buffer (20 mM phosphate pH 7.4, 50 mM KCl, 5 mM $MgCl_2$) and pelleted again for 20 minutes. The final BMC pellet was resuspend in 1 mL of labeling buffer and stored at 4°C for less than two weeks. BMCs were quantified via Bradford Assay against a BSA standard.

Subsection 2.4.3: Transmission Electron Microscopy

BMC samples were diluted in labeling buffer to 0.1 mg/mL and 5 µL applied to a formvar coated copper grid, 200 mesh (FCF200-CU, Electron Microscope Sciences). Samples were air dried for 15 minutes and excess was wicked off. Salt was removed by 3x5 µL additions of deionized water before staining with 3 µL of 2% uranyl acetate for 10 seconds. Samples were imaged using a JEOL 100CXII electron microscope.

Subsection 2.4.4: Labeling Assays

For the 1-dye assay, purified BMCs are diluted to 1.0 mg/mL in labeling buffer and split into a denatured control and native sample. The denatured control has SDS added to 0.5% and is heated at 95˚C for 5 minutes. Both samples were equilibrated to room temperature before an aliquot of AF532 was added to 125 µM in both samples. Aliquots were taken from each at the designated timepoints and quenched by adding βME to 5 mM. Aliquots were taken and dialyzed against deionized water to remove excess AF532. Samples were separated via standard SDS-PAGE and resulting bands were analyzed in ImageJ.

For the 2-dye assay, purified BMCs are diluted to 1.0 mg/mL as above. AF532 was added to 125 µM and BMCs were labeled for 15 minutes at RT before being quenched as above. Samples were then dialyzed against 20 mM Tris-HCl pH 7.4, 0.5% SDS to remove excess βME before being heated at 95˚C for 5 minutes. AF680 was then added to 250 µM and samples were labeled for 90 minutes at RT and quenched as above. Samples were then alkylated with 20 mM 2-chloroacetamide for 1 hour at RT protected from light. AF680 was difficult to remove via dialysis, so samples were subject to buffer exchange columns to remove excess AF680 and 2-chloroacetamide. Samples were equilibrated into 2D-PAGE rehydration buffer (8 M urea, 0.25% CHAPS, 0.5% 3-10 NL ampholytes, 20 mM DTT, trace bromophenol blue) for 2D separation.

For the biotin pulldown, purified BMCs were diluted to 1 mg/mL in labeling buffer and equilibrated to RT. Probe A was added to 125 µM and BMCs were labeled for 5 minutes. Reactions were quenched by addition of βME to 625 µM followed by

denaturation by adding SDS to 0.5% and heating. Once cooled to RT, probe B was added to 1 mM and reacted for 30 minutes. Samples were then dialyzed overnight into pulldown buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl). Dialyzed, labeled samples were incubated with equilibrated NeutrAvidin™ Agarose (Pierce) for 15 minutes on ice. Supernatant was collected and the beads were washed once with pulldown buffer. Bound proteins were eluted by mixing the beads with 1x standard SDS-PAGE sample buffer + 2 mg/mL D-biotin with heating for 1 minute at 98°C.

Subsection 2.4.5: 2D-Electrophoresis and Peptide Mass Fingerprinting

The ZOOM® IPGRunner™ System (Invitrogen) was used for 2D separation. Briefly, samples were diluted into rehydration buffer and approximately 15 µg of protein sample was loaded onto a pH 3-10 NL IPG strip and equilibrated for 6 hours at RT. Samples then underwent isoelectric focusing overnight with stepwise increases in voltage and held for final focusing overnight (16 hours, 2000V). After focusing, strips were retrieved and incubated in 1x NuPAGE LDS Buffer supplemented with 5 mM DTT for 5 minutes prior to separation via 4-20% PAGE. Gels were fluorescently scanned prior to silver staining/destaining of excised spots with SilverQuest™ Silver Staining Kit (Invitrogen).

Excised gel spots were washed once in 25 mM NH_4HCO_3 before being dehydrated in acetonitrile for 15 minutes at RT. Alkylation was skipped since samples had already been alkylated. Dehydrated gel pieces were swelled in 50 µL of 20 ng/µL of Trypsin/LysC (Promega) in 25 mM NH_4HCO_3 for 30 minutes at 37°C then excess supernatant was removed and digestion was allowed to continue overnight at 37°C. After digestion, peptides were desalted and extracted using C18 tips (Pierce) and eluted

in 20 µL of 0.1% TFA, 50% acetonitrile. Samples were analyzed in a Bruker Autoflex Speed MALDI-TOF in DHB/CHCA matrices in linear mode. Peaks were analyzed in MS-Fit using the NCBI non-redundant database for Salmonella species. A mass tolerance of ±2 Da was used and a minimum of 3 peptides was needed for a match.

Subsection 2.4.6: Dynamic Light Scattering

Labeling samples were prepared as described in other sections. BMC samples were diluted to 0.1 mg/mL in labeling buffer and equilibrated to room temperature immediately prior to measurement. Light scattering measurements were performed on a Nanobrook Omni Particle Sizer with 90˚ scatter and analyzed using the CONTIN model. Three biological replicates were measured twice for a total of 6 measurements per condition. Each scan was performed for 90 seconds at RT.

Chapter 3: The Organization of the Pdu BMC Lumen

Section 3.1: Introduction

Subsection 3.1.1: BMCs are Incompatible with Traditional Structure Analysis

Native BMCs, like the Pdu BMCs at the heart of this dissertation, are large (>100 nm diameter, >100 megadaltons) and are somewhat heterogeneous. The structural data on BMCs primarily comes from synthetic BMCs that are far smaller than naturally occurring counterparts (40 nm diameter at most (Sutter *et al.*, 2017)), only require a few different shell proteins, and have no enzymatic cargo. While all BMC structural studies to date have undoubtedly benefited from these simplifications, valuable insight is lost. How do BMC shells that use many different shell proteins assemble? How does cargo interact with the shell in a native setting? How does cargo assemble into the luminal space? What happens, in total, if one wants insight into the construction of a complete, large, native BMC?

The Pdu BMC is large (~140 nm), mildly heterogeneous (Figure 2.6), and is made of at least a dozen different proteins of varying stoichiometries (Figure 2.2 (Yang *et al.*, 2020)). Accordingly, traditional structural approaches, like X-ray crystallography, NMR, and cyro-EM, are unlikely to succeed with assessing the structure of the native Pdu BMC. For example, crystallography would suffer from the sample heterogeneity. Even if the outer shell was regular enough to permit crystal growth, there is far less of a guarantee of that for the inner milieu. A case-in-point comes from encapsulin nanocompartments which, while small like synthetic BMCs,

have been crystallized with cargo; however, the electron densities of the interior could not be assessed (Sutter *et al.*, 2008). Further, crystallographic data does not exist for all possible Pdu BMC proteins, further complicating model development. NMR in contrast, can readily accept dynamic, heterogeneous assemblies but suffers from complicated spectra with larger biomolecules. One of the largest proteins analyzed by NMR to date is the 670 kD 20S proteasome (Bhattacharyya *et al.*, 2014; Jiang and Kalodimos, 2017) but those experiments utilized isotopically labeled leucine, isoleucine, and valine residues while also undoubtedly benefiting from a pre-existing crystallography-based structure for homology modeling (Löwe *et al.*, 1995). In contrast, cyro-EM may prove to be a successful approach, having already been applied to assess the synthetic Hoch BMC shell (Sutter *et al.*, 2017) as well as the similarly sized 125 nm herpes simplex virus type-1 capsid (Yuan *et al.*, 2018). Cryo-EM is also able to differentiate and subclassify conformationally distinct particles. This was recently exemplified by a 2.1-2.7 Å cyro-EM structure of the 6.2 megadalton phycobilisome, which was captured in several previously unknown conformations (Sauer *et al.*, 2021). These analyses, however, may again be better suited for analyzing shells and capsids as opposed to native, filled BMCs.

Subsection 3.1.2: Using Structural Proteomics to Study BMCs

Crosslinking mass spectrometry (XLMS) can circumnavigate many of the limitations of other structural methods. XLMS is an emerging structural proteomics workflow where protein samples are crosslinked to covalently couple spatially close, interacting proteins (Chu *et al.*, 2018). Lysine residues are the usual crosslinking targets of NHS-ester based crosslinkers (Figure 3.1a (Mädler *et al.*, 2009)) although a diverse

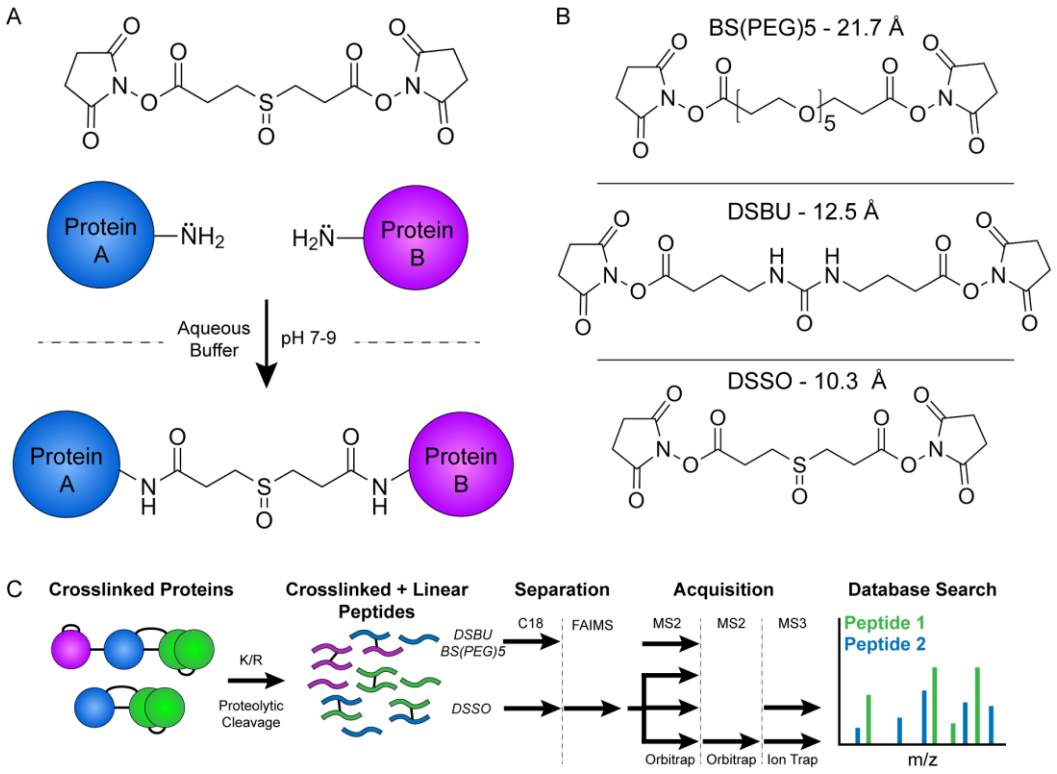


Figure 3.1: XLMS overview and crosslinked used in this study. (A) NHS-ester crosslinkers (BS3, shown) react with primary amines (lysine residues) on protein surfaces forming a peptide bond. If two proteins are close enough, the crosslinker will covalently link them by coupling the proximal lysine residues. (B) The three crosslinkers used to study the Pdu BMC interactome. (C) A schematic of the mass spectrometry workflow to analyze crosslinked samples in this work. Crosslinked proteins are proteolytically cleaved at lysine and arginine residues which results in a mix of crosslinked and linear peptides. Samples crosslinked with DSBU and BS(PEG)5 are separated with a C18 column prior to MS2 data acquisition on an Orbitrap instrument. Samples crosslinked with DSSO are serially separated with a C18 column followed by field asymmetric ion mobility spectrometry (FAIMS) before data was acquired in the MS2, MS2/MS3, and MS2/MS2/MS3 modes.

array of protein crosslinkers are now available to target other residues (Matzinger and Mechtler, 2021). The crosslinks are subsequently analyzed by high-resolution mass spectrometry (O'Reilly and Rappsilbur, 2018). Numerous workflows have been developed to analyze the spectra obtained by a variety of protein crosslinkers (Figure 3.2b). One particularly successful strategy employs crosslinkers which are cleavable within the mass spectrometer (Sinz, 2017). The fragmentation patterns that result from crosslinker cleavage results in specific mass shifts which act a fingerprint for spectra corresponding to crosslinked peptides, thereby triggering further MS^n fragmentation

(O'Reilly and Rappsilbur, 2018; Figure 3.1c). Accordingly, cleavable crosslinkers simplify the database search space for two linked peptides from n^2 to 2n. The XLMS methodology can derive residue-level resolution on where the crosslinks occur, what proteins are crosslinked, and, with the spacer arm of the crosslinker, the distances of those crosslinks. The data that is captured from these analyses can help build interactome maps of protein networks, delineate binding sites (Vreven *et al.*, 2018), and help with model building (Brodie *et al.*, 2017). With enough crosslinks, XLMS data can even report on conformation dynamics (Minteris and Gygi, 2020).

The chemical modification data presented in Chapter 2 reveals that the Pdu BMC is permeable to chemical reagents. Further, Pdu BMCs display a tolerance to these modifications. Altogether, the data indicate that structural information can be extracted from the protein shell and oft neglected lumen, while largely leaving the BMC intact. Spurred by these observations, we employed a battery of three different crosslinkers to capture structural data for the Pdu BMC *in vitro* (Figure 3.2b). These crosslinkers are all amine reactive but differ in their spacer arms lengths and hydrophilic qualities. Crosslinks are identified among both the outer shell, the inner cargo as well as those that detail their connection. This chapter focuses on the results and conclusions drawn from protein crosslinks captured among the interior of the Pdu BMC; Chapter 4 deals instead with the outer shell.

Section 3.2: Results and Discussion

Subsection 3.2.1: Crosslinking Pdu BMCs for Structural Proteomics

Crosslinking of Pdu BMCs was first tested using the common amine crosslinker BS3 (Figure 3.2a). BMCs were purified and diluted to 1.0 mg/mL in phosphate buffer at pH 7.4. This pH is within the appropriate range for NHS ester crosslinking of primary amines (Mädler *et al.*, 2009) while also not being buffered by the primary amine Tris, which would compete with and quench the crosslinking reaction. When the crosslinking reaction is resolved by SDS-PAGE, low mass protein bands disappear while high molecular weight species appear in a concentration-dependent manner (Figure 3.2a, red arrow). The high mass bands partition at the interface between the stacking (4%) and resolving (16%) gels. A control sample was included where BMCs were first denatured with SDS and heat to liberate protein interactions before BS3 was introduced (Figure 3.2a). This sample was not able to appreciably form high molecular weight species, although some band smearing is seen, which likely occurred from

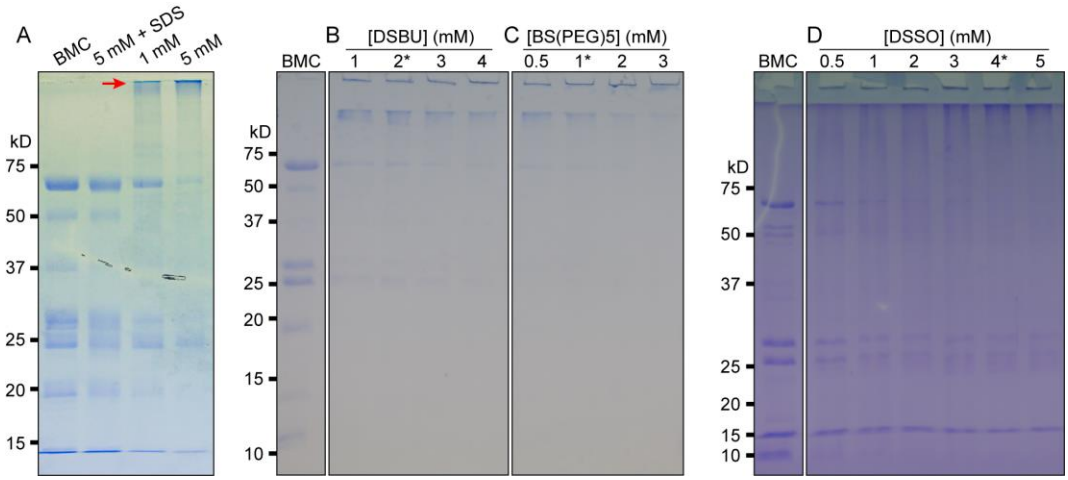


Figure 3.2: Optimizing Pdu BMC crosslinking. (A) Crosslinking was initially tested with the crosslinker BS3. Crosslinking products appear as high molecular weight bands (red arrow). Denatured BMCs cannot form these bands. Pdu BMCs were crosslinked with (B) DSBU, (C) BS(PEG)5, and (D) DSSO. Lanes with asterisks denote the samples sent for XLMS.

intramolecular crosslink formation, thereby affecting migration. This control shows that the crosslinks observed are morphology dependent.

Initial tests with BS3 gave confidence in the conditions necessary to form crosslinks. From here, Pdu BMCs were assayed with the amine-reactive crosslinkers DSBU, BS(PEG)5, and DSSO that were selected for XLMS (Figure 3.1b). DSBU and DSSO have similar spacer arms to BS3 but are cleavable to allow MS3 identifications (Sinz, 2016; Matzinger and Mechtler, 2021), while BS(PEG)5 was selected for its long spacer arm. The concentration of each crosslinker was optimized to (1) see the appearance of high mass species, indicating successful crosslinking, and (2) prevent over-crosslinking which may inhibit downstream proteolytic digestion and sequence coverage (Figure 3.2b, c and d). Visual assessment of the resolved crosslinks show that BS(PEG)5 (Figure 3.2c) was able to achieve the greatest mass shifts, likely from having the longest spacer arm (Figure 3.1b). Regardless, all three crosslinkers were able to sufficiently crosslink BMC samples. Pdu BMCs crosslinked with 2 mM, 1 mM, and 4 mM of DSBU, BS(PEG)5, and DSSO, respectively (Figure 3.2b, c, and d, asterisks) were selected for mass spectrometry analysis. This combination of cleavable and non-cleavable crosslinkers and different spacer arm lengths was coupled to several different modes of data acquisition (Figure 3.1c). Each data acquisition mode identified different crosslinks and allowed deeper peptide and crosslink coverage than any one mode would provide alone.

Subsection 3.2.2: Overview of the Pdu BMC Interactome

The Pdu BMC has yet to have its interactome described globally. Some work has attempted to model the Pdu BMC interactome using co-evolution descriptors (Jorda

et al., 2015) but there is scant biochemical evidence to support many of the predictions. Using XLMS with several different crosslinkers allowed us to capture 230 unique crosslinks within the Pdu BMC system (Figure 3.3a). These crosslinks describe interactions between 16 of the 21 proteins encoded by the *pdu* operon. Crosslinks were not found on PduHUVX, likely a result of poor sequence coverage brought on by their low abundance (Yang *et al.*, 2020). PduN, the vertex protein, was identified with a sole intramolecular crosslink.

The crosslinks can be divided into three categories; (1) those between cargo proteins (161 crosslinks), (2) those between shell proteins (56 crosslinks), and (3) those between cargo and shell proteins (13 crosslinks, Figure 3.3a). When these crosslinks are mapped out, the Pdu BMC clearly segregates into a distinct cargo and shell domain (Figure 3.3b). Most crosslinks within the cargo domain are between PduCDELP with peripheral interactions captured to the proteins PduGOSQW. Results and interpretations concerning the protein shell will be expanded upon in Chapter 4. Tables containing all crosslinks identified can be found in the Appendix.

Subsection 3.2.3: Encapsulation Peptides Act as a Cargo Aggregator

It is not clearly understood how EPs operate in a native setting. Knowledge in this regard is critical for their successful recapitulation in synthetic settings. Accordingly, the crosslinks were assessed for any information directly concerning EPs. Four proteins in the Pdu BMC, PduDELP, are known to contain N-terminal EPs. These four proteins form a tight-knit, central interaction network within the BMC (Figure

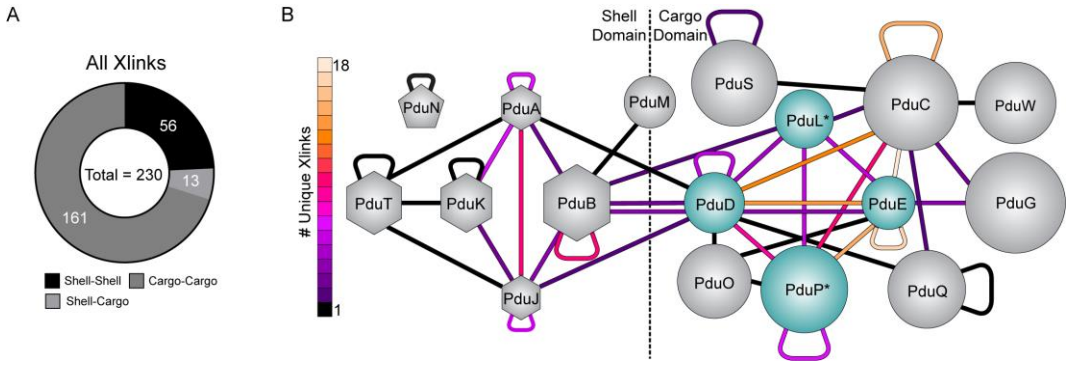


Figure 3.3: A summary of the crosslinks captured. (A) Pie chart of the unique cross-links identified. All cross-links fit one of three categories depending on the type of proteins captured in the link. (B) An interactome heat map describing the crosslinks identified. Crosslinks are depicted as lines connecting proteins colored according to the number of unique cross-links identified between the two proteins. Proteins are scaled by their amino acid length with shell proteins shown as hexagons/pentagons and cargo proteins as circles. Blue cargo proteins contain encapsulation peptides. Asterisks denote cross-linking partners of an encapsulation peptide tagged GFP protein, $GFP^{P1\text{-}18}$.

3.3b, blue circles). PduC, the diol dehydratase major subunit (Daniel *et al.*, 1998), also forms a high number of crosslinks to PduDELP. PduC, however, does not encode for an EP but forms a dimer of trimers with PduDE which do have EPs.

The crosslinks among EP-containing cargo make up most cargo crosslinks identified. Of the 161 cargo crosslinks, 85 are between EP-containing cargo ($cargo^{EP}$, Figure 3.4a), 51 are between cargo with EPs and cargo without, and only 25 were found between non-EP fused cargo (Figure 3.4a). The prevalence of these crosslinks may be understood in terms of protein stoichiometry, where EP-containing proteins are highly abundant within the lumen (Yang *et al.*, 2020), but also in terms of interaction stoichiometry, where EPs form numerous interactions and may aggregate these proteins (Juodeikis *et al.*, 2021). In addition, all EPs are guaranteed at least one crosslinking site at their N-terminus. These factors make their crosslinking more likely.

XLMS data can be broken down further to the residue level. Crosslinks could be identified directly to the EPs, defined as the first 18 amino acids (Aussignargues *et al.*, 2015), of PduDELP. The crosslinks to EPs were further sub-classified as either

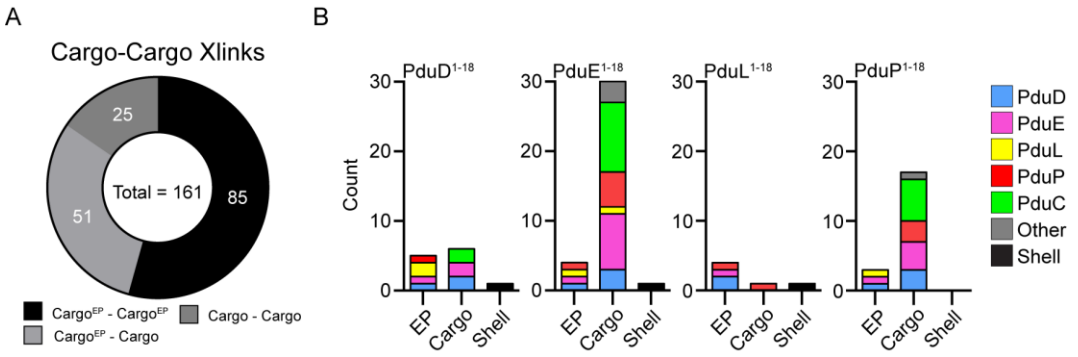


Figure 3.4: The interactions captured among cargo proteins. (A) Summary of the cargo-cargo cross-links captured. Most crosslinks occur among encapsulation peptide containing proteins (cargo^{EP}). (B) An enumeration of the crosslinks occurring among EPs. Crosslinks to EPs are tabulated as occurring directly to another EP, to the main body of another cargo protein, or to shell proteins.

being to another EP directly, to the main body of a cargo protein, or to a shell protein (Figure 3.4b). This breakdown makes it glaringly apparent that, in contrast to the literature, EPs largely do not interact broadly with shell proteins (Figure 3.4b). Non-EP containing cargo are likewise largely excluded from these crosslinks. Rather, EPs prefer to interact with other EPs or with EP-fused proteins. These data are in line with other observations that EPs can act as aggregation factors (Juodeikis *et al.*, 2021; Lee *et al.*, 2016).

One may argue that EP-crosslinks may be enriched if the interior of the Pdu BMC was sufficiently dense. However, the same argument can equally apply for hypothetical EP-shell crosslinks. If EP-shell interactions were crucial, then they too must be in a relatively high concentration. Further, they would be spatially more accessible than most luminal interactions that were observed since they would necessarily lay closer to the exterior of the compartment. Despite this, EP-shell crosslinks are not found in high abundance. This data is instead reflective that EPs mainly drive cargo aggregation, specifically with other EP-tagged cargo.

Subsection 3.2.4: Non-native EP-Tagged Cargo Interacts with Other Cargo Proteins

Whether EPs act as a protein aggregator in a native setting should be reflected in attempts to add new proteins to the Pdu BMC via an EP tag. To test this, we co-expressed Pdu BMCs with an arabinose-inducible GFP that included the EP of PduP on its N-terminus and an SsrA proteolytic tag on the C-terminus. This GFP, deemed $GFP^{P1\text{-}18}$, can be protected from proteolytic cleavage upon encapsulation into the BMC lumen (Jakobson *et al.*, 2015). If purifiable, $GFP^{P1\text{-}18}$-containing BMCs can be crosslinked to discover the interaction partners of non-native cargo.

Induction of $GFP^{P1\text{-}18}$ (+ arabinose) without co-expressing Pdu BMCs (- 1,2 propanediol) in *E. coli* led to mostly dark cells with some cells exhibiting diffuse fluorescence, indicating digestion of the GFP (Figure 3.5a). This contrasted with cells co-expressing Pdu BMCs, which demonstrated bright green foci (at least 1 per cell) indicating that the $GFP^{P1\text{-}18}$ was protected by the BMC. Further, these foci were distributed throughout the cell, suggesting they are not inclusion bodies (Rinas *et al.*, 2017). $GFP^{P1\text{-}18}$-containing Pdu BMC could also be purified like native BMCs. These BMCs demonstrated a similar protein content as native Pdu BMC but included a new band at ~32 kD (Figure 3.5b, red box), corresponding to $GFP^{P1\text{-}18}$ and, perhaps consequently, a slightly reduced amount of PduP (Figure 3.5b, black box). Purified $GFP^{P1\text{-}18}$ Pdu BMCs were then crosslinked using DSBU (Figure 3.5c) with 2 mM found to be an appropriate concentration. Samples were then analyzed via XLMS with an identical workflow as their native DSBU-crosslinked counterparts to see if $GFP^{P1\text{-}18}$ will be incorporated into the EP subnetwork.

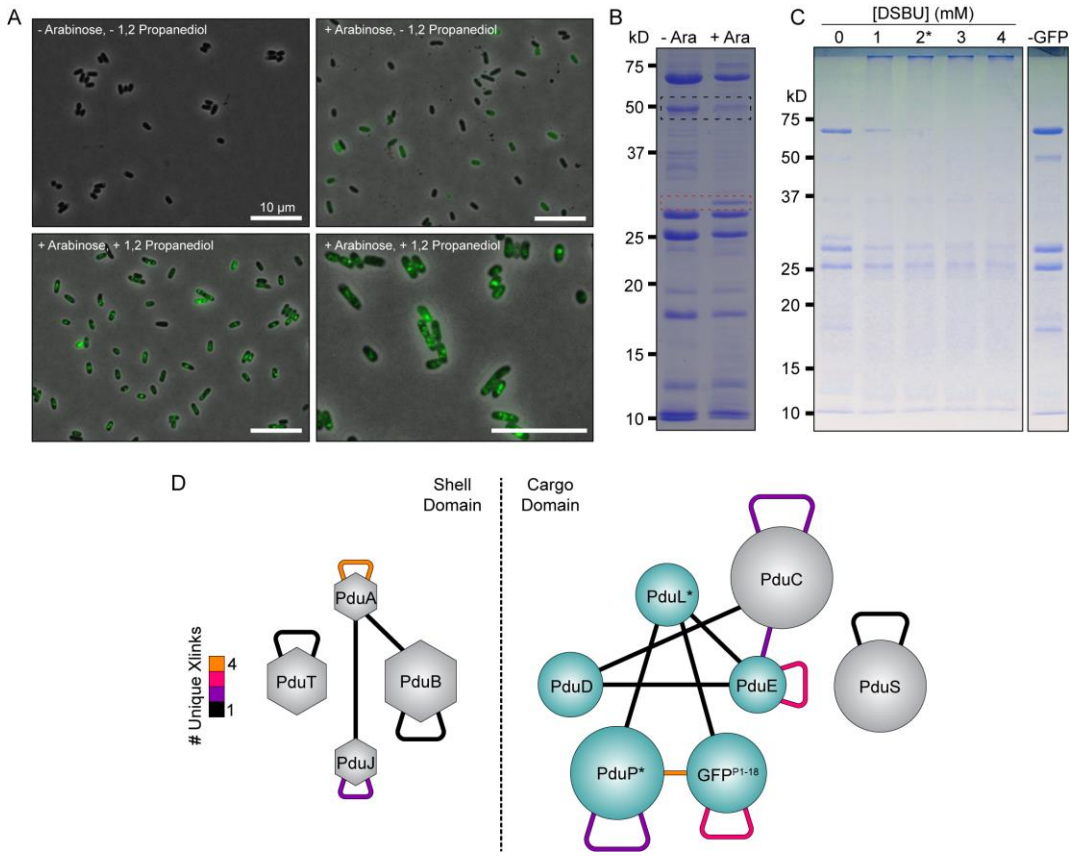


Figure 3.5: Non-native cargo incorporates into a subnetwork of the Pdu BMC interactome. (A) The ability of the GFP^{P1-18} reporter to incorporate into Pdu BMC *in vivo* was tested. This reporter contains a SsrA proteolytic tag that will result in cleavage of non-encapsulated GFP. The reporter can only form bright green foci when protected within the lumen of the Pdu BMC. (B) Purifying Pdu BMCs co-expressing the *gfp* gene result in co-purification of GFP^{P1-18} (red box). The abundance of PduP suffers marginally (black box). (C) GFP containing Pdu BMCs were crosslinked with DSBU. The asterisk denotes the sample sent for XLMS analysis. (D) XLMS analysis of GFP containing Pdu BMCs captured 19 unique crosslinks. Crosslinks among the protein shell are conserved with wild-type Pdu BMCs. GFP^{P1-18} incorporates in the EP-dependent interaction subnetwork.

Nineteen crosslinks were identified from Pdu BMC containing GFP^{P1-18}. GFP^{P1-18} was found to integrate into the EP-driven subnetwork with crosslinks identified to the EP-containing proteins PduP and PduL (Figure 3.5d, asterisks). These crosslinks also included direct links between the EPs of PduP and PduL. Note, the EPs of native PduP and GFP^{P1-18} cannot be differentiated but we contend that they act similarly in a native setting. No crosslinks between GFP^{P1-18} and shell proteins were identified. Perturbation of the luminal cargo network also did not appear to influence the

crosslinks captured among the protein shell as crosslinks between PduB^{138}-PduA^{55} and PduA^{55}-PduJ^{54} are maintained throughout all datasets (see Appendix). From this we speculate that the shell may scaffold on, but assemble independently of, the inner core. This data supports the notion that EPs are modular and encourage cargo aggregation as the mode of trafficking.

Subsection 3.2.5: Piggybacking Interactions Are Observed in the Pdu BMC

Cargo proteins without known EPs were identified in 51 crosslinks to those with EPs in a native setting (Figure 3.4a). Among these is the reactivation factor PduG (Toraya and Mori, 1999), which is known to interact with the signature complex PduCDE by displacing PduD (Mori *et al.,* 2005; Shibata *et al.,* 2005). Expectedly, PduG only forms cross-links between PduC and PduE but not PduD (Figure 3.4b). Other exceptions also include the cobalamin reductase PduS, all interacting with the signature complex PduCDE, and the corrinoid adenosyltransferase PduO with PduDEP (Figure 3.4b). Given that PduDEP all contain EPs, it is tempting to speculate that PduO may itself contain an unidentified EP, although more crosslinks/biochemical data would help back this argument. It should also be noted that PduGOS all work to recycle B_{12} within the BMC lumen (Figure 2.2a), so it is unsurprising to find them located within the crosslinking range of the B12-dependent diol dehydratase PduCDE.

Intermolecular crosslinks were likewise identified with the non-B_{12} recycling enzymes PduW and PduQ. PduW is a propionate kinase (Palacios *et al.*, 2003), which has recently been proposed to be located outside of the inner lumen (Yang *et al.*, 2020) while PduQ is a propanol dehydrogenase that helps internally recycle NADH along with PduP (Cheng *et al.*, 2012). These two proteins formed several crosslinks to the

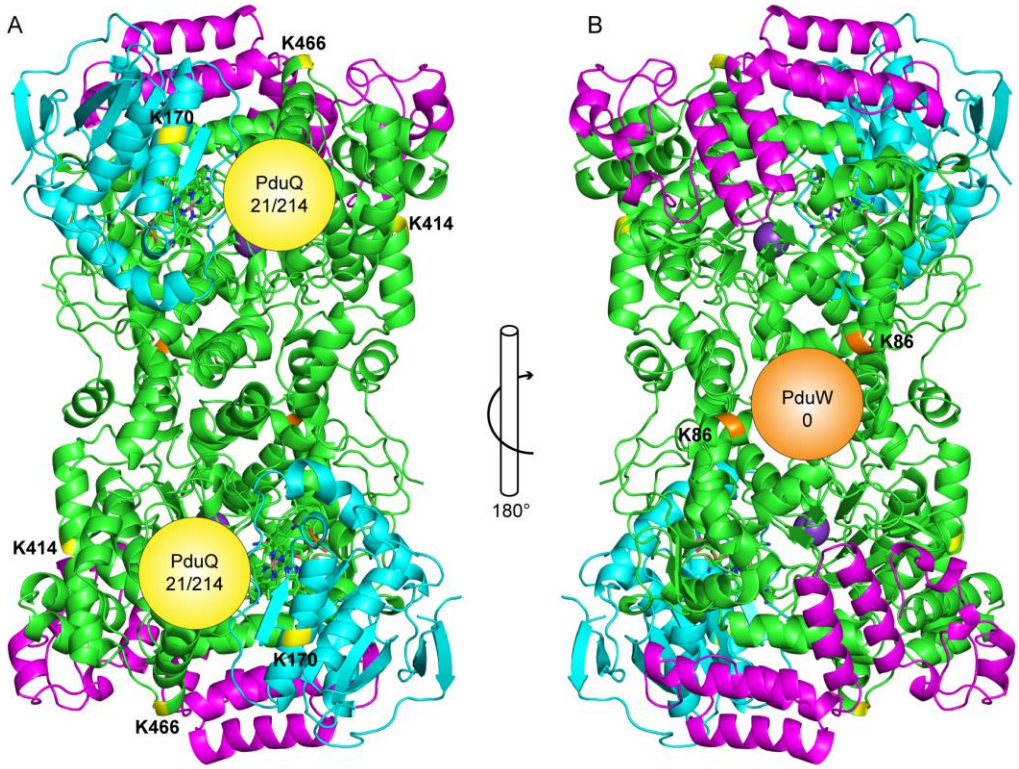


Figure 3.6: XLMS reveals sites of piggybacking on the signature complex PduCDE. Crosslinks to PduQ and PduW are mapped into a homology model of *S. enterica* PduCDE (1DIO (Shibata *et al.*, 1999)). PduC, green; PduD, cyan; PduE, magenta; PduQ yellow; PduW, orange. PduQ crosslinks to yellow lysines on the signature complex and PduW crosslinks likewise to orange colored lysines. PduQs relative positioning is triangulated between PduC^{414,466} and PduD^{170} (A) on the opposite face of PduW (B). The opposing locations of the crosslinks can be interpreted as sites of so-called "piggybacking" interactions, wherein non-EP tagged cargo can assemble into the Pdu BMC by interacting with proteins that do have EPs, such as PduDE within the signature complex.

signature complex PduCDE (Figure 3.4b) that may indicate discrete binding sites. To investigate this, a homology model of *S. enterica* PduCDE was built and the crosslinking positions to PduQW were mapped onto it (Figure 3.6). The three crosslinks between PduCDE and PduQ triangulated to a specific area on either of the poles of the signature complex (Figure 3.6a), specifically PduC^{414,466} and PduD^{170}. PduW, meanwhile, forms a crosslink to PduC^{86} on the opposite face of the complex from PduQ (Figure 3.6b). Due to the dimerization of PduC near K86, there are two redundant K86 residues in close proximity. Further, the presence of PduW among cargo

crosslinks across several datasets/biological samples indicates it is indeed part of the luminal composition.

These results indicate that the signature complex may act as a hub for many piggybacking interactions. Namely, PduGOQSW were captured interacting predominantly with PduCDE. However, this data does not offer a more detailed mechanistic explanation of this phenomena. However, prior to our study, the protein-protein interactions that guide PduQ, PduO, or PduS to the inner lumen remain were unknown, although PduQ had been suggested to interact with PduP (Cheng *et al.*, 2012) and PduS with PduT (Cheng and Bobik, 2010; Jorda *et al.*, 2015; Parsons *et al.*, 2008). These results are also consistent with the notion that the B_{12} recycling factors PduGHOS do not form a larger B_{12}-recycling complex within the BMC lumen but instead interact individually with the signature complex. The GTPase PduV has also been shown to be a part of the Pdu BMC (Parsons *et al.*, 2010; Yang *et al.*, 2020) and potentially interacts with the shell protein PduU (Jorda *et al.*, 2015). The data reported here does not corroborate these findings, likely owing to the low abundance of both PduU and PduV (Yang *et al.*, 2020).

Section 3.3: Conclusions

XLMS analysis of the Pdu BMC luminal interaction network led to new insight into how it is natively assembled. One critical finding is that EPs tend not to interact with the protein shell, but rather other EP-fused protein cargo (Figure 3.4b). This phenomenon appears to be a conserved mode of action, as a non-native cargo containing an EP was found to make similar contacts (Figure 3.5d). Crosslinks that include EPs account for nearly a third of the cross-links we identified, yet only three

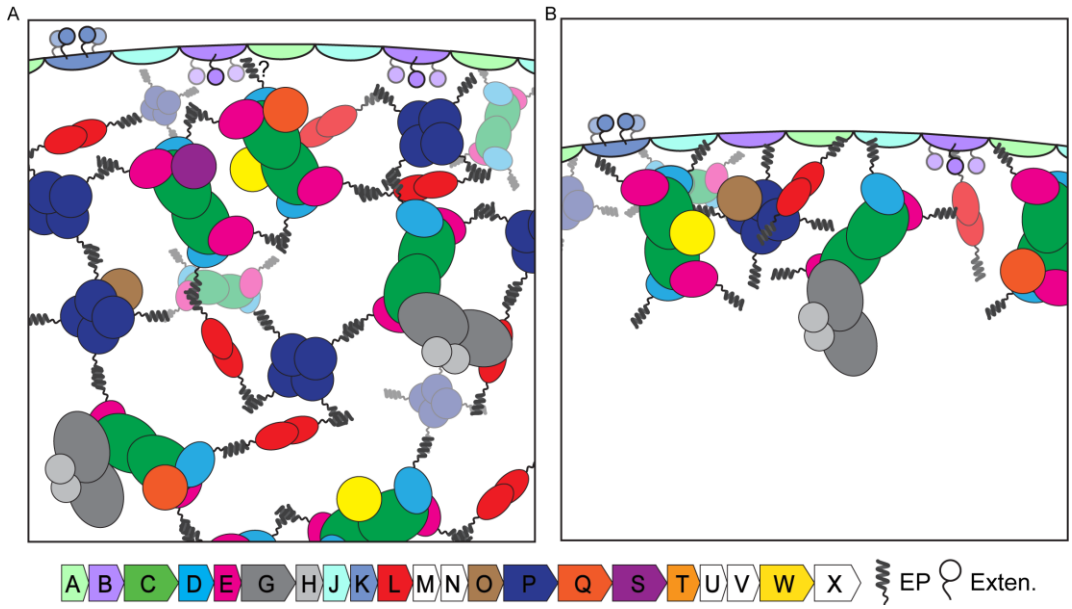


Figure 3.7: Models for the native assembly state of the Pdu BMC lumen. (A) A model where the Pdu BMC lumen is filled cargo proteins extending throughout the lumen. This assembly state is governed by EPs forming a mismatched, multivalent interaction network. The luminal space is largely disorganized and may be like a shell-constrained biomolecular condensate. (B) A model where EP-tagged cargo largely interacts with the protein shell as opposed to each other. This mode of assembly entails most enzyme localizing to the interior surface of the shell. The data reported here support the model shown in panel A.

were to shell proteins. The hypothesis going into this work was the opposite; if interactions between EPs and the protein shell were commonplace and located at the easily accessible shell-lumen interface, then shell-cargo cross-links should dominate. Our data instead calls into question the currently understood mechanism of action for EPs.

The ability for EP-containing cargo to aggregate is one of the only consistent rules dictating the assembly state. These results support a distinct vision of the assembled Pdu metabolosome where EP-fused cargo proteins act as a random aggregate extending throughout the BMC lumen with other accessory proteins piggybacking along (Figure 3.7a). This contrasts with some models which dictate cargo proteins scaffold exclusively along the interior face of the protein shell (Figure 3.7b).

It is important to note that none of the EP-containing proteins in the Pdu BMC is a monomer. PduP is likely a tetramer (Zarzycki *et al.*, 2017), PduL is at least a dimer – interestingly influenced by the presence of its EP - (Erbilgin *et al.*, 2016), and PduCDE exists as a dimer of trimers (Masuda *et al.*, 1999). The widespread interactions captured between these proteins, coupled to their mismatched multivalency, invites comparisons to the 'stickers and spacers' model used to describe biomolecular condensation (Banani *et al.*, 2017; Kahn *et al.*, 2021). From this, we speculate that the inner lumen might be a disordered biomolecular condensate that is constrained by an exterior protein shell (Figure 3.7a). This may suggest that efforts to synthetically manipulate BMCs, such as incorporation of designer cargo, may need to consider using multiple EP fusions with mismatched valences.

These results not only describe the native luminal assembly state but also may hint at the biogenesis of the Pdu BMC. From our aggregate data, we speculate that the Pdu BMC is a plastic system governed by a multitude of multivalent interactions. At this stage, our data is consistent with a model where the inner cargo first undergoes LLPS to form an EP-driven pro-metabolosome (or 'core-first assembly', (Kerfeld *et al.*, 2018)) prior to shell envelopment (expanded on in Chapter 6). This model shares features with the assembly pathway that has been proposed for Rubisco during carboxysome assembly (Wang *et al.*, 2019). A prediction of this overall model is that if the cargo undergoes LLPS, it should exhibit a strong reliance on mismatched valences of protein-protein interaction points, such as EPs. Given this, it is interesting that, anecdotally, the BMCs containing $GFP^{P1\text{-}18}$ resulted in less than half the yield per purification preparation, compared to wild-type BMCs. Perhaps the monomeric state

of GFP^{P1-18} resulted in dead-end network extensions, thereby perturbing the proper assembly pathway. Designing multivalency in conjunction with EPs, therefore, may be a viable bioengineering target. This may encourage heterotypic interactions over the homotypic interactions (Riback *et al.*, 2020) that led to the insoluble aggregation that complicated early native cargo protein purification efforts (Leal *et al.*, 2003; Masuda *et al.*, 1999; Shibata *et al.*, 2010).

Section 3.4: Methods and Materials

Subsection 3.4.1: Crosslinking Pdu BMCs

Crosslinkers were prepared as smaller single-use aliquots to extend their utility and cut down on costs, like the protocol described in Chapter 2.4.1. BMCs are diluted to 1.0 mg/mL in labeling buffer and equilibrated to room temperature. Regardless of crosslinker used, all were equilibrated to room temperature then added immediately such that the final concentration of DMSO never exceeds 2%. Crosslinking reactions proceeded for 45 minutes at room temperature with gentle rocking protected from light. Quenching was performed by adding Tris-HCl pH 8.0 to a final concentration of 5 mM.

Subsection 3.4.2: Mass Spectrometry of Crosslinked Pdu BMCs

Crosslinked samples were acetone precipitated and resuspended in 5% SDS prior to reduction in 200 mM DTT for 1 hour at 57°C. Samples were then alkylated with 500 mM iodoacetamide for 45 minutes in the dark at RT and quenched by addition of 200 mM DTT. 50 µg of sample were then cleaned and digested in S-Trap™ Micro Spin Columns (ProtoFi) according to manufacturer's protocols. In brief, samples were acidified with 5 µl of 12% phosphoric acid and proteins precipitated by six times

dilution with S-Trap buffer (100 mM TEAB in 90% MeOH, pH = 7.1) and captured on porous micro S-Trap spin column beads. S-Trap columns were washed three times with 150 ul of S-Trap buffer followed by digestion in 20 ul of 50 mM ammonium bicarbonate solution containing 1 ug of trypsin (Promega). Digestions were completed in 1 hour at 47˚C and peptides were eluted, dried, and resuspended in 0.5% acetic acid. Digested peptides were eluted in 40% acetonitrile, 0.5% acetic acid, followed by 80% acetonitrile, 0.5% acetic acid, dried, and resuspended in 0.5% acetic acid and stored at -80°C until analysis.

For DSBU and BS(PEG)5 samples, approximately 7 µg of the sample was analyzed. Samples were separated online iusingn an EASY-nLC 1200 (Thermo Fisher Scientific) and with a flow rate of 200 nL/min and a gradient of eluted in 2-80% acetonitrile in, 0.5% acetic acid over 220 minutes into an Orbitrap Fusion Eclipse spectrometer (Thermo Fisher Scientific). Full mass spectra were acquired with a resolution of 60,000 (at m/z 200), AGC target of 4e5, max. ion time of 50 ms, and scan range of 400-2000 m/z. MS2 spectra were collected with a cycle time of 3 seconds in the ion trap using rapid scan with an with a resolution of 30,000, AGC target of 5e4, max. ion time of 70 ms, one microscan, 1.6 m/z isolation window and, HCD using stepped 15, 30, and 45% normalized collision energy. and CID of 25%. A fragment mass filter for 85.0527 and 111.032 to select crosslinked peptides for MS3. MS3 spectra were collected with IT set to rapid, AGC of 2x104 max. ion time of 70 ms, one microscan, 2.5 m/z isolation window, and HCD of 27%. MS/MS spectra were analyzed using Sequest and XlinkX within Proteome Discoverer 2.53. A mass tolerance of 10 Da was set for MS1 and 0.5 Da for MS2. Crosslinks were searched using non-cleavable

acquisition strategy, 1.5 signal to noise ratio, FDR of 0.01, and variable crosslink modification on lysine of +138.068 for DSS and BS3 and +302.137 for BS(PEG)5. The following modifications were allowed: fixed: carbamidomethyl on cysteines (+57.021); variable: DSS/BS3 Tris on Lysine (+259.142), DSS/BS3 hydrolyzed on lysine (+156.079), BS(PEG)5 Tris on lysine (+424.218), BS(PEG)5 hydrolyzed on lysine (+320.147), oxidation on methionine (+15.995) and deamidated on asparagine (+0.984). While the cross-linker BS3 was used, it led to the identification of redundant cross-links reported within other datasets. The mass spectrometry raw files acquired at the Proteomics Laboratory at NYU are accessible at https://massive.ucsd.edu under accession MassIVE MSV000088460.

DSSO treated samples were run on an Orbitrap Eclipse mass spectrometer equipped with a Thermo Ultimate 3500 UPLC and FAIMS unit operating under 3 different CVs (-50, -60, and -75). Chromatography was performed using mass spectrometry grade reagents (H_2O and 0.1% FA for Buffer A and 95% ACN with 5% H_2O 0.1% FA for buffer B). A 75 cm Thermo Fisher Easy-Spray column was utilized for separation. A 200 minute gradient was utilized for separation at a 0.3 uL/min flow rate (0 min 5% B, 10 min 10% B, 150 min 70% B, 158 min 80% B, 161 min 90% B, 171 min 90% B, 172min 5% B, 200 min 5% B). MS1 was collected in the orbitrap at a resolution of 60,000 scanning from 375-1500 m/z with standard AGC target and MIPS active. An intensity threshold of 2.04e4 was utilized along with charge state inclusion between 2 and 8. Dynamic exclusion was held to 45 min with a mass tolerance of 20 ppm. For stepped collision energy MS2 acquisition, HCD assisted collision energies of 25% and 30% were used. MS2 scan were collected in the orbitrap at a resolution of

30,000. A normalized AGC target of 200% was used with a maximum injection time of 70 ms. For the MS2/MS3 acquisition, a fixed CID collision energy of 25% was used and MS2 scans were collected in the orbitrap at 30,000 resolution at standard AGC target and 70 ms max injection time. MIPS was then used and a targeted mass difference of 31.9721 m/z was used for the DSSO crosslinker. A charge state filter of 2-6 was utilized followed by precursor ion exclusion with a 50 ppm mass width. 2 MS3 scans were performed in the Ion Trap operating under Rapid scan mode using a fixed CID energy of 35%. An AGC target of 200% were used with a maximum injection time of 120 ms. For the MS2/MS2/MS3 acquisition, a fixed CID collision energy of 25% was used and MS2 scan were collected in the orbitrap at a resolution of 50,000. MIPS was then used and a targeted mass difference of 31.9721 m/z was used for the DSSO crosslinker. A secondary MS2 scan was collected post targeted mass identification in the orbitrap at a resolution of 30,000 utilizing ETD fragmentation. A defined mass range of 150-2000 m/z was used with a maximum injection time of 100 ms. For the MS3 scan, charge state filtering was performed post targeted mass identification to include charge states between 2 and 6 followed by precursor ion exclusion with a 50 ppm mass width. 2 scans were obtained in the ion trap operating under Rapid scan rate with fixed CID collision energy of 35%. An AGC target of 200% was used with a maximum injection time of 120 ms. All data was processed in Thermo Proteome Discoverer 2.5 using the Xlinkx node for crosslink identification. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Deutsch *et al.*, 2020) via the PRIDE (Perez-Riverol *et al.*, 2019) partner

repository with the dataset identifier PXD029675. Data reported here all have an Xlinkx score greater than 30.

Subsection 3.4.3: Purification of Pdu BMCs Containing GFP^{P1-18}

Pdu BMCs containing GFP^{P1-18} were sourced from the same strain also containing EYK054 (Jakobson, 2015). Growth and purification proceeded as in Chapter 2.4.2, except growth media volume was doubled and also contained 0.005% L-arabinose to induce GFP^{P1-18} expression. EYK054 was a gift from Dr. Danielle Tullman-Ercek (Addgene plasmid # 69512).

Subsection 3.4.4: Microscopy of E. coli expressing GFP-Containing Pdu BMCs

For in vivo imaging of BMCs, R995 + PduST co-transformed with EYK054 were grown overnight then diluted 1:100 in fresh 2xYT supplemented with 0.5% 1,2-propanediol and 0.005% (w/v) L-arabinose. Cultures were grown for 5.5 hours at 37°C before being pelleted and resuspended in 1x PBS. Samples were then blotted onto an agarose pad and imaged on a Zeiss Axio Observer Z1.

Chapter 4: Structure and Organization of the Pdu BMC Shell

Section 4.1: Introduction

Subsection 4.1.1: The Native Topology of BMC Shells is Unclear

Efforts to delineate the structure of BMC shells have greatly expanded since carboxysomes were first confirmed to have an icosahedral geometry (Schmid *et al.*, 2006). Some major breakthroughs included the high-resolution X-ray structures of BMC-H proteins (Kerfeld *et al.*, 2005; Tsai *et al.*, 2007), which tessellate to create the facets (Figure 1.2d), and BMC-P types, which cap the vertices (Tanaka *et al.*, 2008). These latter structures of BMC-P shell proteins allowed researchers to fully model and estimate a triangulation of T = 75 for carboxysomes (Tanaka *et al.*, 2008). However, these early models could not differentiate between a concave-out or convex-out orientation for the protein shell. Such aspects are critical for understanding the interactions that are pivotal towards cargo encapsulation. It would not be until 2017 when researchers solved the structure of the 40 nm diameter synthetic Hoch BMC (Figure 4.1a (Sutter *et al.*, 2017)) which revealed that the concave face of all shell components orient outwards. Later high-resolution structures of other small, synthetic BMCs confirmed this orientation (Figure 4.1 (Sutter *et al.*, 2019; Kalnins *et al.*, 2020; Tan *et al.*, 2021).

This story, however, is complicated by research completed *in vivo* on naturally occurring BMCs. These studies directly tied shell proteins in interacting with cargo EPs in the Pdu BMC system (Fan *et al.*, 2012); specifically with a C-terminal tail on the

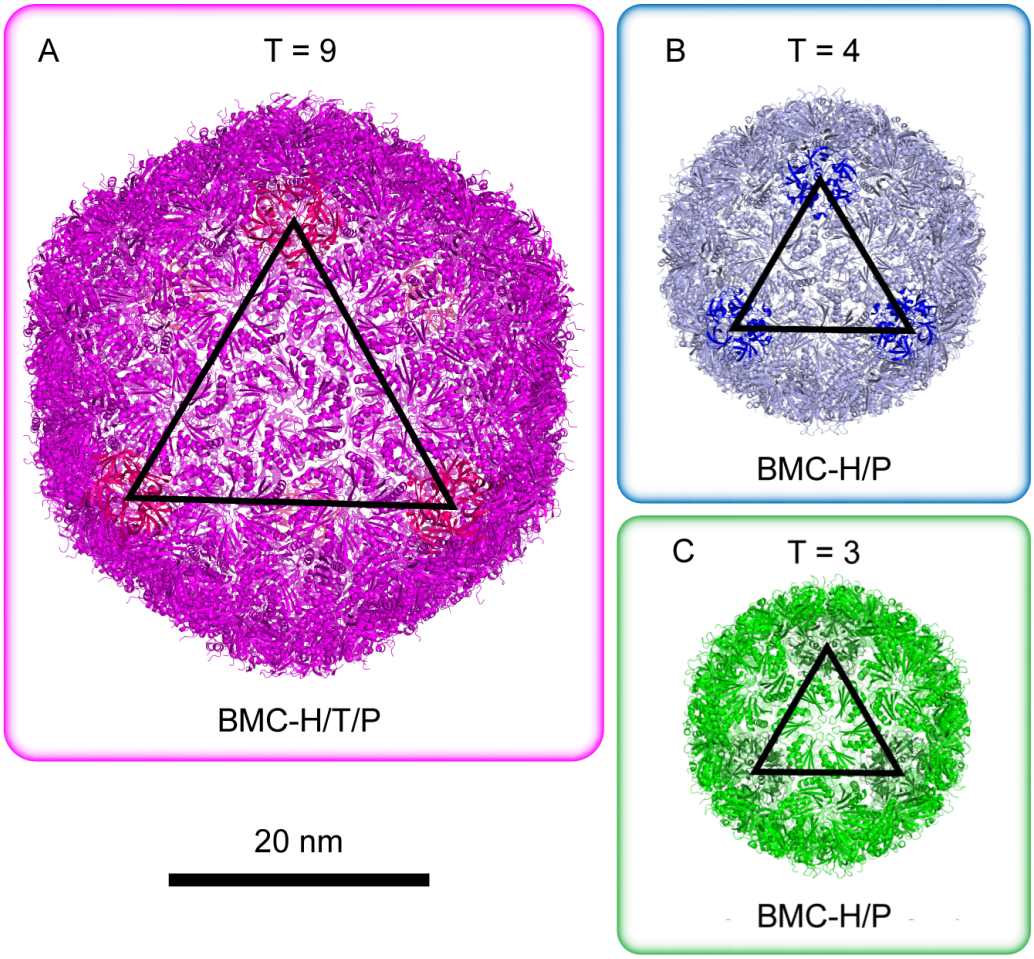


Figure 4.1: The construction of known, empty BMC shells. (A) The Hoch BMC (5V74, Sutter *et al.*, 2017), a proteobacterial BMC of unknown function, was the first structure solved. Other synthetic BMCs have been solved like simplified α-carboxysomes (Tan *et al.*, 2021) with different triangulation numbers (7CKC, B, and 7CKB, C) within the same BMC preparation.

BMC-H protein PduA, which lays on the concave face (Crowley *et al.*, 2010). Since the presumption is that cargo would be on the interior, then the concave-facing tail of PduA would likewise face the interior. This was later supported by docking simulations of all Pdu EPs to the concave surface of the PduA hexamer (Jorda *et al.*, 2015). These results question whether naturally occurring BMCs organize differently than their synthetic, empty counterparts. Clearing this discrepancy will be crucial for new BMC bioengineering efforts.

Subsection 4.1.2: The Assembly Principles of Large BMC Shells is Unclear

The assembly principles of BMC shells have been identified thanks to breakthroughs in cryo-EM imaging of various small, empty shells. These systems all reveal that BMC shells are built by facets which are extended sheets of BMC-H and/or BMC-T type proteins. The edge-edge interactions that guide assembly owe to shape complementarity and by the pairing of antiparallel lysine residues that form a specific hydrogen bonding pattern to dictate orientations (Figure 1.2c). These principles presumably scale to full-size BMCs that contain cargo.

The above inferences all come from work completed on synthetic BMCs that benefited from being small and composed of only a few protein components. What is not as clear, however, is how larger, complex BMC shells form. For example, the Pdu BMC at the heart of this dissertation contains 8 shell proteins; namely PduABB'JKNTU (PduB' is a truncated translation product of PduB). A minimal set of 6 proteins, PduABJKNU, has been noted to form complete empty shells (Parsons *et al.*, 2010) which is in stark contrast to some synthetic systems which require as few as 2-3 different shell proteins. Further, the BMC-H proteins PduA and PduJ are nearly identical in sequence and have been noted to be largely redundant (Kennedy *et al.*, 2021; Chowdhury *et al.*, 2016), although PduJ is >3-fold more abundant (Yang *et al.*, 2020).

Why some BMCs shells require more components and encode for apparently redundant homologs is not well understood. Synthetic BMC shells can even exhibit a mix of triangulation values within the same preparation (Figure 4.1b, c (Sutter *et al.*, 2019a)). Clearly, much more needs to be learned about these systems. This chapter

reports crosslinking and footprinting results among the protein shell of the Pdu BMC. These results give acute insight into the construction and sidedness of the shell of this large entity.

Section 4.2: Results and Discussion

Subsection 4.2.1: The Convex Surface of the Shell Faces Luminally

It is critically important to understand which face of the protein shell interacts with cargo in a native context. Knowledge in this regard would give rational engineering efforts a starting place to design new interaction surfaces. We scanned the crosslinking data for any connections between shell and cargo proteins. Overall, 13 shell-cargo crosslinks were identified to the shell proteins PduABJ (Figure 4.2a). These crosslinks were subdivided on the basis of which surface these proteins the crosslinks occurred on, based on crystallographic data (Figure 4.2b). When subdivided this way, 11 of these crosslinks occurred on the convex face while only one was to the concave surface. The last crosslink, to PduM, was to the disordered N-terminus of PduB which cannot be placed as either concave or convex laying based on crystallographic data (Figure 4.2b, gray circle) although crosslinks explained in Chapter 4.2.2 do place it on the convex face. PduM has been suggested to be a shell protein although it shares no sequence homology to BMC folds (Sinha *et al.*, 2012). The exact function of PduM remains enigmatic.

The BMC-H proteins PduA and PduJ were both only found to crosslink to the diol dehydratase medium subunit PduD on their convex face (Figure 4.2b, red circles).

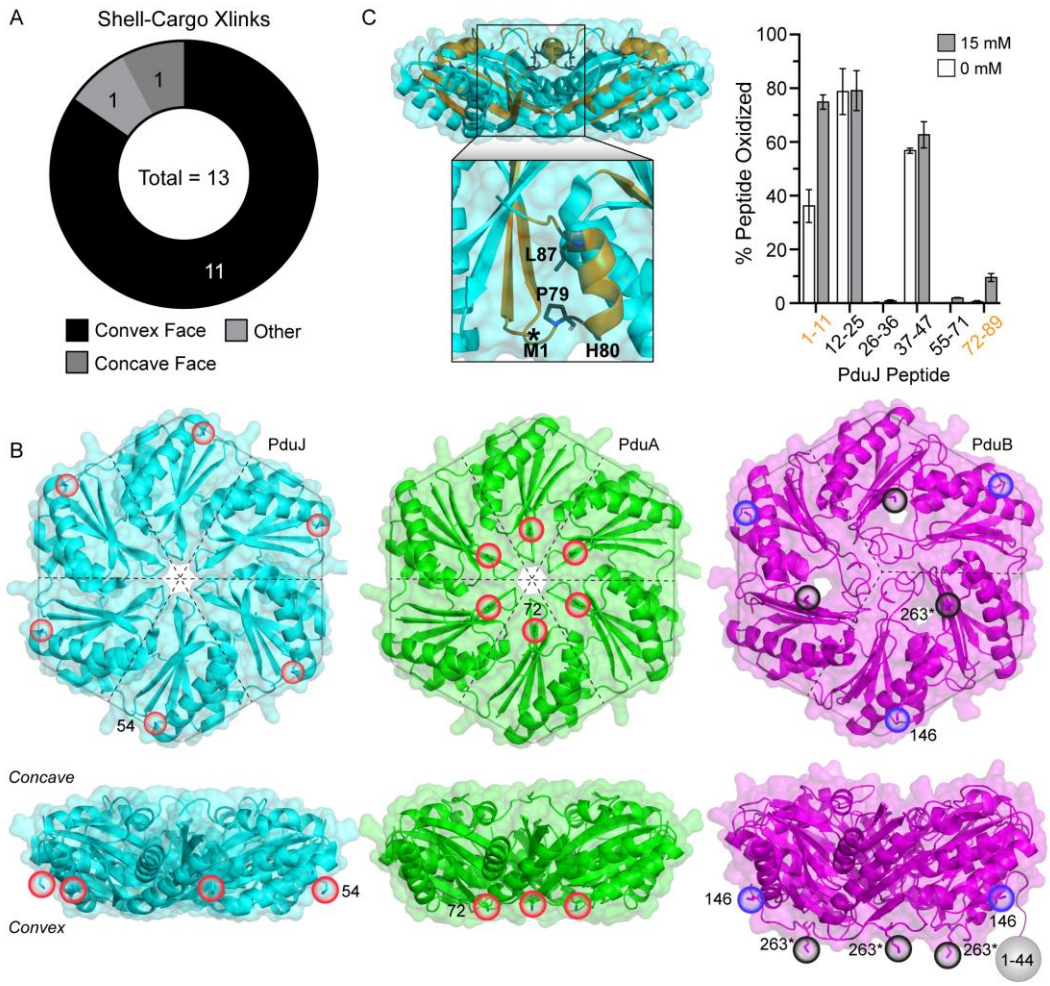


Figure 4.2: The convex face of the protein shell faces luminally. (A) An account of which face of shell proteins that enzymatic cargo crosslinked to. (B) Cargo crosslinking sites were mapped onto the shell proteins PduJ (cyan, 5D6V), PduA (green, 3NGK), and PduB (magenta, homology model of 4FAY). Residues are circled, with circle colors representing what cargo were linked. Red, PduD; blue, PduE; black, PduDEL; gray, PduM. Asterisks denote PduB^{263} which crossed directly to encapsulation peptides of PduE and PduL. (C) Pdu BMCs were subject to hydroxyl-radical footprinting to independently establish the sidedness of shell proteins. PduJ (shown) is used as a model due to excellent sequence coverage. Orange peptides had the largest response to oxidant. Black residues, on the concave face, carried most of the oxidation change. Asterisks denotes the N-terminus which is not in the structure of PduJ.

Meanwhile, PduB formed crosslinks to PduDEL also along its convex face. Specifically, PduB^{146} crosslinked to PduE which PduB^{263} crosslinked to all PduDEL including to each of their EPs (Figure 3.4b, Figure 4.2b, asterisks). These results do support the notion that EP containing cargo will interact with the protein shell, although to a far lesser extent than to other cargo as noted in Chapter 3. These data also suggest

that PduB may be a hub for shell-cargo interactions based on the diversity of crosslinks captured to PduB. This connection has been noted previously in a study that suggested that the disordered N-terminus of PduB faces inwards and interacts with cargo, effectively acting as an anchor point (Lehman *et al.*, 2017). The authors found that deletion of the N-terminus of PduB seriously hampers cargo packaging but not shell formation (Lehman *et al.*, 2017). This understanding is expanded on in Chapter 6.

Enumerating the locations of shell-cargo crosslinks along specific faces of shell proteins suggests that the convex faces lay luminally. This agrees with high-resolution structures like that of the Hoch BMC (Figure 4.1a). Further, the sidedness data for PduABJ all agree with one another; an expected result given the assembly principles that dictate matching orientations. However, these interpretations presume (albeit a reasonable presumption) that the cargo is located on the interior. We sought to apply hydroxyl-radical footprinting (McKenzie-Coe *et al.*, 2021) to further confirm the shell orientation in a crosslinking-independent manner (Figure 4.2c). In this experiment, we measured the oxidation of surface-exposed amino acids using hydroxyl radical generation via the Fenton reaction (Koppenol, 1993). The percent oxidation was measured using mass spectrometry. Peptides that had the most change (>10%) were deemed exposed (Figure 4.2c, orange). Using PduJ as a case study (due to excellent sequence coverage), our footprinting results showed that the concave-presenting residues on PduJ (M1, L79, P79, and H80; black residues, Figure 4.2c) exhibited the most oxidation change and therefore are oriented outwardly. The data presented here unambiguously attribute the convex face of the protein shell with laying luminally.

Subsection 4.2.2: The Termini of PduBK Exist on Opposite Faces of the Shell

A total of 56 crosslinks were identified strictly between shell proteins (Figure 4.3a). These crosslinks were classified based on which face of the protein shell that were crosslinked together. For example, crosslinks between proteins that happened on both their concave faces are deemed concave-concave. 20 concave-concave crosslinks, 9 convex-convex, 5 concave-convex, and 22 crosslinks classified as other were identified (Figure 4.3a). Crosslinks classified as "other" occurred to the extensions on the PduB trimer and PduK hexamer whose orientations are unknown. The extension on PduB is located on the N-terminus and has been implicated in anchoring the shell to the inner cargo (Lehman *et al.*, 2017). The presence of this extension differentiates PduB from the shorter PduB'. Meanwhile, the extension on PduK is on its C-terminus and may contain an iron-sulfur cluster binding motif (Crowley *et al.*, 2010).

Crosslinks were captured between the terminal extensions of PduB and PduK to specific faces of neighboring shell proteins. Tabulating these crosslinks enabled us to assign a spatial location to these termini, either being luminal or cytosolic. The N-terminal extension of PduB formed 6 crosslinks to the convex faces of other shell proteins and only one crosslink to a concave surface (Figure 4.3b). In contrast, the C-terminal extension on PduK was found in 10 crosslinks to concave surfaces and only once to a convex surface (Figure 4.3b). Consequently, we assign the N-terminus of PduB with laying luminal to interact with cargo while the C-terminus of PduK presents outwardly to interact with cytosolic factors. This interpretation for PduB agrees with previous studies that implicated the N-terminus with interacting with cargo proteins (Lehman *et al.*, 2017). PduK, meanwhile, may assist in coordinating electron transport

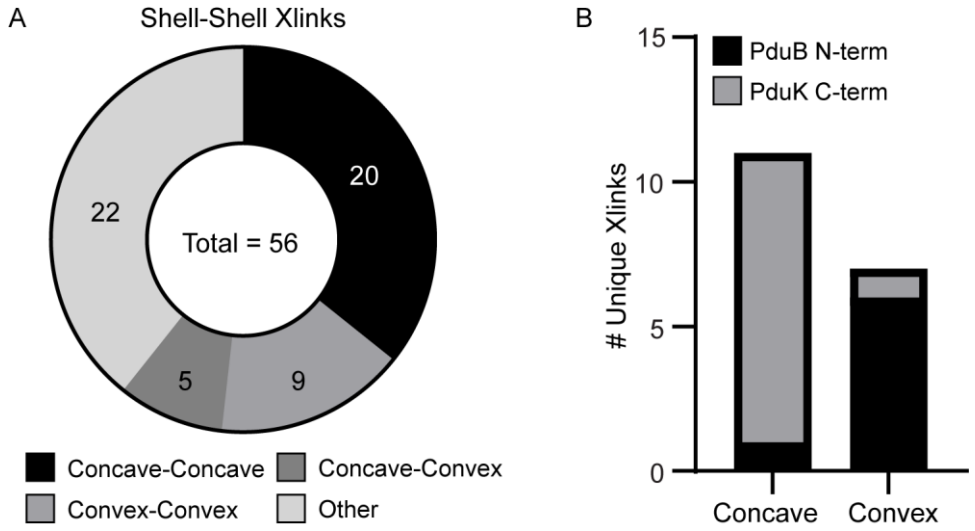


Figure 4.3: The relative locations of shell protein terminal extensions. (A) An account of which faces of shell proteins crosslinked to each other. "Other" crosslinks are to the N-terminus of PduB or C-terminus of PduK whose structures are not known. (B) Tabulating which faces of other proteins that the termini of PduB and PduK crosslink to. PduB's N-terminus crosslinks mostly to the convex face of other shell proteins, and PduK's C-terminus crosslinks along the concave face.

to the inner lumen from the cytosol. This speculation is tentatively supported by crosslinks captured between PduK and the trimer PduT, which itself is known to coordinate an iron-sulfur cluster at its pore (Crowley *et al.*, 2010; Chowdhury *et al.*, 2019). Together, these two shell proteins may facilitate electron transport through the shell layer.

Subsection 4.2.3: The C-Termini of Major BMC-H are Dynamic/Flexible

Numerous crosslinks were established between the major BMC-H proteins PduA and PduJ. The sequences of PduA and PduJ are over 80% identical and 93% identical (Figure 4.4a) making some regard them as potentially degenerate (Chowdhury *et al.*, 2016; Kennedy *et al.*, 2021; Yang *et al.*, 2020). These two proteins, in terms of raw subunit abundance, together make up slightly over 60% of the Pdu BMC (44.7% PduJ monomers, 15.8% PduA monomers (Yang *et al.*, 2020)). Understanding how

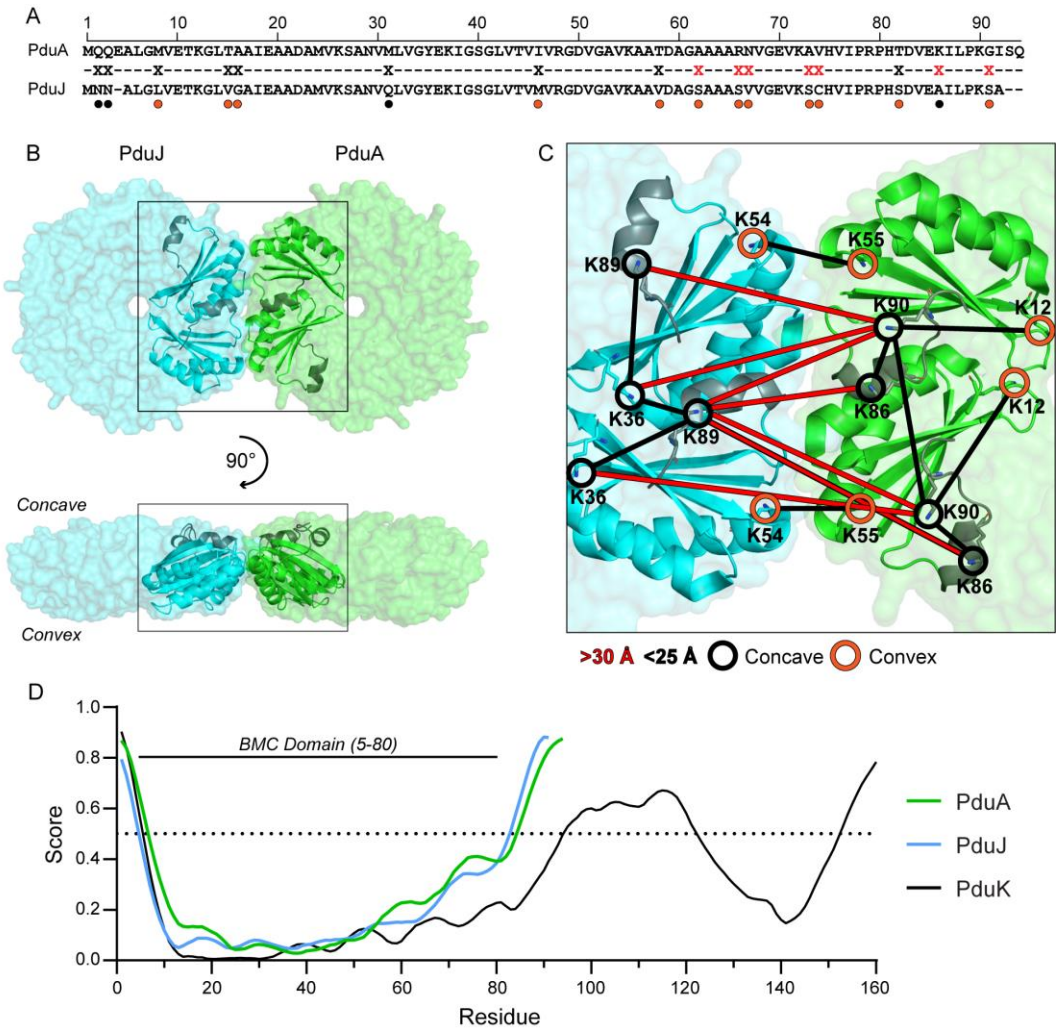


Figure 4.4: The C-terminus of PduA and PduJ are disordered. (A) An alignment of PduA and PduJ. Black X's denote conserved divergence, red X's denote major divergence. Black dots represent concave-facing residues and orange represent those along the convex surface. (B) The model of PduA (3NGK, green) and PduJ (5D6V, cyan) used to evaluate the crosslinks (C) between them. The crosslinks are shown as lines connecting circled residues. (D) IUPred3 disorder analysis of the shell proteins PduA, PduJ, and PduK (control). Scores higher than 0.5 are interpreted as being disordered. The C-terminus of PduA and PduJ are more disordered than the equivalent region of PduK, indicating they are disordered.

these two proteins interact in a native context is therefore crucial to understanding the assembly of the BMC shell.

The crosslinks between PduA and PduJ were carefully evaluated for any insight. This was done by taking the separate high-resolution crystal structures of PduA and PduJ and aligning them edge-to-edge as they would exist in a facet (Figure 4.4b).

Crosslinks were then manually mapped onto the aligned structures with a distance cutoff of 30 Å deemed acceptable for the crosslinkers used (~12 Å spacer, 7 Å per conjugated lysine). Nearly all crosslinks agreed on sidedness (*i.e.* the linking of concave or convex pairs of lysines). This distance cutoff allows for crosslinking of PduA[55] and PduJ[54] directly along the interaction interface on their convex sides (Figure 4.4c). Other intramolecular crosslinks among PduA and PduJ are agreeable with this cutoff (Figure 4.4c, black lines).

However, most intermolecular crosslinks between PduA and PduJ are spaced too far for the crosslinkers used (Figure 4.4c, red lines). All these crosslinks were to the C-terminal helical tails of PduA and PduJ (Figure 4.4c, darker helices). One such crosslink had PduA[90] on its C-terminal tail crosslink to PduJ[36] near its central pore. PduA[90] was also captured in crosslinks to PduA[12], which is located near the pore of PduA on the opposite face of the hexamer (Figure 4.4c). It is unlikely that the model used to evaluate the crosslinks (pairing PduA and PduJ side-to-side) is skewing results given what is known on how BMC-H proteins assemble in BMC facets. Rather, we speculate that the C-terminus helix of these two proteins is actually disordered and flexible in solution allowing the contacts that are observed.

This speculation is warranted. Recent molecular simulations of carboxysome hexamer CcmK2, a homolog of both PduA and PduJ, suggest that its C-terminal tail is highly flexible (Faulkner *et al.*, 2020). Indeed, sequence analysis of these PduA and PduJ with IUPred3 (Erdös *et al.*, 2021), an online disorder prediction server, also supports this (Figure 4.4d). Here, each residue is assigned a disorder score with scores over 0.5 being considered disordered by default. The IUPred analysis shows that PduA

and PduJ are anticipated to be well ordered within the central BMC domain but are much less ordered within their C-termini (Figure 4.4d). PduK, a BMC-H protein anticipated to have an ordered C-terminus, serves as a control.

These data together allow us to hypothesize that the C-termini of the major Pdu BMC hexamers, PduA and PduJ, are flexible/disordered in solution. The distance of the contacts made to their termini may indicate that these termini may be an unappreciated mode of shell-shell contact that is important for assembly. This interpretation may explain why previous data implicated the C-terminus of PduA with cargo interactions despite laying on the concave face (Fan *et al.*, 2012). The author's disruptions of the C-terminus of PduA may have perturbed its ability to assemble into the BMC shell. Further mechanistic investigations into the terminal extensions of BMC shell proteins, their ubiquity, and their bioengineering potential are warranted.

Subsection 4.2.4: The BMC-T Protein PduB Hosts Edge-Specific Contacts

PduB is another major component of the Pdu BMC shell. This trimer can be translated into full-length PduB or as a truncation deemed PduB' which lacks the N-terminal 37 amino acids. This N-terminal extension has already been implicated by others with interacting with cargo proteins (Lehman *et al.*, 2017) and confirmed to be facing the interior by data provided here in Chapter 4.2.2. The data here also shows the PduB makes the highest number of cargo crosslinks and is linked to a greater diversity of cargo than other shell components (Figure 4.2b). This data shows that PduB may indeed act as an anchoring hub between the protein shell and inner cargo. The fact that PduB may be the key to understanding the shell-cargo connection underscores the need to understand how PduB assembles in the protein shell.

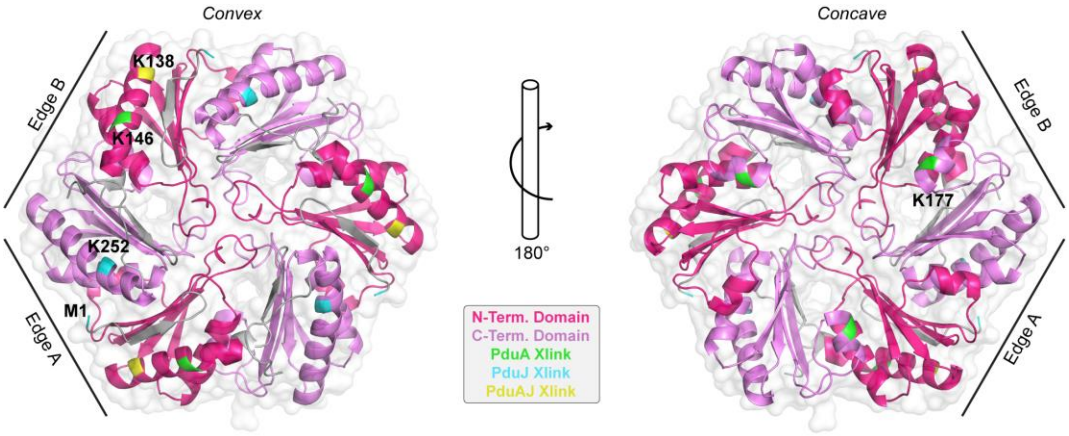


Figure 4.5: PduB accommodates specific interactions with PduA and PduJ. PduB, shown as a homology model of 4FAY (Pang *et al.*, 2012) has two different edges; A and B. Specific residues are colored according to whether it crosslinked to PduA, PduJ, or both.

Among the shell proteins, PduB was only identified in crosslinks to the hexamers PduA and PduJ (Figure 3.3a). A single crosslink was also identified between the N-terminus of PduB and PduM. PduM is a protein with an unknown structure and no sequence homology to other proteins. It has been tied to BMC assembly and is in some instances considered a shell protein (Sinha *et al.*, 2012). Due to the dearth of information about PduM, we instead focused on the crosslinks between PduB and the well-characterized shell proteins PduA and PduJ.

Trimeric shell proteins, like PduB, have been speculated to have evolved through gene duplication of the BMC domain. This duplication allows for the protein to have two distinct edges (Figure 4.5, edges A and B) and, therefore, the potential to accommodate two different sets of interactions. Crosslinks between PduB and PduAJ were evaluated with this in mind and mapped onto a homology model of *S. enterica* PduB built from a high-resolution structure of *L. reuteri* PduB (Figure 4.5 (Pang *et al.*, 2012)). Mapping the crosslinks onto the model reveals that PduA and PduJ form specific crosslinks to specific edges of the PduB trimer. For instance, PduA crosslinks to PduB^{146} and PduB^{177} both along one specific edge of the trimer (Figure 4.5, green).

In contrast, PduJ crosslinks to PduB^{1} and PduB^{252} along the other edge (Figure 4.5, cyan). This is surprising since PduB^{146} and PduB^{252} occupy equivalent positions within the pseudo-hexameric structure, yet PduA only links to the former and PduJ to the latter, emphasizing the specificity. Both PduA and PduJ are captured in crosslinks to PduB^{138} (Figure 4.5, yellow) likely due to its position at a vertex making it equidistant and accessible to both proteins.

Trimeric shell proteins making different contacts has been observed before. For instance, the Hoch BMC (Figure 4.1a) has a trimer situated at the center of each facet that is surrounded by equivalent hexamers. While the hexamers are the same, unlike in the Pdu BMC, the trimer can accommodate different angles along each edge depending on the relative positioning of the pentameric cap (Sutter *et al.*, 2017). The data reported here shows that PduB may have to also accommodate two different sets of interactions and coordinate with PduA and PduJ separately. This is highlighted by the absence of crosslinks between PduB to other shell proteins. Further, all the crosslinks captured between PduB peptides can be most easily explained as originating from intramolecular contacts as opposed to intermolecular to another PduB trimer. These data are consistent with the notion that BMC-T subunits may be interspaced by BMC-H within the shell facets, representing a fundamental repeating unit of the Pdu BMC shell (Figure 4.7a).

The structural basis of the specific edge interactions we observed can be not described from our data. The differences present between PduA and PduJ (Figure 4.4a, X residues) lay predominantly along the convex surface (orange circles) as opposed to the edges that interact with other shell proteins. These sequence differences, in conjunction with specific interactions with PduB, may hint at specific roles for these

supposedly degenerate proteins. It may be the case that PduB coordinates PduA and PduJ to accommodate different angles of interactions, such as for PduB homolog in the Hoch BMC, in relation to other shell components nearby. The sum of these interactions has been understood to be the basis of curvature in the Hoch BMC model system but need to be expanded beyond our analysis for the Pdu BMC. Understanding the differences may help to rationally design curvature and therefore the size of BMC particles.

Subsection 4.2.5: Modeling the Facets of the Pdu BMC

Piecing together the puzzle of the Pdu BMC shell requires (1) knowledge of the individual protein interactions and (2) an estimation for the size of the facets. This latter aspect gives the appropriate frame size with which to nestle the individual shell protein tiles into and places constraints on the combinatorial possibilities. Fortunately, several high-resolution synthetic BMCs structures have been solved. These structures give insight into not just their size, but also their triangulation number.

Triangulation numbers typically refer to the number of subdivisions within the facet of a viral capsid (Prasad and Schmid, 2012) but this definition can apply to any icosahedral structure. The fact that the triangulation (T) is proportional to diameter (d) of a particle squared ($T \propto d^2$) means that if one knows the diameter of an icosahedron, then T can be solved for if the proportionality is known. Estimating the triangulation number for the Pdu BMC is key to attempting to piece together the facet substructure.

The relation between triangulation and diameter in all BMC is likely imposed by the basic structural components, the individual shell proteins. Since the shell proteins are so well conserved in structure, this means that one should be able to fit the

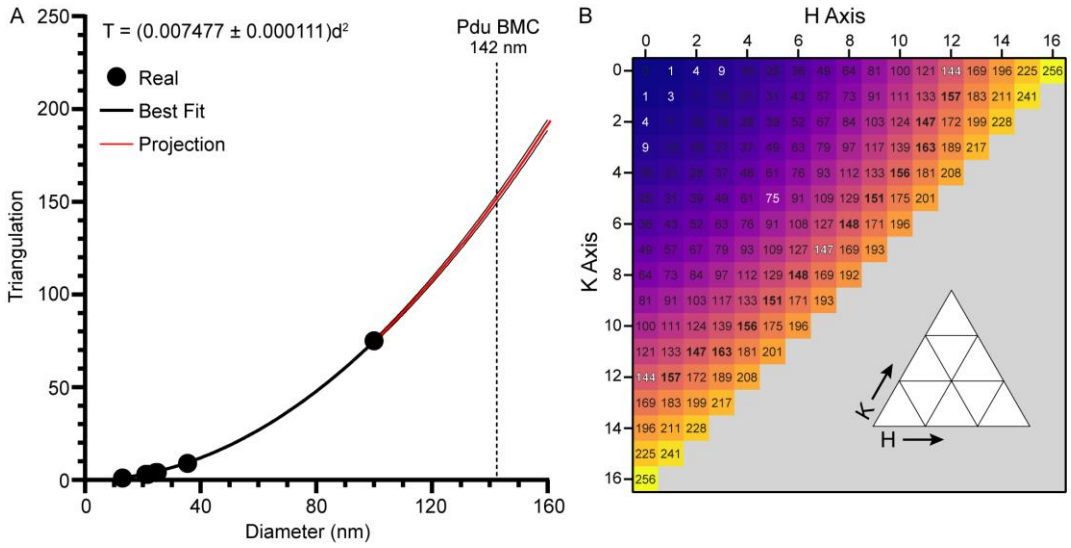
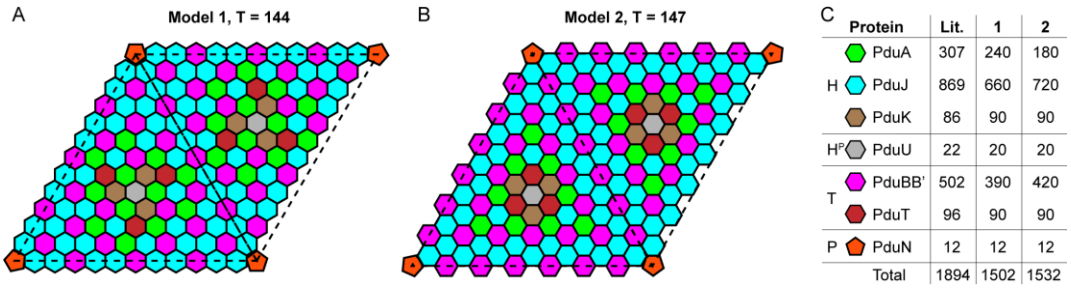


Figure 4.6: Estimating the triangulation values of the Pdu BMC. (A) Known BMC structures were plotted according to their triangulation value and diameter. An equation was fit (black line) and projected (red line) to the diameter of the Pdu BMC. The thin black lines about the projection represent the upper and lower bounds calculated by the error in the proportionality constant. (B) Heat map of all possible triangulation values. White numbers represent known BMC triangulations. Bolded values are those considered within the appropriate range given a diameter of 142 nm. White, bolded values are deemed the most ideal values. Inset shows how H and K values plot onto an imaginary icosahedron facet.

relation between triangulation and diameter for known BMC and solve for the proportionality constant. When BMC structures are plotted in this manner (including the carboxysome widely regarded as having a T = 75, d = 100 nm (Tanaka *et al.*, 2008) a proportionality constant of 0.007477 ± 0.000111 is given (Figure 4.6a). Unfortunately, the Pdu BMC does not exist on this curve given its diameter of 142 nm (Figure 2.6b). Extrapolating out (Figure 4.6a, red line) gives the Pdu BMC a T = 150. Given the errors in the proportionality constant (Figure 4.6a, thin black lines) associated with this analysis, T values between 138 and 163 were considered as options for the Pdu BMC.

While T is proportional to d^2, it is also equal to $H^2 + HK + K^2$ where H and K are positive integers denoting the distance from the origin in the H and K axis - much like X and Y in cartesian space (Figure 4.6b, inset). Due to this, not every integer value



	Protein	Lit.	1	2
H	PduA	307	240	180
	PduJ	869	660	720
	PduK	86	90	90
H^P	PduU	22	20	20
T	PduBB'	502	390	420
	PduT	96	90	90
P	PduN	12	12	12
	Total	1894	1502	1532

Figure 4.7: Models for the Pdu BMC shell. (A) The basic unit of the Pdu BMC facet is BMC-T proteins surrounded by BMC-H proteins, resulting in an interspaced arrangement of BMC-T types. (B) Shell protein abundances presented in the literature and models for a T = 144 facet (C) and T = 147 facet (Model 1 and 2, respectively).

between the 138 to 163 estimation is a possibility. All possible triangulation values for H and K up to 16 were assessed (Figure 4.6b) and possible triangulations were determined (bolded). Known BMC triangulations lie principally along the single-component or diagonal axis. For this reason, the triangulation values of 144 (either H or K = 12) and 147 (H = 7, K = 7) (Figure 4.6b, white bolded) were deemed most likely to represent the Pdu BMC shell.

The dimensional constraints offered by the estimated triangulation allows shell protein "pieces" to be fit and the Pdu BMC facets modeled (Figure 4.7). These models are informed by both the cross-linking data herein and by stoichiometric data reported by Yang et al (Figure 4.7b). Both models (for T = 144 and 147) have commonalities. Briefly, they (1) both have trimers interspaced by hexamers (Figure 4.7a), (2) a central region dictated by PduU, and (3) facet edges predominantly composed of PduJ and PduB (Figure 4.7c, d).

While no crosslinks were observed to PduU, a circularly permuted BMC-H protein, its copy number indicates that there exists one copy per facet. Presumably, PduU exists at the center of each facet to coordinate specific mid-facet interactions. These models also reflect the alternating arrangements of PduA and PduJ around PduB

captured in the crosslinking data (Figure 4.5) albeit not at every available position dictated by known stoichiometry (Figure 4.7b). Lastly, these models anticipate that PduT and PduK are connected. This was reflected in the crosslinking data (Figure 3.3b) but also suggested by the near 1:1 ratio of their abundances (Figure 4.7b). There connection is further reinforced by the cytosolic C-terminal extension of PduK containing an iron-sulfur cluster binding site (Crowley *et al.*, 2010) and PduT coordinating such a cluster at its central pore (Crowley *et al.*, 2010); the two may interact to facilitate electron transport across the shell.

Altogether, the models presented here are consistent with the literature except based on stoichiometry (Figure 4.7b). The protein abundances determined by Yang *et al.* (Yang *et al.*, 2020) are consistent with Pdu BMCs with T = 192/196 and a diameter of >160 nm. The authors themselves measure the Pdu BMCs in their study as having a 110 nm diameter - even smaller than the 142 nm we report (Figure 2.6b). Given that shell proteins are known to form an array of higher-order structures (Uddin *et al.*, 2018), it may be that these structures co-purify with BMCs and skew the results. Regardless, the models presented here are consistent with (1) the sizes we report, (2) the interactions we capture, and (3) scaled-down abundances of the shell components.

Section 4.3: Conclusions

Subsection 4.3.1: Shell-Cargo Links are Consistent with LLPS

Our data suggest that the trimeric protein PduB may act as a hub for shell-cargo interactions. This is consistent with prior studies that showed that perturbation of the N-terminus of PduB leads to aberrant cargo packaging (Lehman *et al.*, 2017). If true,

we speculate that the luminal facing N-terminus of PduB may act as a cognate EP to anchor/seed shell formation around the inner cargo. The distributive arrangement of BMC-T proteins that we proposed (Figure 4.7) and the need for two isoforms of PduB are consistent with this. Recent modeling suggested that increased interactions between shell and cargo can lead to premature cargo encapsulation and aberrant structures (Mohajerani *et al.*, 2021). In this context, the combination of PduB and PduB' interactions with the inner cargo might therefore be needed to permit proper condensate maturation prior to shell envelopment. Perhaps the protein shell can be viewed as a stabilizing factor that scaffolds on the inner core while introducing a more-or-less predestined curvature and size, although it cannot be determined from our data whether a concomitant or core-first model is best (Figure 1.3). Here, in this model, the Pdu BMC may form through biomolecular condensation but once assembled it exists actively as a shell-stabilized entity. We consider our data to be consistent with this overall model.

Subsection 4.3.2: The Organization of the Protein Shell

Our data was closely examined for any new insights. We confirm that native BMC shells have the same topology as synthetic, empty shells (concave out), thereby reaffirming the modularity of these systems. Our analysis also suggested that the Pdu BMC shell may employ a fundamental repeating unit, consisting of a cyclical arrangement of BMC-H proteins around BMC-T types (Figure 4.7a). Here, BMC-T proteins punctuate the facets and are spaced by BMC-H types. Our data specifically implicates PduA and PduJ in alternating positions around PduB (Figure 4.5). Overall, our shell model is like the structural model presented for the Hoch BMC (Sutter *et al.*, 2017), wherein the center of each facet is occupied by a BMC-T protein surrounded by

BMC-H. In contrast, our model differs from an alternate arrangement that was proposed from measurements of stoichiometric data (Yang *et al.*, 2020), which proposes that PduB occupy the edges of each face.

The models presented in Figure 4.7 are consistent with the sum of the literature data reported herein. Additionally, these models may explain the basis of heterogeneity observed in purified BMC samples. We predict that the central region of each facet is governed by PduU and enables sheet extension, thereby making this region relatively conserved in terms of protein components. The size of each facet, however, may be dictated by expansions beyond this central region as to (1) minimally conserve the central interactions while also (2) maintaining the separation of BMC-T proteins. Given this, Pdu BMC may have a minimal size of ~105 nm (T = 81) or ~100 nm (T = 75) based on whether Model 1 (Figure 4.7c) or Model 2 (Figure 4.7d), respectively, is considered. This speculation is supported by the literature, where Pdu BMCs below 100 nm in diameter are rarely reported (Kennedy *et al.*, 2020). Conversely, too much facet extension beyond the central region may be destabilizing, giving rise to size limitations.

Which triangulation model (Figure 4.7) is correct and whether all facets within the same metabolosome need to be the same is unknown. However, some synthetic BMCs have been observed with just H or both H and K triangulation components (Tan *et al.*, 2021) but why certain configurations are preferred is not understood. Understanding these aspects, including delineating the roles of shell proteins in complicated BMCs, are critical for BMC shell bioengineering. Future research in this

area will lead to controllable BMC shells with a range of sizes available for specific needs.

Section 4.4: Methods and Materials

Subsection 4.4.1: Mass spectrometry of crosslinked Pdu BMCs

Refer to chapter 3.4.2 for details.

Subsection 4.4.2: Hydroxyl-Radical Footprinting of Native Pdu BMCs

Pdu BMCs were purified as described in Chapter 2,4,2. Aliquots of 45 µL of 5 µM samples in 20 mM Tris-HCl pH 8.0 buffer were used for each reaction condition in triplicates. Samples were reacted with 15 mM, 7.5 mM, 3.25 mM, and 0 mM of iron (II) sulfate heptahydrate and hydrogen peroxide at 2-fold excess, respectively, for 2 minutes. The reaction was quenched with 50 µL of 0.1 ng/µL catalase, 68 µM thiourea, and 34 µM methioninamide. The samples were subsequently incubated with 50 µM Poros Beads for 2 hours at 4°C. The beads were then loaded onto C4 ZipTips, washed 3 times with 0.1% TFA, 3 times with 0.5% acetic acid and eluted with 40% acetonitrile followed by 80% acetonitrile in 0.5% acetic acid and subsequently concentrated in a SpeedVac concentrator. Samples were resuspended in 50 µL 100 mM ammonium bicarbonate, reduced with 20 mM dithiothreitol for 1 hour at 57°C, alkylated with 50 mM iodoacetamide for 45 min at room temperature in the dark, and digested with 1 µg of trypsin (Promega) overnight at room temperature. Next, the peptides were desalted with C18 ZipTips and resuspended in 0.5% acetic acid.

An aliquot of the peptide mixture was loaded onto an Evosep tip and separated on an EV-1106 Evosep column using the 15SPD Evosep method. MS1 scans were

acquired at 60,000 resolution (@ 400 m/z), 50 msec maximum injection time, AGC target of 4e5, and scan range of 400 to 2000 m/z. MS2 scans were acquired at 30,000 resolution (at 400 m/z), 200 msec maximum injection time, AGC target of 5e4, isolation window of 2 m/z, one microscan, and HCD using 30% normalized collision energy. Raw files were searched using Byonic (Protein Metrics) using the following settings: The MS2 spectra data was searched with a 10 ppm mass tolerance for MS1 and MS2, allowing for 2 missed cleavages. We allowed the following modifications: carbamidomethyl +57.021 on C, -17.0265 on Q, -18.010565 on E, oxidation +15.9949 on A, E, I, K, L, Q, R, V, C, D, F, H, M, P, W, Y, dioxidation +31.9898 on C, M, P, W, and Y, trioxidation +47.9847 on W, Y, P, and C, carbonylation +13.9793 on L, D, I, V, P, R, K, E, and Q, carbon monoxide loss -27.9949 on D and E, carbon dioxide loss -43.9898 on D and E, decarboxylation -30.0106 on D and E. Along with the documented oxidative histone modifications of -23.0160, -22.0320, and +4.978930 and cysteine modifications of -41.0265, -9.0358 and -25.0316 (Xu and Chance, 2007). The area under the curve for each observed peptide and modified isoform was calculated in a semi-automatic fashion using Byologic (ProteinMetrics) and the % oxidation for each peptide calculated. The mass spectrometry raw files acquired at the Proteomics Laboratory at NYU are accessible at https://massive.ucsd.edu under accession MassIVE MSV000088460.

Chapter 5: Pdu BMCs Exhibit Dynamic Qualities *in vitro*

Section 5.1: Introduction

Subsection 5.1.1: Purified Shell Proteins Assemble in Real Time

BMC shells are regarded as unyielding, unbending, and unpliable. These structures cannot be persuaded to budge. In short, they are static. As icosahedrons, BMCs are believed to consist of 12 vertices, 30 edges, and 20 facets. The facets are generally modeled as flat sheets (although they have some curvature) where individual shell protein tiles tessellate to extend the structure (Figure 1.2b). This mode of facet assembly was first speculated to be biologically relevant when the first crystal structures of the carboxysomes BMC-H proteins CcmK2 and CcmK4 were crystallized (Kerfeld *et al.*, 2005). The high-resolution structures of these proteins revealed that they form extended sheets within the crystal lattice. This is also the case for the BMC-H protein PduA (Crowley *et al.*, 2010).

While the crystal forms have afforded a wealth of insight into BMC assembly, they still represent just a static image of shell proteins. Newer developments have begun applying high-speed atomic force microscopy (AFM) to study BMC-H sheets in real-time and, importantly, in solution. These AFM studies have found that BMC-H protein sheets are, contrary to the static crystal structures, highly dynamic and assemble and disassemble in real time (Sutter *et al.*, 2016). The assembly occurs predominantly at the edges of the sheets where proteins are not as “locked in”. However, individual

hexamers can also dissociate from the middle of the sheets leaving behind holes which can nucleate further disassembly. The rates of subunit exchange and sheet size can be affected by mutations along the interaction edge, pH, salt concentration, and type of salt (Faulkner *et al.*, 2019). These observations, while applying strictly to purified, homogenous shell proteins, begin to set the stage for the possibility that some classes of BMC have a dynamic protein shell.

Subsection 5.1.2: Pdu BMCs Exhibit Unanticipated Permeability

The possibility that the Pdu BMC shell may be dynamic is tentatively supported by the results reported herein so far. To recap, the probes used in Chapters 2-4 (AF532, crosslinkers, *etc*.) are larger than the shell protein pores should be able to accommodate (Figure 2.1). In fact, these probes were initially selected for that very reason. Despite their large size compared to the pores, these probes were still able to access the inner lumen (Figure 2.4a) albeit at a slower rate than the outer protein shell (Figure 2.4b). This fact allowed us to gather crosslinking data on not just the outer shell, but also the inner cargo too.

These results suggest that the Pdu BMC shell exhibits a degree of flexibility that enables the probes to gain access to the interior space. However, the exact nature of this flexibility is not yet resolved. It may stem from "cracks" within the protein shell or there being an intentional equilibrium of shell proteins that leaves holes within the facets. We wanted to gain further insight into this curiosity *in vitro*.

Subsection 5.1.3: Methods to Image BMCs Have Stagnated

Studying the potential dynamics of the Pdu BMC requires in solution approaches that are severely lacking for studying BMC systems. More generally, synthetic deployment of BMC biotechnology hinges on a proper understanding of their morphology and in-solution behavior as well. One aspect that limits the ability to directly study BMCs is their size. Typically, BMCs are between 40-300 nm in diameter. These dimensions exist just at or below the Abbe diffraction limit for visible light. Due to this, transmission electron microscopy (TEM) has been the method of choice for decades as the wavelengths of the electrons can be made to be much shorter than those in visible light microscopy (Kumar and Sinha, 2021). Accordingly, TEM analysis benefits from high resolving power and relatively simple sample preparation but is hindered by dehydration and staining methods that can warp the native conformations of these samples (Kennedy *et al.*, 2020). Despite this, TEM has helped researchers understand BMC pH and temperature stability (Kim *et al.*, 2014), visualize non-native BMC-based nanotubes (Hagen *et al.*, 2018b; Noël *et al.*, 2015; Uddin *et al.*, 2018), as well as studying synthetic, empty BMC systems (Mayer *et al.*, 2016; Kalnins *et al.*, 2020; Jorda *et al.*, 2016). TEM, however, cannot attest to in-solution behavior and suffers from sample preparation artifacts in addition to being low throughput. Other optical techniques, such as dynamic light scattering (DLS), scanning electron microscopy (SEM), and cryo-TEM have been applied but their use has not been widely adopted and each has benefits and drawbacks (Kennedy *et al.*, 2020).

Newer methods of BMC analysis have been developed using AFM (Faulkner, 2017). AFM was previously mentioned to be used to study the facet formation of

purified shell proteins (Sutter *et al.*, 2016) but has also been leveraged to interrogate purified beta-carboxysomes (Faulkner *et al.*, 2017; Rodriguez-Ramos *et al.*, 2018). AFM benefits from high resolution, the ability to analyze particles in solution, and one can gain kinetic data, but it requires uncommon instrumentation and specialized training to operate.

Fluorescence confocal microscopy is a common optical technique that has yet to be widely adapted to study BMCs *in vitro* as it relies on visible light and is limited by the diffraction limit. However, new detector technology, such as the Airyscan detector (Huff, 2015), can allow for 1.7x higher resolution than traditional confocal microscopes. Confocal analysis of BMCs *in vitro* could benefit from observing in-solution properties while enabling detection of single BMC particles. We report a laser scanning confocal method that can be used to directly image Pdu BMCs. This method allows for single-particle detection and, due to using differently tagged BMCs, colocalization events and can therefore report on equilibrium properties of the Pdu BMC.

Section 5.2: Results and Discussion

Subsection 5.2.1: The Outer Shell Layer Influences Dye Permeability

We first sought to investigate whether the Pdu BMC can be modified to bias probe access even more for the exterior of the compartment. For this, we turned to chemical crosslinking believing it will rigidize the Pdu BMC thereby perturbing the labeling profiles explored in Chapter 2. This experiment scheme (Figure 5.1a) has several steps; (1) Pdu BMCs are crosslinked followed by (2) fluorescent labeling with

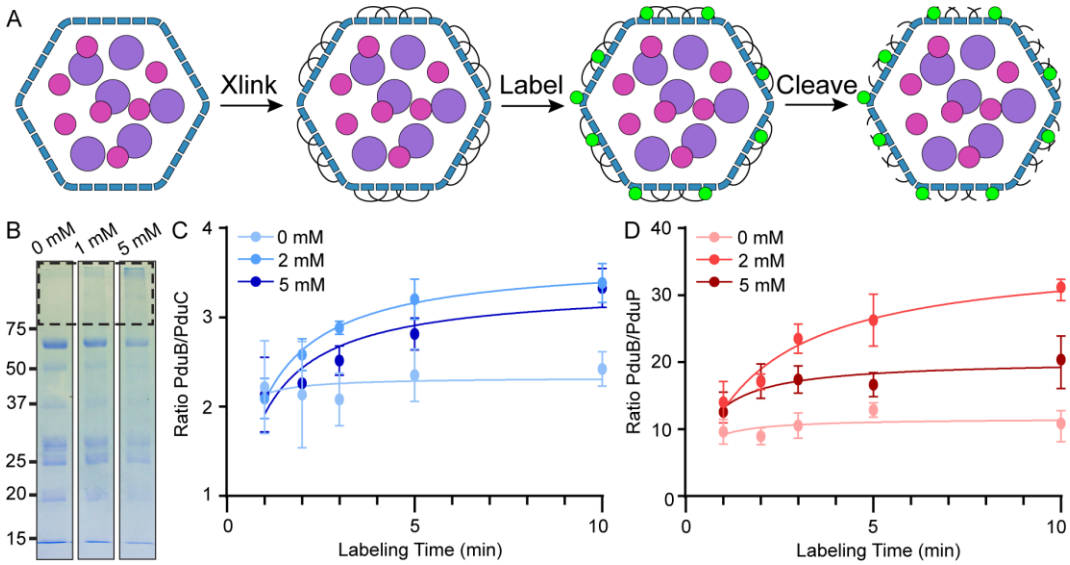


Figure 5.1: The permeability of the Pdu BMC can be chemically altered. (A) Experimental schematic where BMCs are first crosslinked *in vitro* with sulfo-EGS. After crosslinked BMCs, are exposed to the fluorescent label AF532 for a series of timepoints. The crosslinker is then cleaved prior to SDS-PAGE separation. (B) The crosslinker sulfo-EGS leads to the appearance of crosslinked species (dashed box) in a concentration dependent manner prior to cleavage. When the assay is carried out, the fluorescence for the major shell protein band for PduB was quantified and compared to that for the major cargo proteins PduC (C) and PduP (D) as a function of crosslinker concentration.

timepoints taken before finally (3) cleaving the crosslinker. This last step is critical to resolve individual proteins by SDS-PAGE, which will be used to quantify the labeling of shell and cargo proteins.

The crosslinker sulfo-EGS (ethylene glycol bis(sulfosuccinimidyl succinate)) and the fluorophore AF532 were used for this assay. Sulfo-EGS is a water-soluble, amine-reactive crosslinker that includes a hydroxylamine-cleavable linker arm. An amine-reactive crosslinker was chosen to enhance crosslinking sites while not competing with the thiol-reactive AF532 that was established in Chapter 2. When applied to Pdu BMCs, this crosslinker leads to dose-dependent high-molecular weight band formation (Figure 5.1b, black dashed box). The appearance of these bands indicates that crosslinking is successful and that any changes in labeling profiles would result from the crosslinking.

Pdu BMCs were crosslinked with either 0, 2 or 5 mM sulfo-EGS. Crosslinked Pdu BMCs were then subject to fluorescent labeling with AF532 in triplicate with timepoints taken after 1, 2, 3, 5, and 10 minutes. Each crosslinker in each sample was then cleaved with hydroxylamine and then resolved with SDS-PAGE. The labeling reaction was then quantified as a ratio of the fluorescent intensities of the PduBB' (shell protein) bands to either PduC (Figure 5.1c, cargo protein) or PduP (Figure 5.1d, cargo protein). Quantifying in this way helps control for sample-to-sample loading discrepancies. These proteins specifically were chosen due to forming robust, distinct bands in the SDS-PAGE profile (PduBB' = 25/28 kD, PduC = 63 kD, PduP = 50 kD, Figure 5.1b).

The results show that the labeling reaction becomes more biased for the shell protein PduBB' compared to the two cargo proteins when the BMC is crosslinked (Figure 5.1c, d). Pdu BMCs treated without sulfo-EGS display a constant ratio between the shell and cargo while sulfo-EGS treatment can enhance shell labeling, relative to the cargo, by >50%. The 2 mM sulfo-EGS condition appears to outperform the 5 mM condition, perhaps due to over-crosslinking of the outer shell inhibiting access to thiols on PduBB'. Specifically, PduB has a cysteine on both its concave and convex surfaces, so over-crosslinking could limit access to the interior facing thiol. Regardless, these results suggest that the permeability of the Pdu BMC can be perturbed via crosslinking which limits probe access to the interior cargo. These results may infer that crosslinking corrects for heterogeneity within the protein shell or stabilizes the shell into position thereby increasing its ability to act as a barrier.

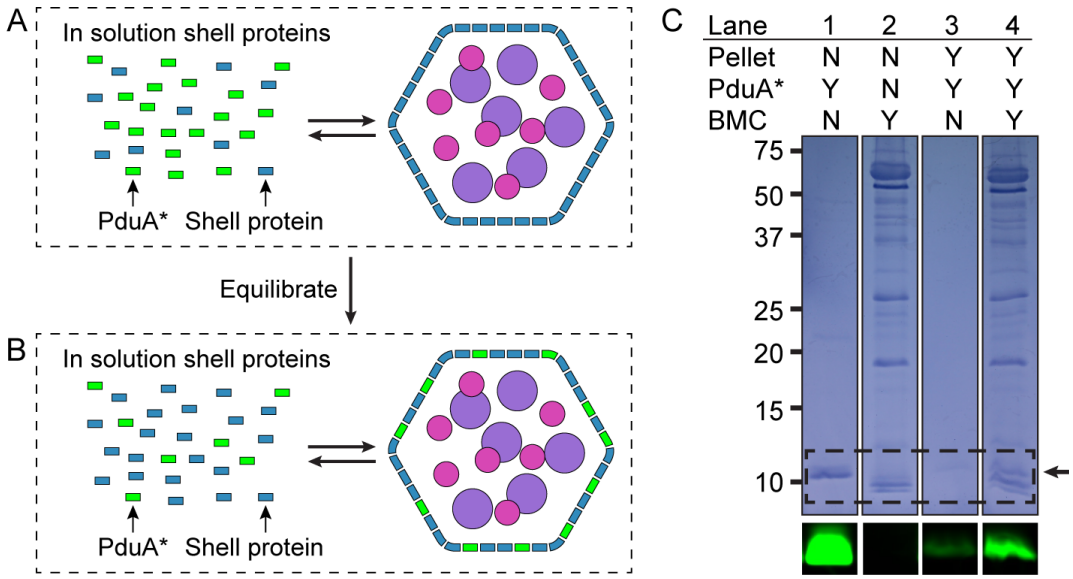


Figure 5.2: The protein shell interacts with whole Pdu BMCs *in vitro*. (A) The Pdu BMC may exist in an equilibrium between being "complete" and having shell components in solution. (B) The shell components could have the innate ability to exchange/scaffolds with other shell proteins within the BMC complex. (C) A PduA pulldown assay that uses Pdu BMCs as bait. Labeled PduA (PduA*) was incubated with or without Pdu BMCs and either pelleted or not. The black arrow denotes PduA*. The dashed box shows the area fluorescently scanned (bottom) confirming the presence of PduA*.

Subsection 5.2.2: Purified Shell Protein PduA Will Colocalize to Pdu BMCs

The above results combined with AFM studies (Sutter *et al.*, 2016; Faulkner *et al.*, 2019) invite an exciting prospect - that the protein shell is dynamic and the facet subunits exchange with other subunits in solution. We sought to explore this possibility by seeing if purified shell proteins will colocalize to Pdu BMC *in vitro*. We hypothesized that BMCs exist in an equilibrium with individual shell protein subunits that exist free in solution. (Figure 5.2a). We reasoned that if we supplied an exogenous BMC-H, PduA, labeled with a fluorophore (PduA*) to a solution of Pdu BMCs that Pdu* will interact with and potentially incorporate into the BMC shell (Figure 5.2b). Pdu BMCs can then be pelleted as a facile pulldown method to test for PduA* enrichment.

The purified PduA* exists in solution as a set of two bands (Figure 5.2c, lane 1). The minor upper band represents the small fraction of labeled PduA* which results in a mass shift from the major band. The fluorescent label and additional hexahistidine tag on PduA* give it a unique mass compared to the banding profile exhibited by native Pdu BMCs (Figure 5.2c, lane 2). Pdu BMCs and Pdu* were equilibrated and then BMCs were pelleted (lane 4). The Pdu BMC now includes the band for PduA*, which was confirmed via its fluorescence. PduA* can pellet on its own (lane 3) although this effect is greatly exacerbated when equilibrated with Pdu BMCs. The inherent insolubility of PduA is expanded upon in Chapter 6.

Still, PduA* will preferentially pellet with the Pdu BMC, indicating an interaction between the two. Imaginably, PduA is incorporated into the single-layered Pdu BMC shell. These results suggest that the shell layer itself is dynamic in a native context. It may be the case that the facets of the Pdu BMC behave in a dynamic manner as observed with BMC-H sheets (Sutter *et al.*, 2016; Faulkner *et al.*, 2019).

Subsection 5.2.3: Pdu BMCs Can be Imaged by Laser Scanning Confocal Microscopy

The observation that Pdu BMCs may be dynamic and exchange parts in solution required further investigation. However, it can be difficult to observe exchange between two relatively identical entities. For example, the exchange of components between, say, BMC A and BMC B is impossible to notice if BMC A and B are indistinguishable. For this reason, we developed an imaging approach based on laser scanning confocal microscopy. Confocal imaging allows sets of Pdu BMCs labeled with different fluorophores to be distinguishable by emission wavelength.

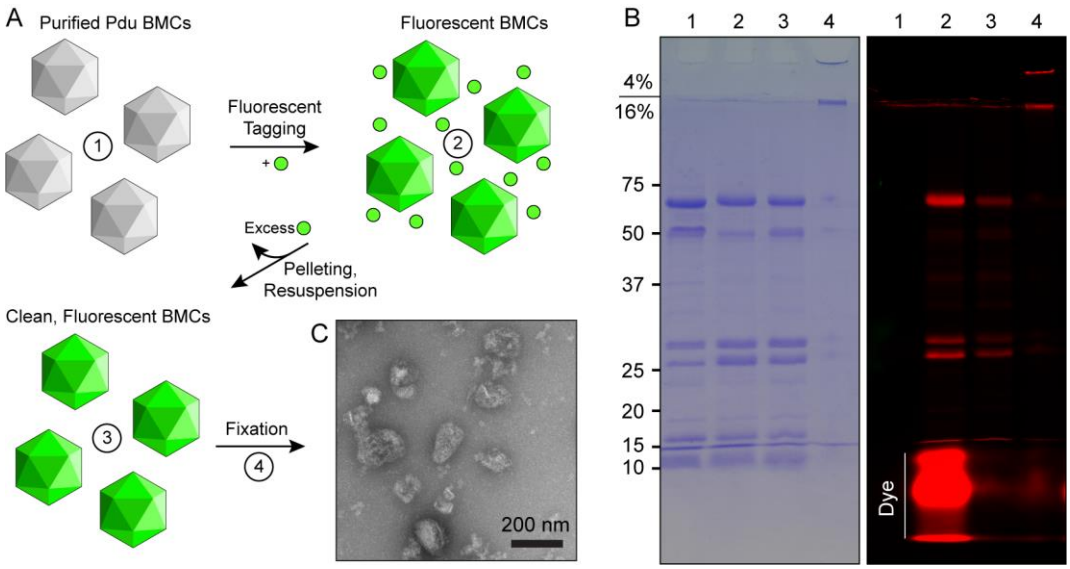


Figure 5.3: The preparation of Pdu BMCs for imaging with a laser scanning confocal microscope. (A) A scheme showing the steps in the preparation. Numbers correspond to the lanes in (B). (B) Samples at various stages of preparation were separated with SDS-PAGE to evaluate the scheme. (C) Fluorescently labeled BMCs were visualized with TEM to confirm they maintain their morphology.

A scheme was developed to prepare Pdu BMC for laser confocal imaging (Figure 5.3a). Here, Pdu BMCs are labeled with a fluorophore, like in Chapter 2, before the labeling reaction is quenched and labeled BMCs are then pelleted and resuspended. This step enables facile removal of excess dye. Labeled Pdu BMCs can then be fixed. This scheme was carried out and each step was analyzed with SDS-PAGE (Figure 5.3b). The SDS-PAGE confirms robust labeling (lane 2) and excellent removal of excess dye (lane 3). Labeled Pdu BMCs were then highly fixed in the presence of 2% PFA and 5 mM crosslinker (BS(PEG)5), which results in nearly all protein matter being unable to enter the resolving gel (lane 4). TEM analysis of labeling Pdu BMC confirmed that they maintain their typical morphology (Figure 5.3c).

Pdu BMCs were prepared with either AF488, AF680, or both as their fluorescent label. These samples were then imaged with a laser scanning confocal microscope equipped with an Airyscan detector which enables 1.7x higher resolution

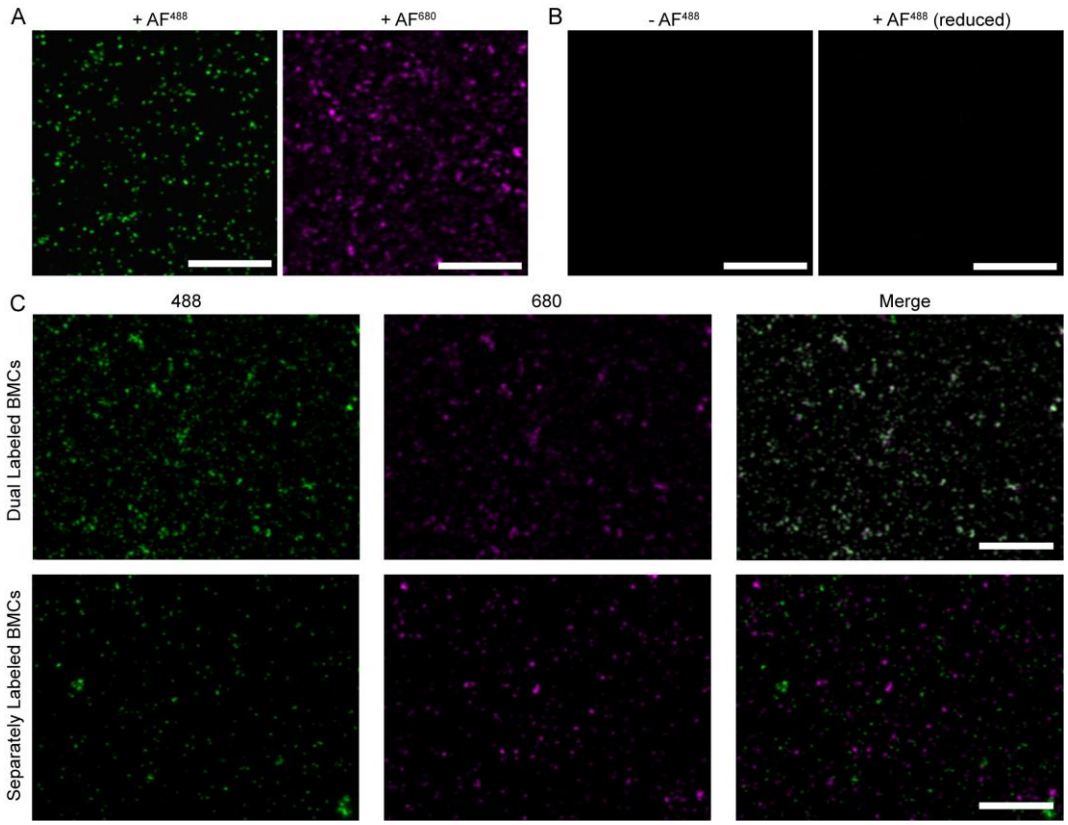


Figure 5.4: Pdu BMCs can be imaged with a laser scanning confocal microscope. (A) Pdu BMCs labeled with either AF488 or AF680 lead to small, distinct foci that settle onto the coverslip. (B) Unlabeled BMCs do not fluoresce (left panel). In addition, Pdu BMCs exposed to pre-reduced label (right panel) likewise do not fluoresce. (C) Testing the resolution of the instrumentation. (Top row) Pdu BMCs co-labeled with both AF488 and AF680 leads to white foci indicating high colocalization of green and purple signal. (Bottom row) Pdu BMCs separately labeled with either AF488 or AF680 remain as separable, distinct foci. All scale bars, 5 µM.

than otherwise similar scopes (Huff, 2015). Those prepared with either AF488 or AF680 both appear at small foci that settle onto the coverslip (Figure 5.4a). AF488 labeled BMCs, unsurprisingly, result in more defined foci thanks to the lower emission wavelength. Unlabeled BMCs do not fluoresce at all and neither do Pdu BMCs treated with AF488 that was pre-reduced with DTT (Figure 5.4b), indicating that the label itself is not depositing onto the coverslip. Dual-labeled Pdu BMCs colocalize extremely well (Figure 5.4c, top row) which indicates that the Airyscan detector can image colocalization despite the exceedingly small size of the BMCs. Pdu BMCs independently labeled with AF488 or AF680 were also mixed to test the microscope's

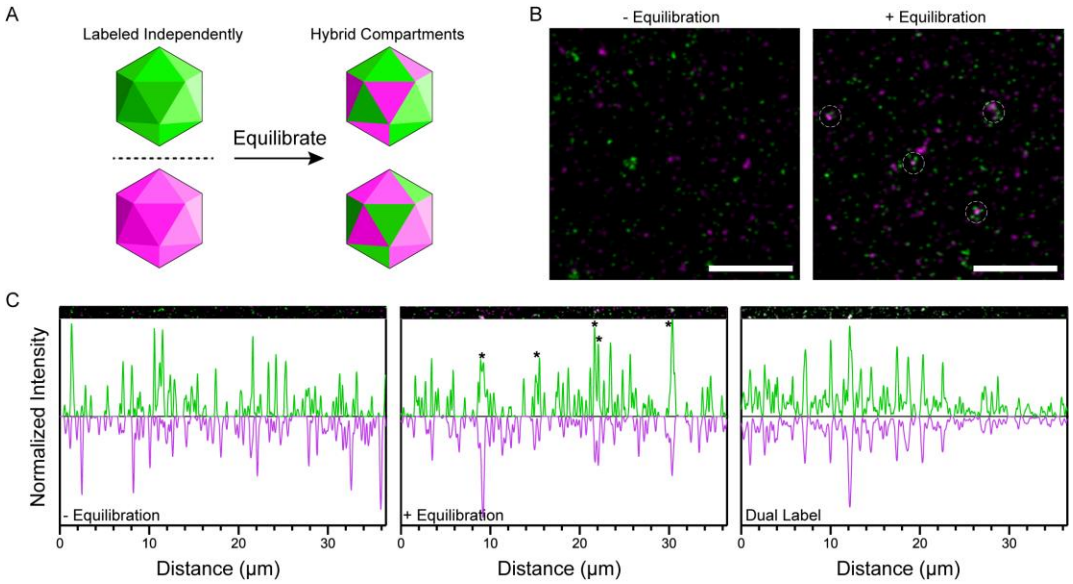


Figure 5.5: Confocal microscopy allows BMC colocalization events to be observed. (A) Schematic of experimental design. Separately labeled BMC batches (Alexa Fluor 680, Alex Fluor 488) were mixed, equilibrated, and imaged by laser scanning confocal microscopy. (B) To examine whether BMC preparations exhibited colocalization of BMC factors, samples were either imaged by confocal microscopy without being equilibrated first or were equilibrated for 30 minutes on ice before being imaged. Equilibrated samples revealed the presence of white foci (circled). Scale bars, 5 μM. (C) Plotting signal intensity from slices of micrographs. Colocalization events (asterisks) could be detected for equilibrated samples and for positive control samples that had been labeled with both fluorophores.

ability to resolve independent foci. Indeed, a mixed BMC sample included separate green and purple foci (Figure 5.4c, bottom row), indicating that separate foci can be distinguished from one another. These results confirm that Pdu BMCs can be directly imaged using a laser scanning confocal instrument.

Subsection 5.2.4.: Uniquely Dyed Pdu BMCs Will Colocalized When Equilibrated

Confocal analysis allows one to distinguish separate BMC foci and detect colocalization events (Figure 5.4c). We sought to leverage this ability to study the equilibrium properties of BMC in vitro. In this experiment, BMC samples are independently labeled with different fluorophores prior to being mixed and equilibrated (Figure 5.5a). Intervening fixation steps can be used to moderate the potential equilibration and exchange of fluorescently tagged BMC components.

Colocalization events can be detected visually (Figure 5.5b) albeit with difficulty. Green and purple foci overlapping will lead to white foci when merged (Figure 5.5b, right). However, white foci stem primarily from equivalent overlap of green and purple intensities. Therefore, one's own eyes would miss out on unequal overlaps that are nonetheless critical to evaluate the extent of sample overlap.

This frustration led to a more quantitative approach where strips of micrographs had their green and purple signals independently assessed and plotted against each other (Figure 5.5c). When each scan is normalized to the peak intensity within that scan, this method leads to a series of fluorescent "spikes" that correspond to BMC foci which can, in some cases, coincide. Labeled BMCs that were fixed prior to equilibration (Figure 5.5c, left panel) reveal an apparent random distribution of green and purple foci. This can be interpreted as representing a complete lack of protein exchange due to the chemical fixation. This is in stark contrast to dual-labeled BMCs, a positive control, which show a mirror-image distribution in the fluorescent peak (Figure 5.5c, right panel). Between these two extremes lies separately labeled BMCs that were allowed to equilibrate for 30 minutes (Figure 5.5c, center panel). These samples contain a combination of random peaks as well as some that obviously coincide. These results suggest that BMCs can, to some degree, equilibrate their protein components.

While the above analysis represents a clear improvement over flawed visual assessment, it still suffers from some pitfalls. For one, the slices from micrographs may be cherry-picked. It also suffers from noise and limited resolution. For that reason, we elected to pursue van Steensel cross-correlation analysis (van Steensel *et al.,* 1996).

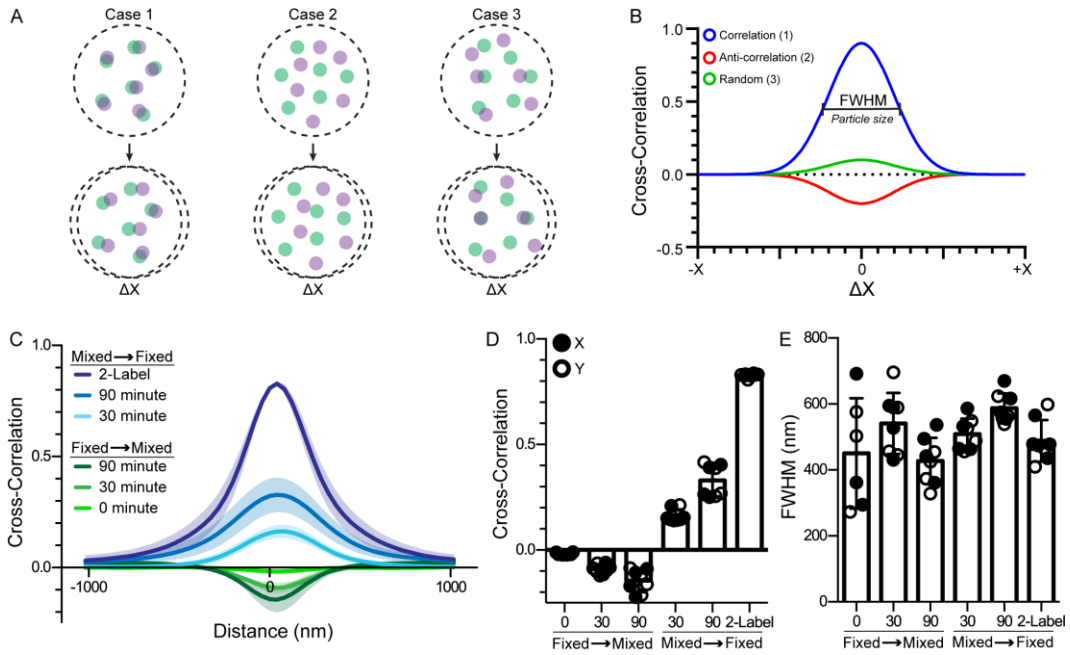


Figure 5.6: Cross-correlation analysis reveals that proteins from Pdu BMCs colocalize. (A) The three different scenarios for particle distributions on a microscope slide can be identified with van Steensel cross-correlation analysis (B). (C) Pdu BMCs were either fixed prior to ("Fixed → Mixed") or after ("Mixed → Fixed") being allowed to equilibrate, then imaged and analyzed for cross-correlation via the van Steensel method (van Steensel *et al.*, 1996). Four micrographs for each sample were analyzed in both the X and Y directions. (D) The peaks of each curve in (C) are shown in this graph. Data analyzed in the individual X and Y directions are shown. (E) The full-width of the Gaussian distributions in (C) at half-max are plotted here, showing that particle size estimates appear approximately equal for all samples.

Van Steensel analysis measures the signal overlap between two different fluorescence scans on a pixel-by-pixel basis prior to shifting one scan relative to the other in the positive or negative X direction (Figure 5.6a). The signal cross-correlation can then be plotted as a function of the pixel shift which can then be converted to literal distances (Figure 5.6b). Classes of correlation can be defined by the shapes of the curves where positive distributions imply correlation (with higher value being more correlated), peaks near zero indicating a random particle distribution, and negative values equating to anti-correlation and mutual exclusion of particles. The full width at half max (FWHM) of the curves also yields information on the approximate particle size. This analysis method is considered particularly useful for very small particles where direct

colocalization is difficult to establish (Bolte and Cordelieres, 2006) and can be applied to full fields of view, limiting selection bias.

Van Steensel analysis was carried out for labeled BMC samples that were either allowed to equilibrate for different times or prevented by fixation prior to equilibration. Four unique fields of views were analyzed per sample in both the X and Y axis in case isotropic assumptions for BMCs are not valid. This resulted in 8 measurements per sample that were averaged and plotted with their standard deviations. Unsurprisingly, a positive control sample composed of dual-labeled BMCs had a high extent of colocalization (Figure 5.6c) peaking near 0.85 (Figure 5.6d). Differently labeled BMC samples that were equilibrated for 30 or 90 minutes prior to fixation and imaging likewise exhibited positive correlation (0.18 and 0.33, respectively) (Figure 5.6c,d). These samples show that equilibration positively influences cross-correlation albeit to a lesser extent than direct dual-labeling. In contrast, differently labeled BMC samples that were fixed prior to equilibration for similar timepoints displayed anti-correlation (Figure 5.6c). This indicates that the intervening fixation prevents exchange of fluorescent protein matter between BMCs, leading to the mutual exclusion of green and purple particles. Lastly, while sample treatment differed between samples, all had an approximate diameter of 500 nm (Figure 5.6e). While larger than a typical Pdu BMC, these results are unsurprising given the potential for point spread interference at these scales. This effect could be further limited using lower emission-wavelength fluorophores and faster scan rates.

These results advocate for the controversial contention that Pdu BMCs exchange parts. Fixation negatively affects this due to limiting the freedom of tagged

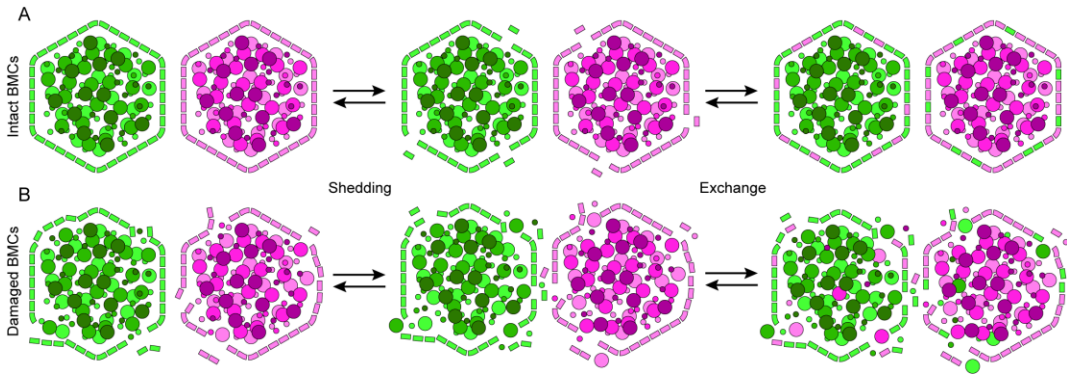


Figure 5.7: Models for how Pdu BMCs behave in solution. (A) Pdu BMCs are dictated by an equilibrium with loose shell proteins in solution. These shell proteins can exchange and replace missing pieces within the shell facets. (B) Conversely, the Pdu BMCs used in these studies may exist in a damaged state. These BMCs non-specifically slough off debris exchange and intermix. This model is expanded upon in Chapter 6.

proteins. However, these results cannot directly attest to whether this exchange is due to just the outer shell components, although we present a pilot experiment that shows purified shell proteins will interact with native Pdu BMCs in vitro (Figure 5.2c). Some may note that the BMCs tested are damaged and likely intermixing in a non-specific manner, although DLS and TEM results indicate that Pdu BMCs survive the fluorescent labeling process (Figure 2.6b, Figure 5.3c)

Section 5.3: Conclusions

The data reported here heavily suggest that the Pdu BMC is a dynamic system. Accordingly, we present two models for how the Pdu BMC behaves in solution. In one model, shell proteins exist in an equilibrium between the protein shell and the outside solution (Figure 5.7a). These shell proteins are then free to exchange to other BMCs beyond that from which they originated. The second model presents a dilemma where the Pdu BMCs used here are in a damaged state (Figure 5.7b). Pdu BMCs in this state may non-specifically shed both cargo and shell proteins which intermix. These two models may be difficult, if not impossible, to distinguish between. Some evidence

reported here tentatively discredits the second model. First, it assumes that the Pdu BMC cannot "fix" itself if it is damaged despite all the required protein components being present. Second, this argument relies on the presumptive notion that the Pdu BMC is static and any observed behavior to the contrary is necessarily due to structural abnormality. Instead, we believe that a flexible structure is perhaps a key part in the design of the Pdu BMC.

The results reported here would be interesting to contrast with the much more rigidly structured carboxysomes. These results should also be contextualized for biological significance. It may be that BMCs exist on a continuum of structural plasticity where evolutionary forces select for rigidity in the oxygen-blocking carboxysomes. At the other end of that spectrum may exist the Pdu BMC, where the forces at play are not as intense and structural flexibility is allowable or, perhaps, even intentional.

BMCs such as the Pdu BMC may be dynamic in order to be highly responsive to environmental conditions. In short, they appear when they are needed and are dismantled, through an unknown mechanism, when their substrate is depleted. Carboxysomes, in contrast, are essential for cyanobacterial growth and are directly inherited during cell division thanks to exquisitely controlled cellular positioning (Kerfeld and Melnicki, 2016; Rillema *et al.*, 2021). In fact, a parent cyanobacteria will see its carboxysomes directly diluted amongst its daughter cells over numerous generations if those cells have their own carboxysome expression blocked (Hill *et al.*, 2020). This suggests that the robust, unyielding structure of the carboxysome is a requirement to ensure (1) limited oxygen exposure to Rubisco and (2) guarantee

inheritance for survival. The Pdu BMC does not meet these criteria, perhaps explaining its livelier qualities. The observations reported here are likely the tip of the iceberg for this emerging topic and will encourage discussions on controllable dynamism within BMC platforms.

Section 5.4: Methods and Materials

Subsection 5.4.1: Dye Permeability Assay

BMCs were prepared in buffer (20 mM sodium phosphate pH 7.4, 50 mM KCl, 5 mM $MgCl_2$) at a concentration of 1.0 mg/mL. Sulfo-EGS crosslinker was then added to the indicated concentrations and samples were allowed to crosslink for 30 minutes at room temperature. After crosslinking, AF532 was added to a final concentration of 125 μM and the reaction was carried out at room temperature. At the indicated time points, reaction aliquots were rapidly quenched in 1x standard SDS-PAGE loading buffer supplemented with 50 mM DTT, to end the AF532 labeling, and 500 mM hydroxylamine, to cleave sulfo-EGS. Samples were then incubated at 37°C for 4 hours protected from light prior to separation via standard SDS-PAGE. Gels were imaged on an Amersham Typhoon gel imager and quantification was carried out in ImageJ.

Subsection 5.4.2: Purification and Fluorescent Labeling of 6xHis PduA S93C

The plasmid encoding PduA S93C was sub-cloned via Gibson assembly, sequence confirmed, and transformed into T7 Express to create strain DT074. Overnights were used to inoculate 500 mL of 2xYT supplemented with antibiotic to an OD600 and grown at 37°C to an OD600 of 0.6-0.8. PduA was expressed by addition of IPTG to 0.5 mM and incubated at 30°C for 4.5 hours. Cultures were then pelleted,

frozen/thawed, and resuspended in Buffer A (50 mM Tris-HCl pH 8.0, 1 M NaCl, 5 mM DTT, and 10 mM imidazole) and lysed via sonication. Insoluble matter was pelleted for 20 minutes at 13000xg, 4°C. The soluble lysate was then equilibrated with 1 mL of equilibrated IMAC resin for 30 minutes at 4°C before being applied to a gravity column. Flowthrough was drained off and the beads were washed with 5 mL of Buffer A, 5 mM of Buffer B (Buffer A + 35 mM imidazole) and PduA S93C was eluted in 5 mL Buffer C (Buffer A + 250 mM imidazole). Eluted protein was dialyzed against 20 mM Tris-HCl pH 7.4, 100 mM NaCl, 10% glycerol. Purified PduA were labeled with AF532 at a ratio of 1:2 AF532:PduA for 2 hours on ice. Purified and labeling proteins were stored at -20°C for later use.

Subsection 5.4.3: PduA Pulldown Assay

PduA* was incubated either with or without 1.0 mg/mL Pdu BMCs for 30 minutes on ice at a concentration of 50 μM in BMC buffer (20 mM sodium phosphate pH 7.4, 50 mM KCl, 5 mM $MgCl_2$). After incubation, samples were pelleted at 20000xg for 15 minutes. Supernatant was removed and BMC pellets were resuspended in buffer and samples were separated on SDS-PAGE.

Subsection 5.4.3: PduA Pulldown Assay

Pdu BMCs were labeled with AF488 as described in Chapter 2.4.4. TEM sample preparation and subsequent imaging was carried out as described in Chapter 2.4.3.

Subsection 5.4.4: TEM of Fluorescently Labeled Pdu BMCs

Pdu BMCs were labeled with AF488 as described in Chapter 2.4.4. TEM sample preparation and subsequent imaging was carried out as described in Chapter 2.4.3.

Subsection 5.4.5: Preparation of BMCs for Confocal Imaging

Pdu BMCs were prepared to 1.0 mg/mL in buffer (20 mM sodium phosphate pH 7.4, 50 mM KCl, 5 mM MgCl2). Thiol-reactive fluorophore is then added to 125 μM and BMCs were labeled for 15 minutes on ice protected from light. Labeling is then quenched with the addition of DTT to 5 mM. Labeled Pdu BMCs (either pre- or post-equilibration with different labeled BMCs) were mixed 1:1 with a freshly prepared 4% paraformaldehyde, 10 mM BS(PEG)5 solution in 1x PBS. Fixation occurred at room temperature for 30 minutes. Labeled, fixed BMCs were then pelleted at 20000xg for 10 minutes at 4°C and resuspended at their original volume of buffer.

Subsection 5.4.6: Confocal Imaging and Data Analysis

Samples were imaged on a Zeiss LSM 980 Airyscan at room temperature using a 63x objective with immersion oil. Briefly, 20 μL of sample solution was pipetted directly onto a #1.5 coverslip (Fisherbrand, 12544E). Initial focus was achieved by focusing on the edge of the sample droplet under transmission mode before switching to fluorescence and focusing down to the coverslip surface containing BMC foci. The foci are plentiful and bright enough to allow live calibration of the Airyscan detector in super resolution mode. The fastest possible image scan time was always used with a minimum pixel scan time of 1.00 μs. All captured and reported images had standard Airyscan processing applied. Captured, unedited images were all analyzed in ImageJ.

Where the results indicate, the JACoP plugin (Bolt and Cordelieres, 2006) was used to implement van Steensel cross-correlation analysis (Van Steensel *et al.*, 1996) with a pixel shift of 25. Images were converted to 16-bit grayscale for this purpose.

Chapter 6: Confocal Studies on Shell Protein Super-Structures

Section 6.1: Introduction

The protein shells of BMCs self-assemble into a polyhedral structure. Strictly speaking, this ability is dictated by (1) shape complementarity of the protein subunits and by (2) the pairing of edge-facing lysine residues in an antiparallel orientation on adjacent protein tiles (Sutter *et al.*, 2017). However, proper BMC shell maturation also depends delicately on also the balance of what shell proteins are available as well as their relative stoichiometries. For instance, it has been noted that empty Pdu polyhedra can be derived *in vivo* from a combination of the proteins PduABJKN; PduTU are not necessary (Parsons *et al.*, 2010). The Hoch BMC (Figure 5.1a) can likewise assemble from different combinations of shell components expressed from a single synthetic operon encoded the genes preceded by differently tuned ribosomal binding sites (Lassila *et al.*, 2014; Greber *et al.*, 2019). Other BMC, like a recently resolved minimal α-carboxysome shell, can produce distinct populations of BMC with different triangulation values within the same sample (Tan *et al.*, 2021). These results suggest that native BMC exist as an exquisite balance between different shell components.

Changing the relative abundances and classes of shell proteins can also lead to other diverse, non-polyhedral structures. A set of experiments from Parsons *et al.* found that overexpressing different combinations of Pdu shell proteins led to both sheet and filament-like structure *in vivo* (Parsons *et al.*, 2010). These *in vivo* observations also apply to overexpression of some singular shell proteins such as RmmH (Noël *et al.*, 2016), PduA (Havemann *et al.*, 2002; Parsons *et al.*, 2010; Pang *et al.*, 2014) and EutM

(Pitts *et al.*, 2012). Purifying these single-protein complexes from cellular lysates confirmed that they form rigid nanotubes *in vitro* as well, as evidenced by TEM microscopy (Noël *et al.*, 2016; Uddin *et al.*, 2018; Hagen *et al.*, 2018b). The nanotubes present as rigid structures as either single tubes or in parallel-oriented clusters. Further, these nanotubes are typically ~20-70 nm in diameter, several micrometers in length, and are thought to exhibit a concave-out orientation (Uddin *et al.*, 2018) much like native shells as outlined in Chapter 4.

Little is known about how shell protein nanotubes behave in solution. This is due to the reliance on TEM to observe these structures. TEM analysis, albeit immensely useful, strips away the ability to observe solution properties. In addition, TEM cannot be used to robustly study protein colocalization events as different protein subunits cannot be resolved. These limitations led us to leverage confocal microscopy to study how these structures behave in solution. Confocal circumnavigates the above problems presented by TEM by relying on colored differentiation of different samples in solution. In this chapter, we purify and fluorescently label a suite of Pdu BMC shell proteins and image them with TEM and confocal microscopy. The images reveal the morphological differences between certain Pdu shell proteins, how they behave in solution, and suggest how they influence overall BMC morphology. Our results also allow us to speculate on the potential biological importance of these non-native structures, utility, and on the biogenesis pathways for native Pdu BMCs.

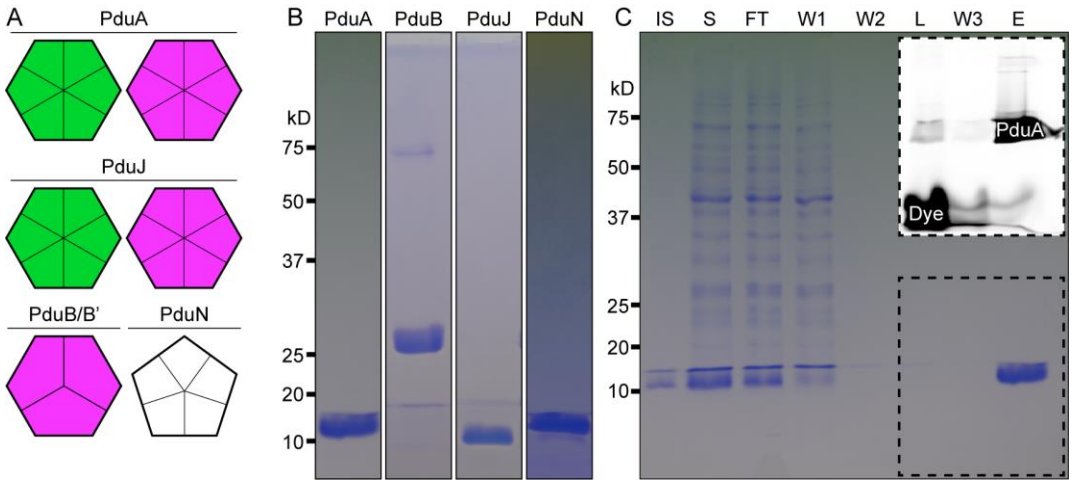


Figure 6.1: The suite of Pdu BMC shell proteins chosen for purification. (A) PduA, PduJ, PduB, PduB' and PduN were chosen to be purified. Proteins are represented as their oligomeric states. Green represents purification with an AF488 label, while purple correspond to AF680. (B) Pdu shell proteins were able to be purified with reasonably high purity. (C) The purification profile for PduA S93C that was labeled with AF680. IS, insoluble fraction; S, soluble fraction; FT, flowthrough; W1, wash 1; W2, wash 2; L, eluant post-labeling; W3, wash 3; E, elution. Inset was fluorescently scanned to reveal the extent of AF680 removal from the eluted protein fraction.

Section 6.2: Results and Discussion

Subsection 6.2.1: Purification and On-Column Fluorescent Labeling of Shell Proteins

To begin, several strains were created to overexpress select Pdu shell proteins for nickel purification. The shell proteins selected were the highly abundant shell proteins PduA, PduB, PduJ (Yang et al, 2020) as well as the vertex protein PduN (Figure 6.1a). PduA and PduB were selected due to being previously implicated in nanotube formation (Uddin *et al.*, 2018; Havemann *et al.*, 2002; Li *et al.*, 2021) while PduJ was selected as it is an often ignored, yet more abundant (Yang *et al.*, 2020), homolog of PduA whose nanotube formation propensity remains enigmatic. The vertex protein PduN, in contrast, is lowly abundant but may act to prime the overall BMC polyhedra (Li *et al.*, 2021). All proteins used an N-terminal hexahistidine tag except PduN which used a C-terminal tag, in accordance with previously published

purification efforts (Wheatley *et al.*, 2013). Note, the PduA used in this chapter is an S93C mutant. This mutation adds a cysteine residue that can be selectively modified with a fluorophore, outlined below.

Initial purification trials on PduA yielded a high amount of insoluble protein (not shown), likely a result of its propensity to self-assemble into larger structures that pellet with other insoluble cellular material. Indeed, this issue is mentioned commonly throughout the literature. We reasoned that this could be limited by directly affecting one of the key shell protein interactions; the edge-facing lysines that form an antiparallel hydrogen bonding arrangement with adjacent shell subunits (Figure 1.2b). We presumably disrupted this interaction by increasing the NaCl concentrations in our buffers from 400 mM to 1000 mM. This small buffer change dramatically increased the yield during purification and was easily applied towards all the shell proteins chosen for purification, resulting in a one-size fits-all purification scheme. Accordingly, PduAJBB'N could be purified to high concentration and reasonable purity with a single nickel purification step (Figure 6.1b).

The last protein preparation step was to tag each purified protein with a fluorophore *in vitro*. To accomplish this, we again turned to maleimide-based conjugation of cysteine residues. Specifically, we wanted to create AF488 and AF680 labeled batches of shell proteins. This method of conjugation, compared to NHS-esters, has the benefit of not disturbing lysine residues which are critical in shell-shell interactions. While this method of labeling is relatively easy to accomplish, the removal of excess dye afterwards proved to be a much more daunting obstacle. Dialysis was futile even after several days. Desalting columns would work and are much quicker,

but they resulted in major sample loss likely from the quasi-insolubility of these proteins. To circumnavigate these issues, we developed a protocol to label the selected shell proteins during the purification process while they were adhered to the nickel resin (Figure 6.1c).

Shell proteins would largely be purified as normal during this workflow up until the wash steps after column binding. At this point, samples were washed with a middling concentration of imidazole to remove non-specifically bound proteins and equilibrate the sample to the correct pH for labeling (Figure 6.1c, W2). Fluorophore would then be added to and equilibrated with the bead-bound proteins for up to 2 hours before excess was drained away (L) and the sample was washed thoroughly (W3).

Altogether, this method results in minimal protein loss, excellent excess dye removal, decent labeling yields (~3%), and allows reuse of the nickel resin. Labeling yields could easily be improved with longer incubation and by simply adding more dye as a typical reaction would only use ~100 nmol of fluorophore to save on resources. This method was used to purify PduA and PduJ labeled with both AF488 and AF680 and PduB and PduB' labeled with just AF680 (Figure 6.1a). PduN was not labeled as we reasoned that a vanishingly low quantity of it would be needed.

Subsection 6.2.2: The BMC-H Protein PduA Forms Rigid Nanotubes

PduA was screened for its ability to form higher order structures with confocal microscopy and TEM. Confocal microscopy confirmed the presence of elongated, dendritic protein clusters and occasionally flat, hexagonal structures (Figure 6.2a). All of these structures were large enough to also be observed under brightfield conditions with a 63x objective with oil immersion (not shown). PduA samples were then subject

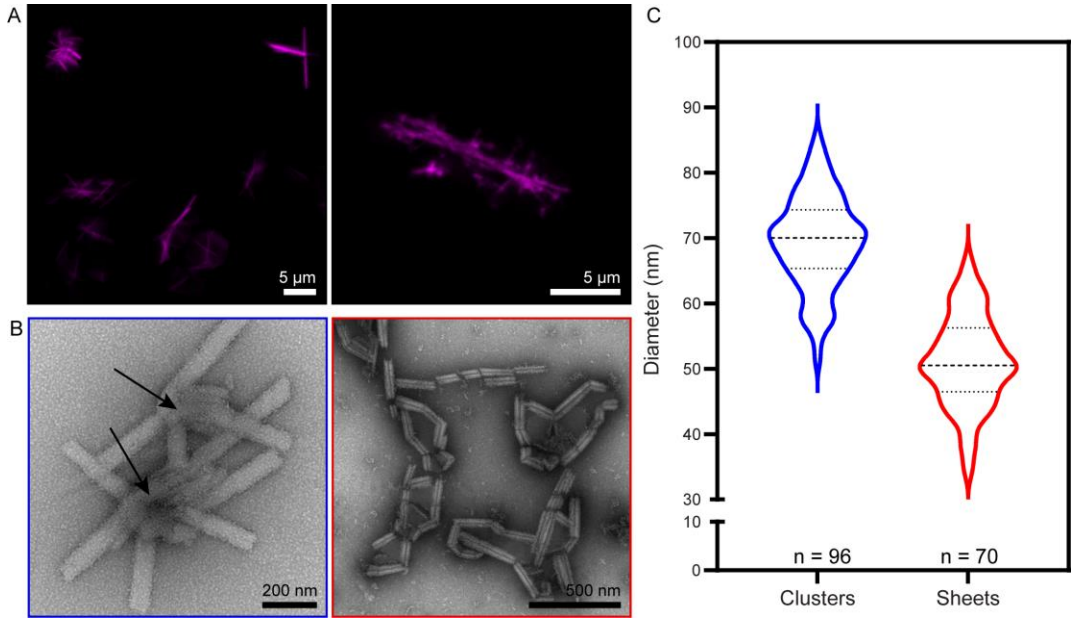


Figure 6.2: Imaging of PduA superstructures. (A) Laser scanning confocal microscopy analysis of Alexa Fluor 680-tagged PduA reveals protein nanotubes. (B) Transmission electron microscopy (TEM) analysis of PduA structures reveals clusters of nanotubes (outlined in blue) or nanotubes associated with sheets (outlined in red). Arrows denote clustering centers. (C) The diameters of PduA nanotubes differ nanotube clusters and sheet-associated nanotubes.

to TEM to gain deeper insight into the nature of these structures. TEM confirmed that the elongated structures observed under confocal were indeed rigid nanotubes - exemplified by the presence of a darker contrasted strip within the center of some structures where uranyl acetate deposited (Figure 6.2b, right).

The nanotubes were observed in two distinct morphologies; one being nanotubes arranged in clusters where nanotubes erupt from a central region (Figure 6.2b, arrow) and the other being nanotubes surrounding and being associated with centralized sheets (Figure 6.2b, right). The former instance recalls analogies to a nucleation site. In contrast, the sheet-associated nanotubes have a more distinct outer edge while the inner edge appears to bleed into the sheet formations in some instances. Far from being just visually distinct, these two nanotube morphologies are also structurally distinct. Nanotubes that appear in clustered arrangements are, on average,

20 nm wider in diameter than those associated with sheets (Figure 6.2c). This may reflect the "bleeding over" observations of the sheet-associated nanotubes, where this phenomenon may be better understood as the nanotubes unraveling into sheets. Likewise, they may also be sheets rolling up into nanotubes. Regardless, sheet-associated nanotubes may be in the process of conformationally switching between tubes and sheets which is reflected by a thinner diameter. We speculate that these two phenotypes either represent two different modes of nanotube generation or two different stages in the nanotube life cycle.

These PduA nanotubes are also structurally distinct, larger, and less uniform than previously reported PduA nanotubes described by Uddin *et al.* (Uddin *et al.*, 2018). The PduA nanotubes of Uddin *et al.* are 20 nm in diameter, highly uniform, and appear largely as distinct nanotubes. While clusters are observed, these clusters have the PduA nanotubes oriented in a parallel fashion as opposed to emerging from a central point as we observed. These differences may stem from slightly different buffer conditions where Uddin *et al.* used Tris-HCl pH 8.0, 50 mM NaCl and we used Tris-HCl pH 8.0, 50 mM KCl, 5 mM $MgCl_2$ (analogous to the buffer reserved for purified Pdu BMCs) although it is not clear if these minor alterations would have any effect. A more convincing explanation can instead be offered by the type of PduA used in each study, where Uddin *et al.* used a serendipitously derived mutant of a typical PduA. This PduA encodes for an additional 23 amino acids past the usual C-terminus (resulting from a stop codon mutation) which resulted in a more soluble form of PduA (Parsons *et al.*, 2010; Pang *et al.*, 2014). Accordingly, we suggest that the nanotubes we describe

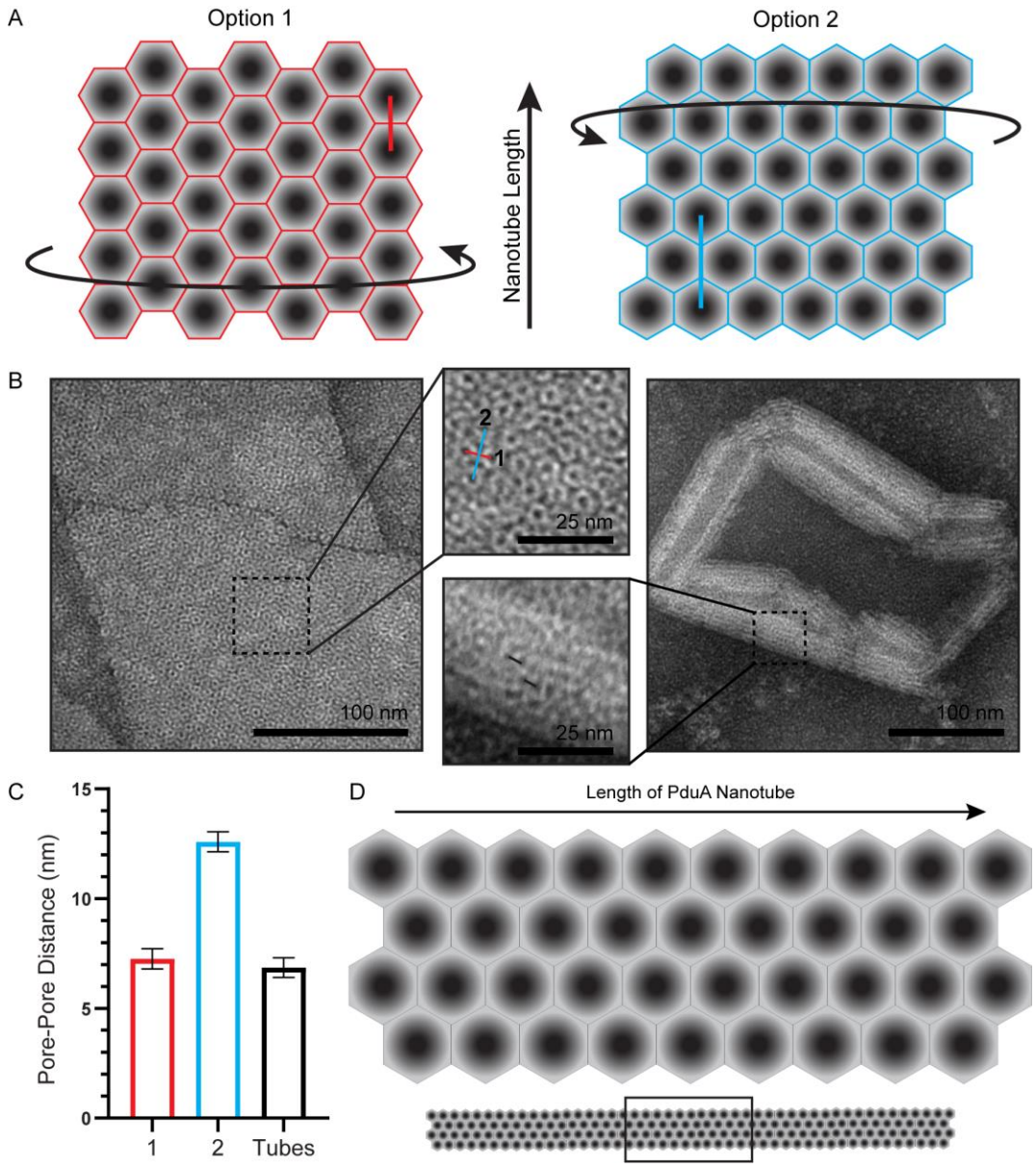


Figure 6.3: Analysis of PduA nanotubes. (A) The individual hexamers can orient one of two ways within the PduA nanotube substructure. (B) TEM of PduA sheets and sheet-associated nanotubes. The pore distances can be measured from TEM micrographs. (C) The measurements from (B) show that PduA nanotubes orient like option 1, shown in panel (A). (D) A model for the composition and orientation of PduA hexamers within the nanotube lattice.

are more biologically relevant, especially when one considers how the C-terminus of PduA may influence shell assembly (discussed in Chapter 4).

In some instances, individual PduA hexamers could be resolved with TEM on both flat sheets and nanotubes. In the sheets, hexamers appear as sets of concentric

circles with a black pore packed in a hexagonal arrangement. We believed the ability to resolve the hexamers may allow us to determine the geometry (or orientation) of PduA hexamers within the nanotubes. The hexamers can only pack in one of two ways; (1) with the pores of adjacent hexamers aligned as close as possible along the long axis of the nanotube (Figure 6.3a, option 1) or (2) with pores being separated by an intervening pair of hexamers (option 2). This latter case necessarily has the hexamer pores measurably further apart than the former option.

Using the flat sheets as a reference (Figure 6.3b), we measured numerous instances of both orientations (Figure 6.3c) and found that the first option had pores spaced by ~7.1 nm (about the width of a PduA hexamer) while the second option had a spacing of ~13.8 nm. While the nanotubes had lower resolution, we were able to measure the distances of adjacent pores (i.e. black patches) along the long axis of the tubes to a distance of ~7.0 nm (Figure 6.2c). This clearly matches with option 1, where PduA hexamers align as observable strips of hexamers along the axis of the nanotube resulting in the observed striations (Figure 6.3d). This addresses a key unanswered question about the construction of these assemblies. This knowledge may be helpful for those seeking to model the assembly of not just shell protein nanotubes, but also the facets and overall assembly of BMC polyhedra.

Subsection 6.2.3: The BMC-H Protein PduJ Forms Flexible Nanotubes

The Pdu BMC operon from *S. enterica* encodes for a BMC-H homolog of PduA called PduJ (Figure 2.2a). PduA and PduJ have largely been considered functionally redundant for shell assembly, although PduJ is more abundant (Yang *et al.*, 2020; Kennedy *et al.*, 2021), due to their high degree of sequence identity (Figure 4.4a).

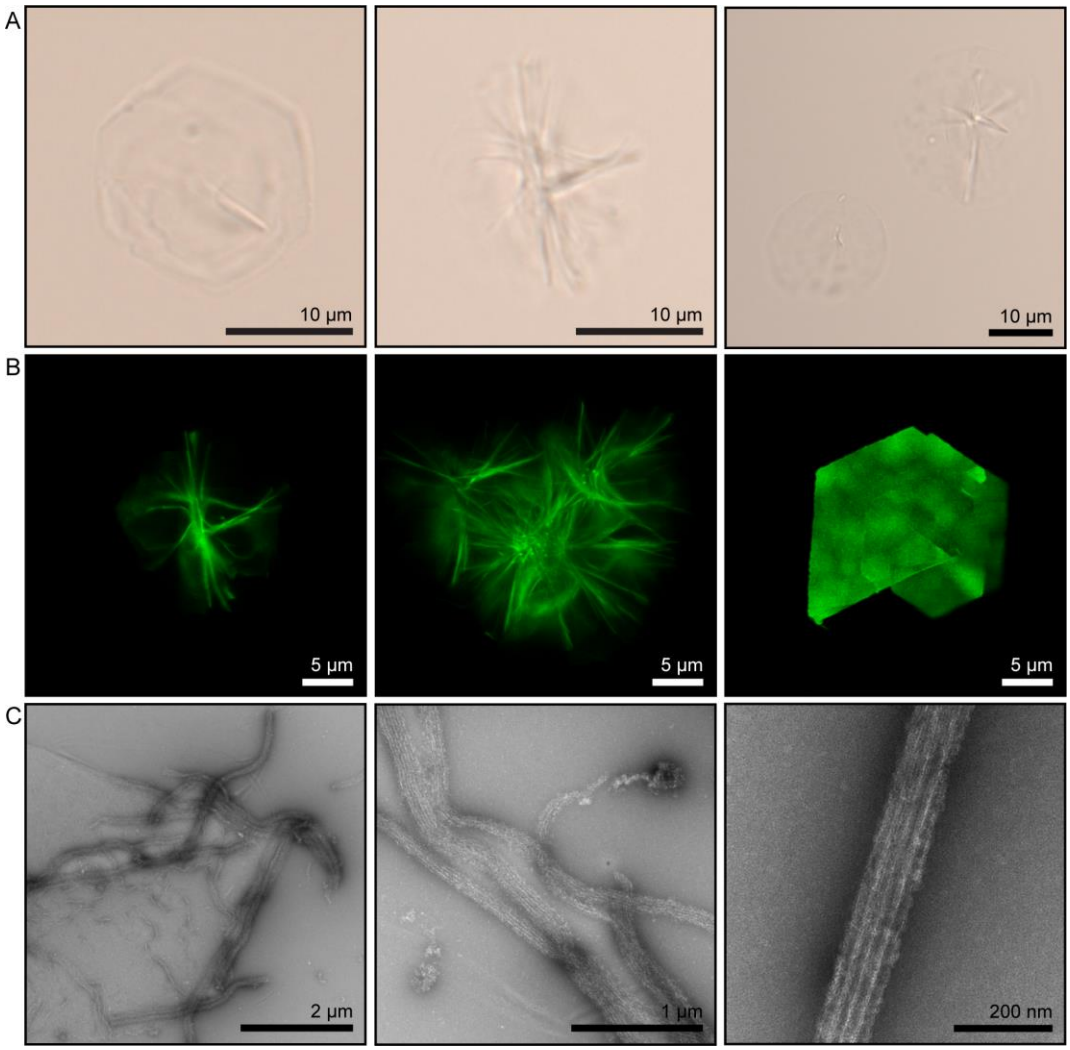


Figure 6.4: The structures formed by PduJ. (A) PduJ forms rosettes and hexagonal sheets that are easily observable with brightfield microscopy. (B) Confocal microscopy reveals the internal structures of these structures. (C) TEM gives deeper insight into PduJ rosettes, which are clusters of flexible nanotubes.

Further, purified PduJ has likewise recently been observed to form nanotubes (Kennedy *et al.*, 2021).

We were able to confirm that purified PduJ does form higher-order assemblies, although with a morphology wholly distinct from both PduA and previously published cases. When viewed simply under brightfield conditions, PduJ forms distinct hexagonal sheets and rosettes as opposed to filamentous structures (Figure 6.4a). Confocal analysis confirms this morphology while revealing the more delicate details of what

appears to be an internal structure to the rosettes (Figure 6.4b). Scanning the depth of field throughout these rosettes reveals that they are three-dimensional entities that are a network of PduJ filaments. TEM confirms that these filaments are sheet-associated nanotubes, similar to some cases for PduA, albeit with a much higher degree of flexibility (Figure 6.4c). When zoomed in, these structures bring to mind the ebb and flow portrayed by aerial views of highway intersections, with nanotubes merged and splintering off of one another. Those that splinter off can result in general protein aggregates (Figure 6.4c, middle panel). The striations observed along the PduJ nanotube axis seem analogous to those observed in PduA nanotubes (Figure 6.4c, right panel) which causes us to believe they assemble with a similar geometry.

Given that the major sequence deviations between PduA and PduJ occur along their convex surfaces (Figure 4.4a), we cannot speculate as to the structural nature of the differences observed. Overall, PduJ forms nanotubes with a similar geometry as PduA nanotubes. However, PduJ nanotubes have a greater degree of flexibility when imaged under the same conditions. This may reflect slight differences in their structural roles for complete BMC assembly, previously suggested by the specific interactions captured with the BMC-T protein PduB (Figure 4.5).

Subsection 6.2.4: The Shell Proteins PduBB'N Form Aggregates

The previous sections explored the structures formed by the two predominant BMC-H proteins, PduA and PduJ, in the Pdu BMC. We next sought to survey the structures formed by the BMC-T protein PduB, its translationally truncated gene product PduB', and the vertex protein PduN. Together with PduA and PduJ, these proteins reflect ~75% of the Pdu BMC shell by copy number (Yang *et al.*, 2020).

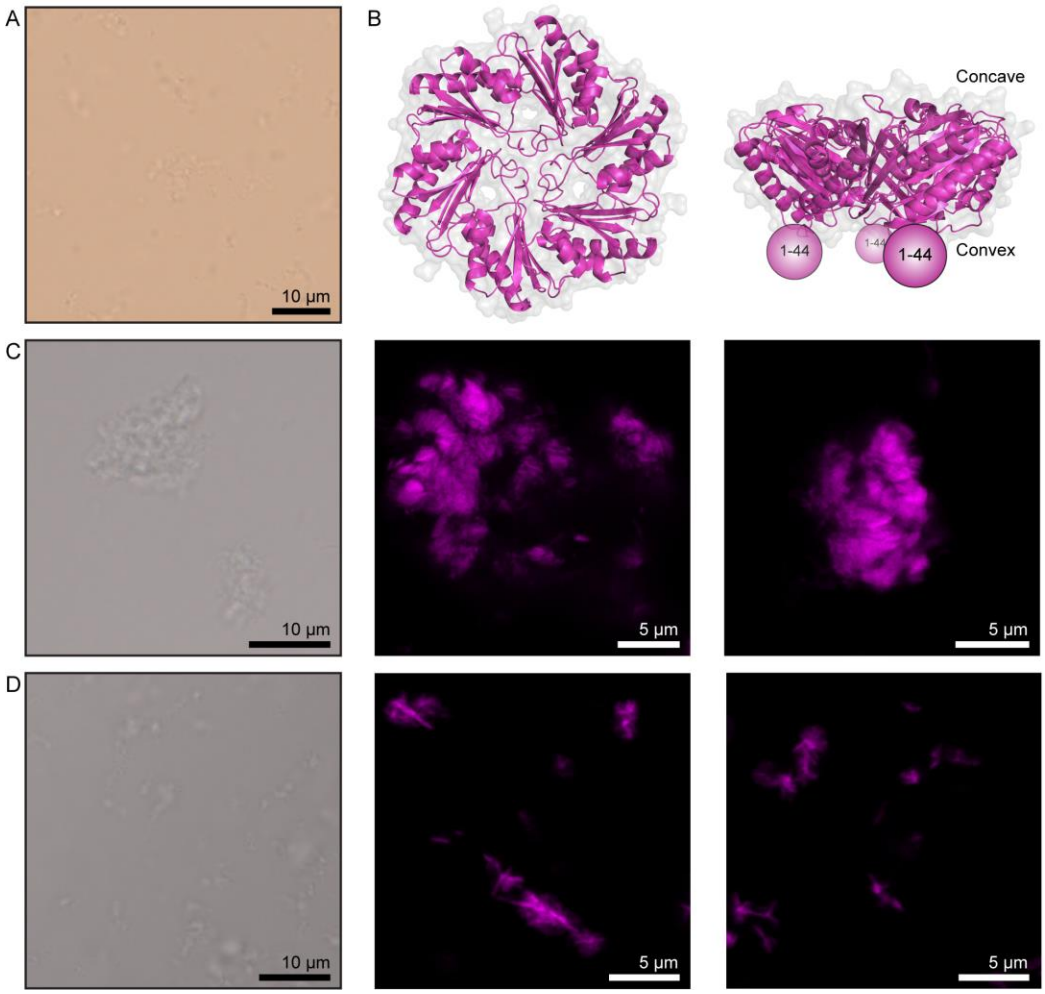


Figure 6.5: The structures formed by the shell proteins PduBB'N. (A) PduN forms general aggregates that are observed with brightfield microscopy. (B) PduB and PduB' differ on the basis of an additional N-terminal extension which PduB' lacks. This extension lays on the convex, interior surface of a Pdu BMC. Structure shown is a homology model derived from 4FAY (Pang *et al.*, 2014). Both PduB (C) and PduB' (D) form protein clusters. The clusters of PduB are much larger than those of PduB'.

Understanding the structures they form may help elucidate their roles in forming the complete Pdu BMC shell.

PduN was purified without a fluorescent tag, so its ability to aggregate was investigated solely by brightfield microscopy. Under these conditions, PduN forms general protein aggregates of various sizes (Figure 6.5a). This observation is unsurprising, as PduN lacks the hexagonal arrangement required for the facet assembly seen in the cases of both PduA and PduJ. Accordingly, PduN can interact with itself

and “assemble”, but these assemblies are likely nonsensical without other shell components.

PduB and PduB’ were both investigated for their ability to form higher-order structures with confocal microscopy. PduB contains an extra 37 residues along its N-terminus compared to PduB’ (Figure 6.5b). This terminal extension has previously been implicated in anchoring the shell to the enzymatic core (Lehman *et al.*, 2017) which is in agreement with our observations of it laying on the convex surface reported in Chapter 4. When visualized with a confocal microscope, PduB appears as extremely large, globular aggregates (>5 μm in diameter, Figure 6.5c). This is in contrast to PduB’ which presents as much smaller aggregates and, in some instances, elongated filaments connecting aggregates (Figure 6.5d). This is in line with previous reports that PduB forms nanotubes (Uddin *et al.*, 2018) although these were reported for the full-length PduB. Regardless, the presence of the N-terminus appears to influence the aggregation propensity of PduB. As such, the N-terminus may be an aggregation factor that interacts with likewise aggregated encapsulated cargo. This idea is further explored in Chapter 6.2.7.

Subsection 6.2.5: Shell Protein Super-Structures are Dynamic

Now that the individual phenotypes of several purified shell proteins were established, we next wanted to establish how combinations of shell proteins influence one another. We began by asking if PduA nanotubes will exchange components. We reasoned that if we equilibrated two sets of PduA, each with a different fluorescent label, that confocal microscopy would be able to establish their colocalization or segregation. We found that separate sets of labeled PduA nanotubes will colocalize

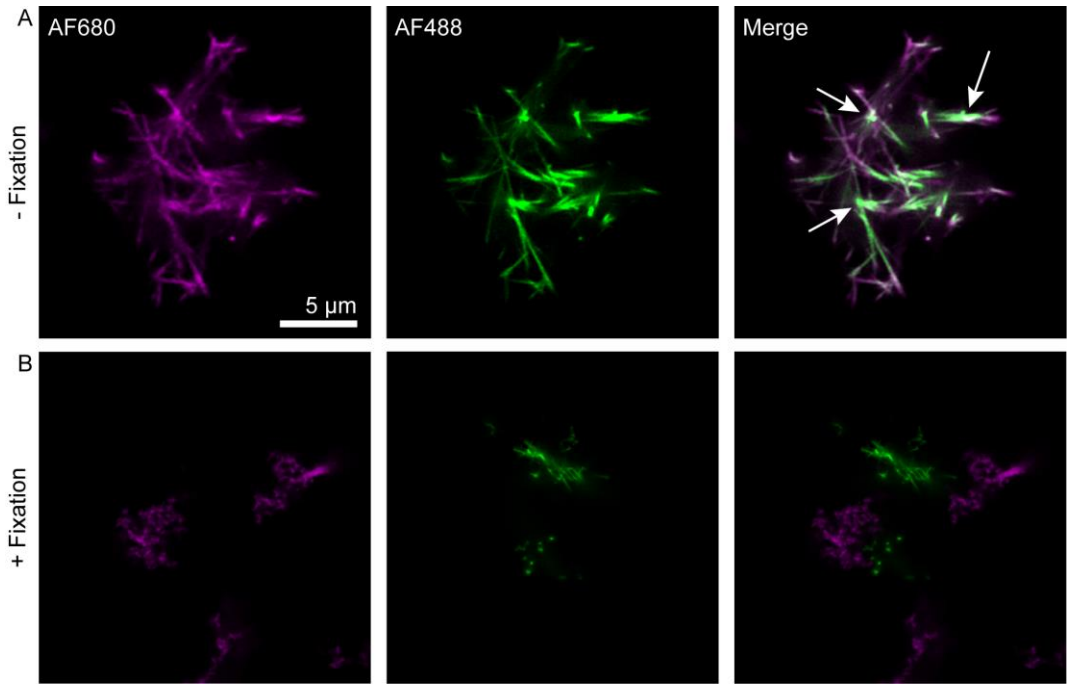


Figure 6.6: Differently labeled PduA nanotubes will intermix. (A) PduA nanotubes were prepared labeled with either AF488 or AF680 then mixed and allowed to equilibrate. Nanotubes will colocalize. This colocalization is especially pronounced in specific clusters (arrows). (B) Nanotubes were prepared similarly as in (A) but were fixed with 2% PFA prior to equilibration. Structures do not colocalize, but rather remain distinct.

heavily with equilibration (Figure 6.6a). However, the colocalization pattern is spotty - some regions of the structures are more heavily biased for green or purple signals. This may indicate that equilibration preferentially occurs at specific sites and that longer incubation times would result in a more even colocalization distribution. As a control, PduA nanotubes that were fixed with 2% PFA prior to equilibration remained as distinct entities (Figure 6.6b). These results indicate that the observed nanotubes are not static objects. Rather, they assemble and disassemble continuously, exchanging parts at the sites of assembly. While the exact positions of this are unclear, the micrographs seem to indicate that colocalization is greatest at intersectional positions (Figure 6.6a, arrows). These may be analogous to the nucleation-like positions revealed by TEM (Figure 6.2b) highlighting that these are potential sites of assembly.

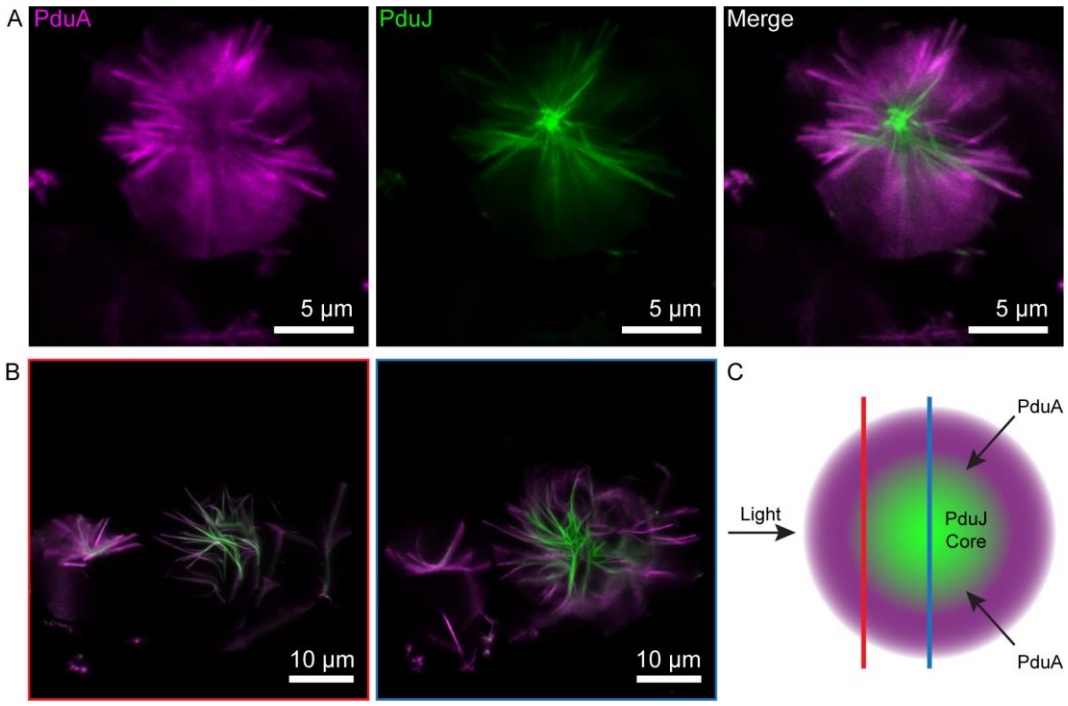


Figure 6.7: Differently labeled PduA and PduJ structure will intermix. (A) PduA was labeled with AF680 and allowed to equilibrate with AF488-labeled PduJ. PduA and PduJ will colocalize to the same rosette but remain somewhat segregated. (B) Changing the depth of field of a rosette reveals different extents of colocalization depending on the focal plane. (C) A model for the rosettes, wherein there is a PduJ core that PduA attempts to infiltrate.

These results establish that shell proteins of the same type equilibrate with each other to form new structures. We next wanted to ask if this likewise applies to combinations of different shell proteins, which can be differentiated under confocal if they contain different labels. We first investigated this by asking if the BMC-H proteins PduA and PduJ will interact and equilibrate. When mixed, PduA and PduJ still establish independent nanotube (Figure 6.7a, bottom) and rosette-like structures (center). However, these structures can display intense segregation of PduA and PduJ components (Figure 6.7a, right panel). This suggests that PduA and PduJ do colocalize and build off of one another. For example, rosette structures, observed solely in PduJ samples, exhibit penetration by PduA which is revealed when the rosettes are scanned at different depths of field (Figure 6.7b). Images taken at the peripherals of the rosettes

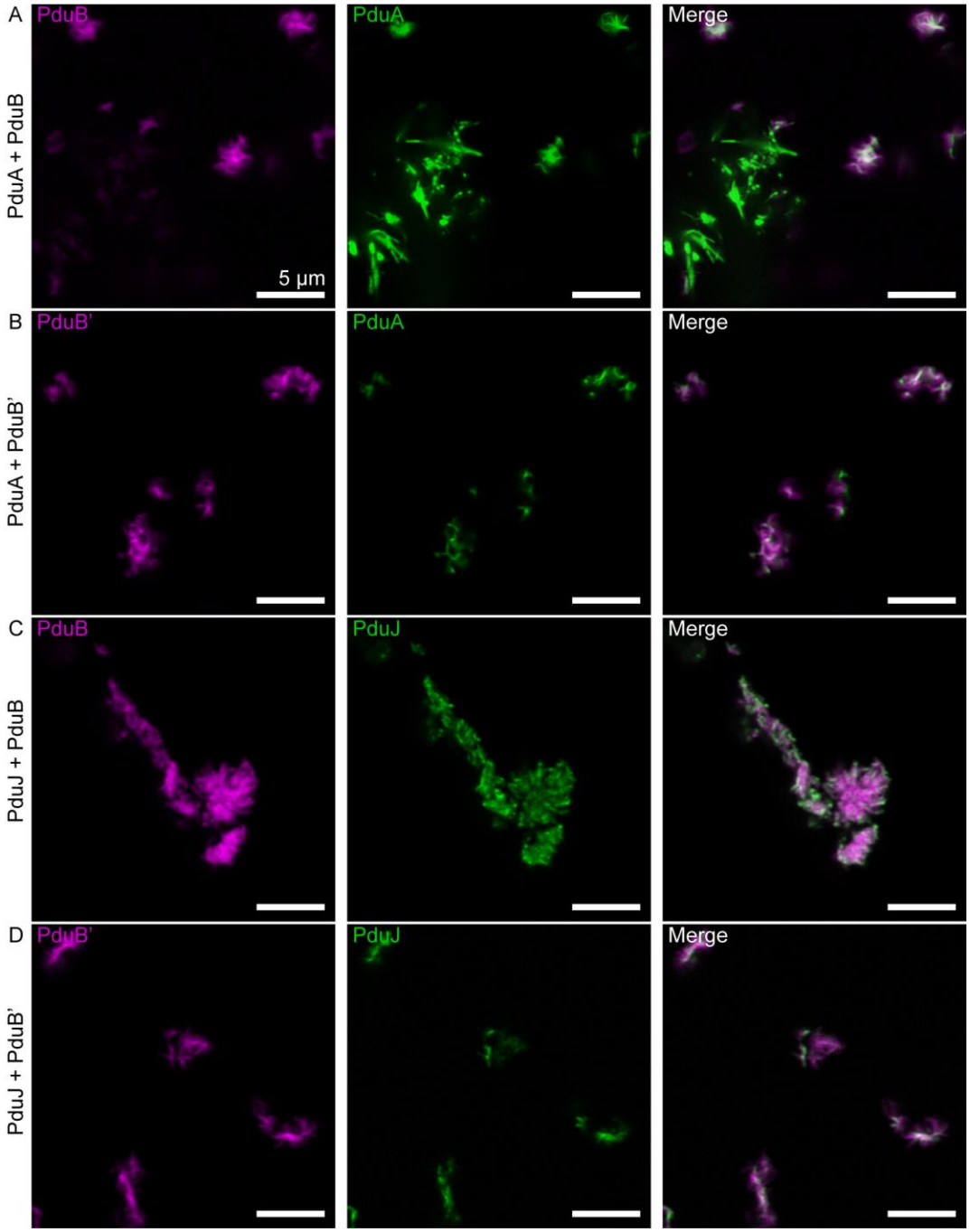


Figure 6.8: The effects of equilibrating BMC-H proteins with BMC-T. (A) PduA mixed with PduB. (B) PduA mixed with PduB'. (C) PduJ mixed with PduB. (D) PduJ mixed with PduB'. All cases show that PduBB' are unable to infiltrate PduA or PduJ structures, but PduA and PduJ can colocalize with PduB and PduB' structures. All scale bars are 5 µm.

show higher colocalization (Figure 6.7b, c; red lines) while images taken deeper into the rosettes show a green PduJ core that nearly entirely excluded PduA (Figure 6.7b, c; blue lines). Since the rosettes are typically assigned to be from PduJ, the retention of

that phenotype despite heavy PduA penetration suggests that these structures, while dynamic, maintain a "memory" of their original conformations.

Finally, we wanted to establish how PduA and PduJ each interact with PduB and PduB'. All mixtures contain distinct PduAJ nanotubes and PduBB' aggregates (Figure 6.8). Colocalization, however, can only be observed at both the PduB and PduB' aggregates - not at the site of PduA or PduJ nanotubes. This suggests that both PduB and PduB', while having different self-aggregation propensities, behave similarly when it comes to assembling with other shell proteins. Namely, PduB and PduB' appear unable to incorporate themselves with the PduA or PduJ nanotube architecture (exemplified in Figure 6.8a). In contrast, PduA or PduJ can generally be associated with PduB or PduB' aggregates. These observations suggest that PduBB' may be excluded from nanotubes due to the extremely high curvature associated with these structures, especially when compared to a full BMC shell.

Subsection 6.2.6: Shell Protein Assemblies Disrupt Native Pdu BMCs

Native Pdu BMCs can be regarded as simply a collection of different shell proteins with a specific structure. Since we established that different shell protein phenotypes can intermix and associate with one another, we hypothesize that this may equally apply to full Pdu BMCs being mixed with purified shell proteins. This experiment would also help further contextualize the results from Chapter 5, where it was found that uniquely labeled Pdu BMCs will cross-correlate when equilibrated (Figure 5.6). We hypothesize that this is mostly owing to a dynamic exchange of shell proteins, but that was not assessed at the time.

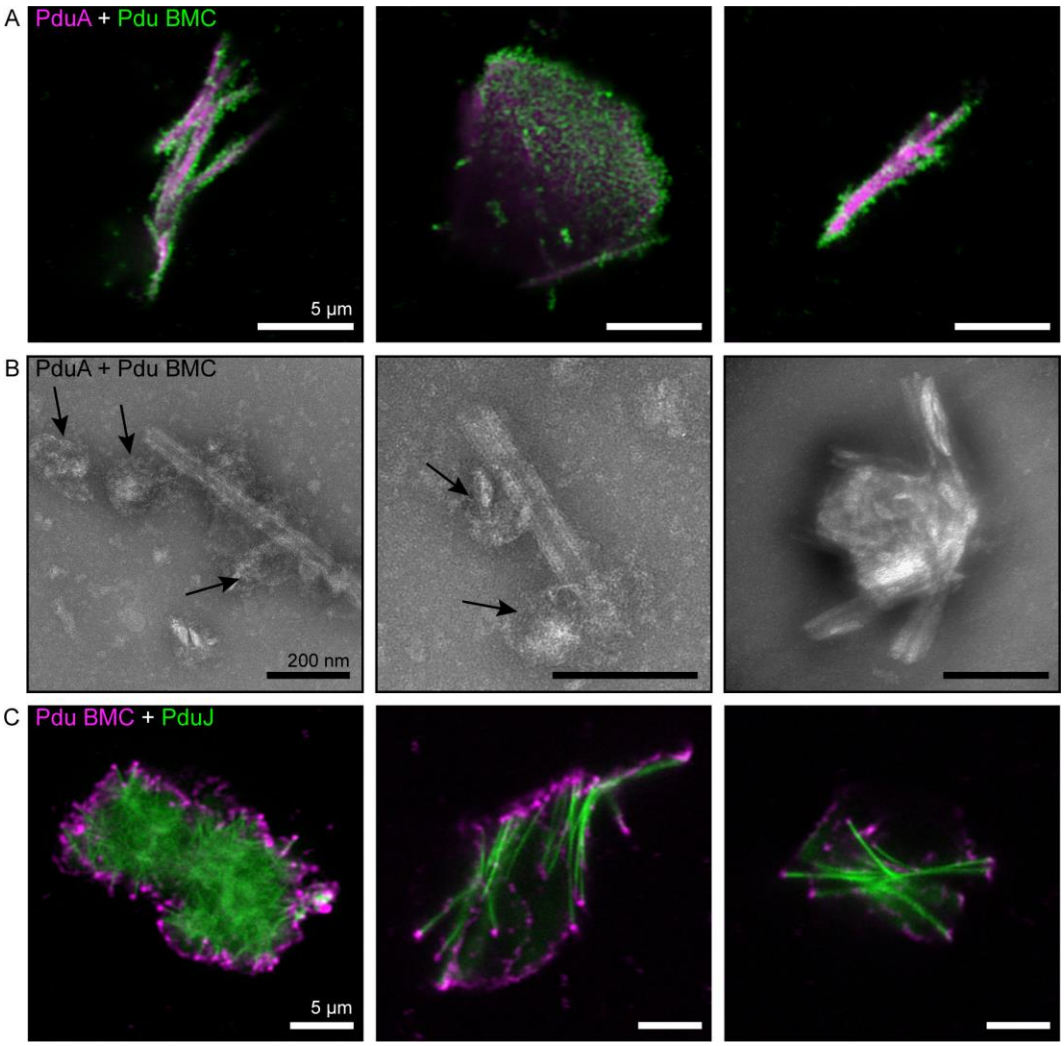


Figure 6.9: Pdu BMCs are disrupted by a shell super-structures. (A) PduA nanotubes and Pdu BMC were equilibrated and imaged with confocal microscopy. Scale bars, 5 µm. (B) Samples from (A) were imaged with TEM. BMCs (arrows) stick to and encompass PduA nanotubes. Scale bars, 200 nm. (C) PduJ was equilibrated with Pdu BMCs. Scale bars, 5 µm.

We began by preparing a solution of fluorescently labeled Pdu BMCs, as outlined in Chapter 5. These BMCs were mixed with and allowed to equilibrate with PduA labeled with a different fluorophore. Confocal analysis shows that these two components can highly colocalize with one another (Figure 6.9a). Specifically, Pdu BMCs appear to dot along PduA architectures, both nanotubes and sheets. This may suggest that Pdu BMCs are directly binding the PduA structures. We turned to TEM to elucidate this interaction more directly (Figure 6.9b). TEM shows that PduA nanotubes

do indeed independently exist and that Pdu BMCs (Figure 6.9b, arrows) associate with and bleed into the nanotubes. This latter point is established by the envelopment of the PduA structures in amorphous protein debris (Figure 6.9b, left panel). These results together indicate (1) Pdu BMCs do exchange, and are affected by, exterior shell components and that (2) Pdu BMC morphology is dictated by a delicate balance of shell protein abundances. The shell proteins themselves are further implicated in this process, as similar results were found with Pdu BMCs equilibrated with PduJ (Figure 6.9c). The Pdu BMCs appear to operate much like PduA, where the protein components attempt to infiltrate the larger PduJ structures. These results further contextualize our findings presented in Chapter 5, that Pdu BMCs are a dynamic system that is actively influenced by its exterior environment.

Subsection 6.2.7: Synthetic EPs have Preferred Shell Binding Partners

Confocal microscopy, to this point, has proven to be a powerful tool to visualize how Pdu shell proteins behave in solution. However, we believed this technique opened an opportunity to investigate how shell proteins interact with EPs *in vitro*. Previously established dogma states that EPs are the domains responsible for interacting with shell proteins to encourage cargo packaging within the shell. Specifically for the Pdu BMC system, the EP for the cargo protein PduP can bind PduA and PduJ (although this link is weaker) while the EP on PduD was shown to bind PduA (Fan *et al.*, 2012). The C-terminus was specifically implicated in hosting this interaction, which we establish operates as a disordered segment that can exist on both sides of the protein shell (see Chapter 4). PduP has also been linked to interacting with another BMC-H, PduK (Lawrence *et al.*, 2014). Further, the N-terminus of PduB has likewise been speculated

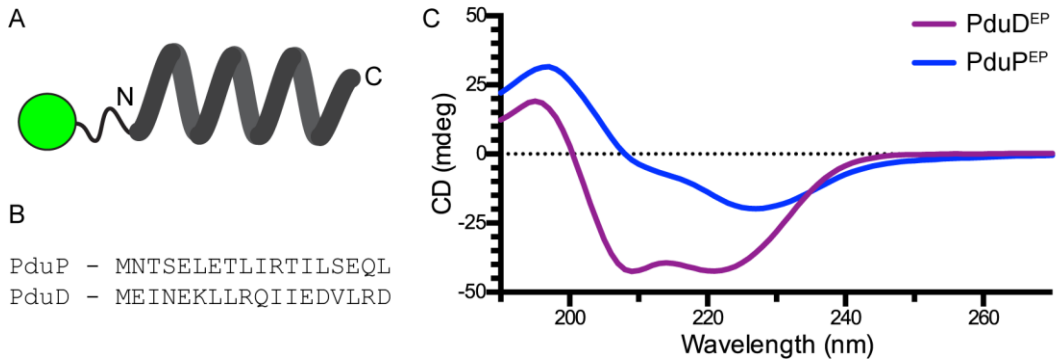


Figure 6.10: The encapsulation peptides (EPs) used in this work. (A) Cartoon schematic of the synthetic derived EPs tagged with fluorescein (green ball). (B) Sequence comparison of the EPs for PduP and PduD. (C) Circular dichroism of the selected encapsulation peptides. Both show α-helical properties.

to act as an anchor between the inner cargo and shell (Lehman *et al.*, 2017). Our crosslinking results agree with this finding, where we found a relatively numerous and diverse swath of cargo crosslinks to PduB and that its N-terminus faces luminally (see Chapter 4). Together this paints a messy picture where it is hard to establish what shell proteins fill what role in cargo envelopment. We believed binding assays would not be appropriate to uncover interaction partners given that shell proteins self-assemble into a range of higher-order structures. Instead, we purchased a synthetically produced EP from the cargo proteins PduP and PduD with an N-terminal fluorescein moiety to use as a reporter with confocal microscopy (Figure 6.10a).

The two EPs for PduP and PduD ($PduP^{EP}$ and $PduD^{EP}$, respectively), were purchased as both fluorescent and non-fluorescent derivatives (Figure 6.10b). Both non-fluorescent EPs were analyzed with circular dichroism to confirm their α-helicity in deionized water (Figure 6.10c). $PduD^{EP}$ forms a quintessential helical structure under these conditions, with the characteristic double minima at 208 and 222 nm (Greenfield, 2009). The curve for $PduP^{EP}$ is not as inspiring as that for $PduD^{EP}$ but still maintains a general α-helical character. This may be a result of insolubility, where the $PduP^{EP}$

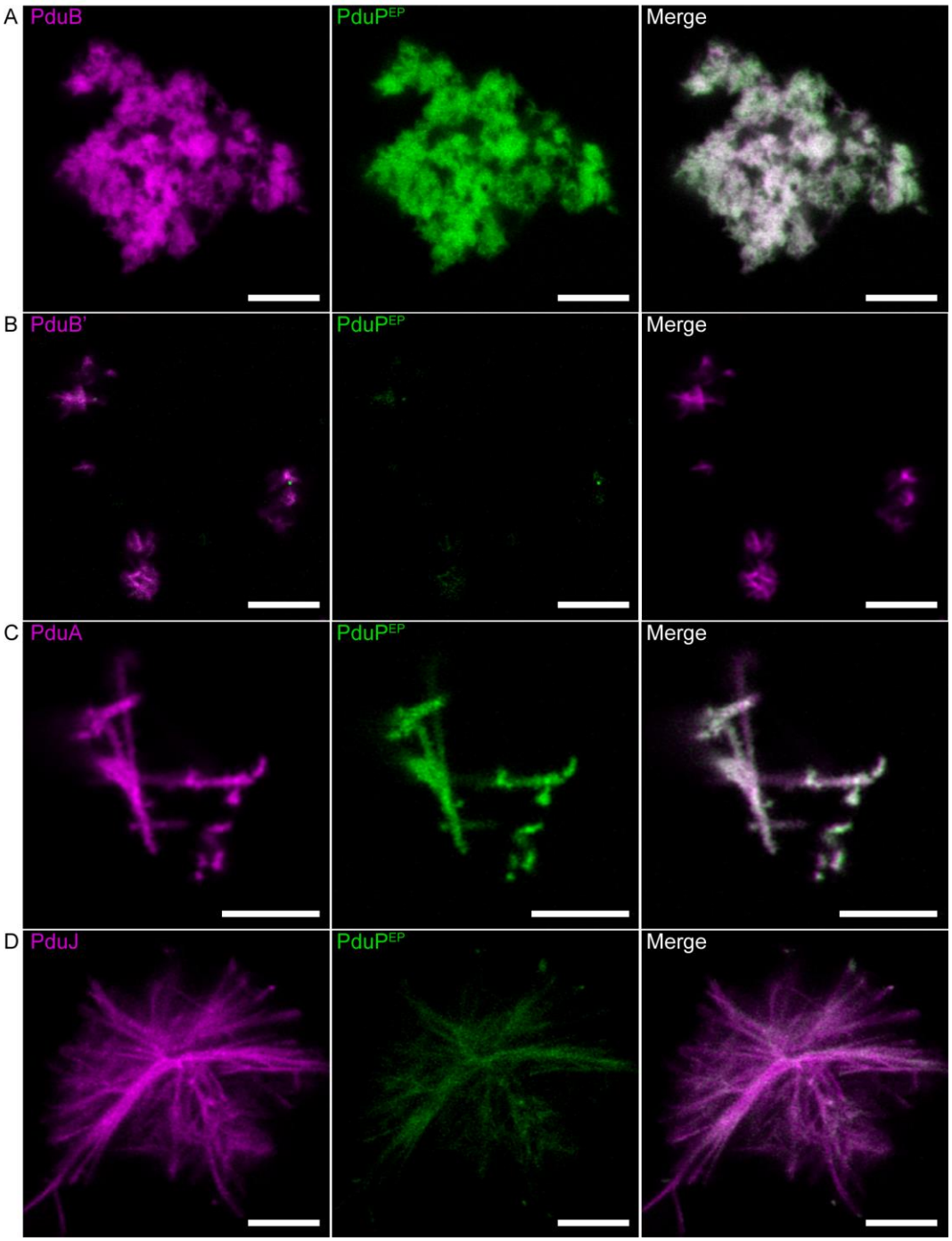


Figure 6.11: Association of the PduP encapsulation peptide with shell protein superstructures. Purified, synthetic, fluorescein-tagged encapsulation peptides (EP) for PduP ($PduP^{EP}$) was equilibrated with either (A) PduB, (B) PduB', (C) PduA, or (D) PduJ. In all instances, samples were equilibrated for 30 minutes on ice before being imaged by laser scanning confocal microscopy. Both EPs behaved similarly; a greater degree of colocalization was observed between the EPs and PduB and PduA. All scale bars are 5 µm.

solution was noticeably opaquer than the $PduD^{EP}$ solution in deionized water. In contrast, both fluorescein-tagged EPs used in subsequent assays were soluble when

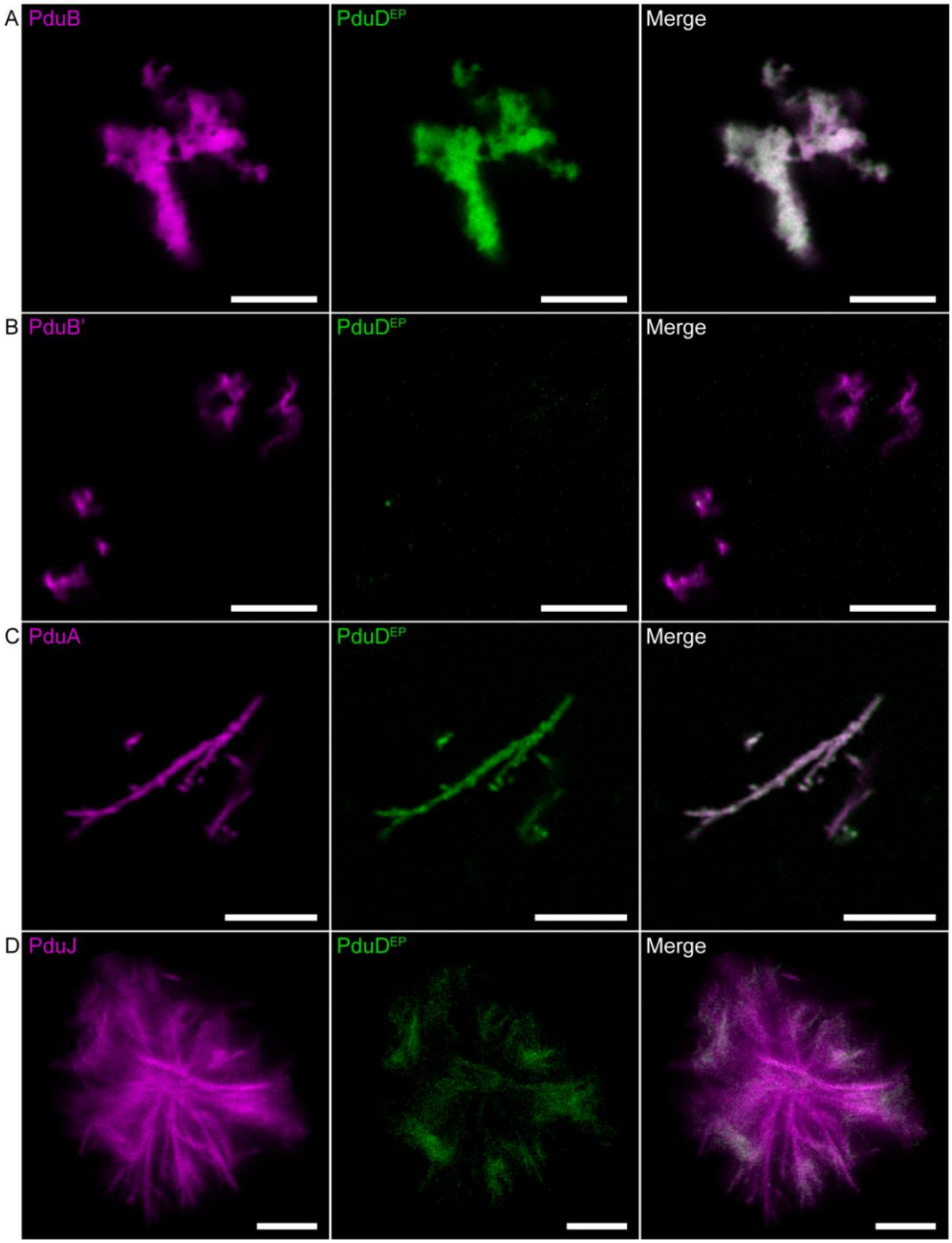


Figure 6.12: Association of the PduD encapsulation peptide with shell protein superstructures. Purified, synthetic, fluorescein-tagged encapsulation peptides (EP) for PduD ($PduD^{EP}$) was equilibrated with either (A) PduB, (B) PduB', (C) PduA, or (D) PduJ. In all instances, samples were equilibrated for 30 minutes on ice before being imaged by laser scanning confocal microscopy. Both EPs behaved similarly; a greater degree of colocalization was observed between the EPs and PduB and PduA. All scale bars are 5 μm.

dissolved in 10 mM Tris-HCl pH 8.0 - a buffer condition which would have precluded

data collection at lower wavelengths due to the absorption of Cl anions. These results indicate that the EPs do forms the anticipated α-helical structures observed previously (Lawrence *et al.*, 2014; Fan *et al.*, 2012).

From here, we moved on to investigate the ability of $PduP^{EP}$ to bind the shell proteins PduABB'J. In this experiment, shell proteins (40 μM) were mixed with a limited concentration of $PduP^{EP}$ (4 μM) and equilibrated on ice for 30 minutes prior to interrogation with a confocal microscope. The results reveal that $PduP^{EP}$ does prefer specific shell proteins (Figure 6.11). First, $PduP^{EP}$ has a clear preference for PduB (Figure 6.11a) compared to PduB' (Figure 6.11b), presumably a result of the luminally facing N-terminal extension on PduB that PduB' lacks. $PduP^{EP}$ is also able to significantly colocalize to PduA nanotubes (Figure 6.11c) in accordance with previous studies (Fan *et al.*, 2012). In contrast, the EP displayed a weakened ability to bind PduJ rosettes (Figure 6.11d). Thus, PduA and PduB may be the preferred targets for PduP while PduJ is tuned for weaker interactions. In all cases $PduP^{EP}$ did not influence the morphology of shell protein structures. $PduP^{EP}$ also did not form observable structures on its own (not shown). These results agree with the literature and represent the first direct instance of shell-EP interactions. This experiment was repeated for $PduD^{EP}$ and yielded similar results (Figure 6.12).

Section 6.3: Conclusions

Subsection 6.3.1: Pdu BMCs are Dynamic Entities

The results shown here further emphasize several of the conclusions drawn in Chapter 5 about BMC dynamics. Namely, BMCs may exist on a continuum where

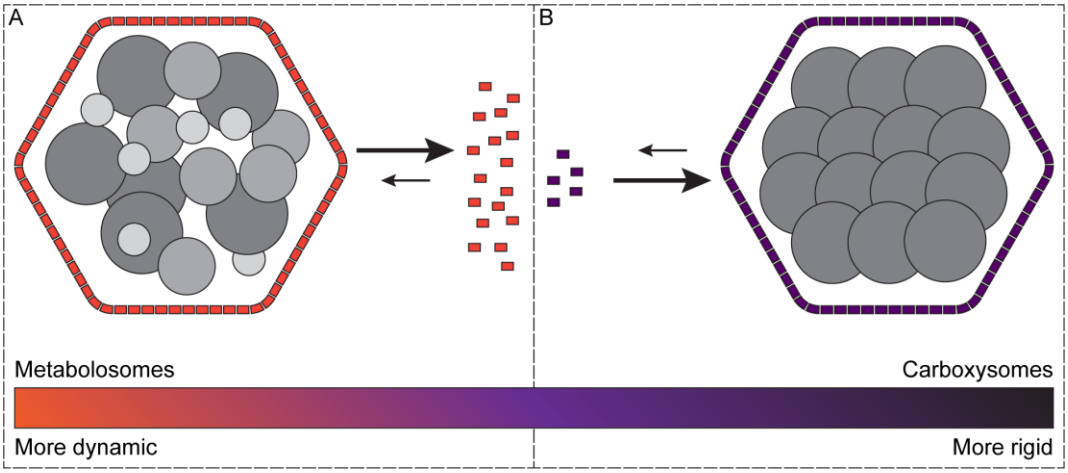


Figure 6.13: Different BMCs may exist along a continuum of dynamic properties. (A) Metabolosomes, like the Pdu BMC, may represent a more dynamic BMC variant. These BMCs may exist in an equilibrium with shell proteins, which dictate the morphology and assembly state of the overall BMC. (B) In contrast, carboxysomes may lie on the other side of the dynamism spectrum. These BMCs are extremely regular and rigid in their composition.

structural dynamism is an evolved adaptation required for specific compartments (Figure 6.13). Pdu BMCs likely represents the more dynamic end of this spectrum, while carboxysomes perhaps exists on the opposing side. The work presented in this Chapter expanded beyond that of Chapter 5 by specifically implicating the protein shells in this process. In our model, BMCs exist in an equilibrium between "complete" BMCs and individual shell components in solution. The equilibrium point may be different for every BMC. This is specifically revealed in our data which suggests that the structure formed by shell proteins can exchange between themselves. Meanwhile, Pdu BMCs essentially act as a collection of shell proteins and can participate in this exchange and become morphologically disrupted. This latter point suggests that BMCs are not just a delicate balance of different proteins, but also that the shell itself mediates the morphology of the BMC rather than the cargo.

The speculation of a sliding dynamism spectrum has implications in synthetic systems. One may choose different BMC models to serve different purposes. For

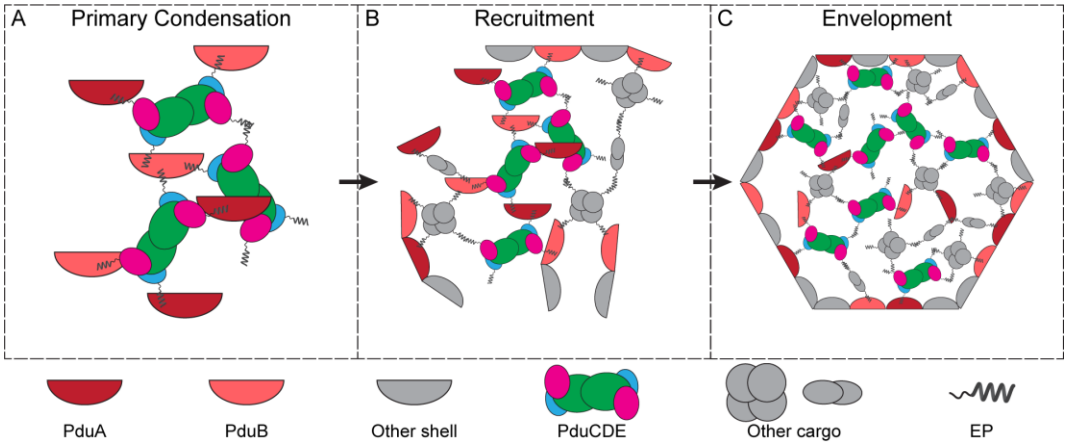


Figure 6.14: EP binding results inform a concomitant model for Pdu BMC biogenesis. (A) The shell protein PduAB are expressed first and begin to condensate with the signature complex, PduCDE. (B) The primary condensate grows by recruiting other protein factors. Other EP-tagged cargo continues to extend the cargo network while shell proteins begin to scaffold along the exterior. (C) The cargo network finishes polymerization once the protein shell completely envelopes it within its interior

instance, a rigid carboxysome model may be suitable to host chemistries that are sensitive to oxygen, required for the hosts survival, or that produces intermediates far more toxic than the typical aldehydes found in metabolosomes. Similarly, carboxysome models may resist the mixing of components, observed for Eut and Pdu BMCs in *S. enterica* (Sturms *et al.*, 2015), and assist in design of orthogonal BMCs which can be co-expressed in a single bacterium. In contrast, a metabolosome like the Pdu BMC may be a good candidate if one wished to establish an environmentally responsive BMC system that arises in response to stimuli then disassembles when the signal dissipates. This model can also help explain the disassembly of metabolosomes, an often-neglected side of the coin. If different BMC are dictated by different extents of equilibrium with shell components, then terminating translation of shell proteins would result in turnover of the BMC. Here, one can imagine a BMC that appears when the desired substrate is present and disassembles when it is no longer needed. There is no

one definitively better BMC chassis. Rather, the choice would depend on the project goals.

Subsection 6.3.2: EP Preferences Hint at a Concomitant Biogenesis Model

Our data also established which shell proteins are the binding partners of the for the cargo protein PduP. Our results do confirm that EPs interact with shell proteins, in addition to other EPs (see Chapter 3), albeit with caveats. It is curious to note that the preferred binding partners of $PduP^{EP}$, PduAB, are also the first two gene encoded by the pdu operon (Figure 2.2a). This detail may betray a critical detail important in the biogenesis of the Pdu BMC. Specifically, PduAB co-aggregate with cargo and subsequently seed shell formation concomitant with cargo condensation (Figure 6.14). In this model, PduA and PduB are co-expressed and aggregate with the signature complex PduCDE in an EP-dependent manner (Figure 6.14a). Other shell and cargo proteins are then recruited to the primary condensate (Figure 6.14b) prior to complete envelopment of the cargo by the protein shell (Figure 6.14c). Here, EPs help aggregate cargo and the shell proteins PduAB. PduAB and then seed shell formation in tandem with the growth of the overall BMC aggregate. This model is entirely consistent with the other data presented in this dissertation as well as the sum of observations throughout the literature.

Section 6.4: Methods and Materials

Subsection 6.4.1: Purification of Shell Proteins

All shell proteins were purified using the same set of conditions. Vectors encoding for the shell proteins PduABB'JN were transformed into T7 Express *E. coli*.

Select colonies were then restreaked and subsequently made into glycerol stocks. For protein production, strains were streaked from glycerol stocks onto LB agar + carbenicillin . Single colonies were then used to produce overnight starter cultures. The next day, starter cultures were used to inoculate 500 mL of 2xYT broth supplemented with carbenicillin and grown to at OD600 of 0.6-0.8. Once at the appropriate density, IPTG was added to 0.5 mM to induce protein expression. Cultures were further incubated at 37°C for 3 hours before pelleted pelleted and frozen at -80°C. Pellets were then thawed, resuspended in 40 mL Buffer A (50 mM Tris-HCl pH 8.0, 1 M NaCl, 5 mM DTT, 10 mM imidazole) supplemented with 5 mM $MgCl_2$, 1 U/mL Baseline-Zero DNase (Lucigen), and 0.5 mM PMSF. Cells were then lysed with sonication and insoluble matter was pelleted by centrifugation at 13000xg for 20 minutes. Soluble matter was then incubated with 1 mL of nickel beads pre-equilibrated in Buffer A for 30 minutes on ice. The flowthrough was then drained, and beads were washed with 5 mL of Buffer A and then 5 mL of Buffer B (50 mM Tris-HCl pH 7.4, 1 M NaCl, 35 mM imidazole). This second wash removes non-specifically bound proteins and equilibrates bound samples for fluorescent labeling. The appropriate fluorophore was dissolved in 1 mL of Buffer B and added to the bead bed where the reaction was allowed to proceed for 2 hours at 4°C. The excess fluorophore was then drained off and the beads were washed again with 2x5 mL of Buffer B supplemented with 5 mM DTT. Labeled proteins were then eluted in Buffer C (50 mM Tris-HCl pH 8.0, 1 M NaCl, 5 mM DTT, 250 mM imidazole) and dialyzed into storage buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM DTT). Aliquots were dispensed and flash frozen for future use.

Subsection 6.4.2: Transmission Electron Microscopy

Shell protein samples were thawed diluted in labeling buffer (20 mM sodium phosphate pH 7.4, 50 mM KCl, 5 mM $MgCl_2$) to 20 μM and 5 μL was applied to a formvar coated copper grid, 200 mesh (FCF200-CU, Electron Microscope Sciences). Samples were air dried for 15 minutes and excess was wicked off. Salt was removed by 3x5 μL additions of deionized water before staining with 3 μL of 2% uranyl acetate for 10 seconds. Samples were imaged using a JEOL 100CXII electron microscope.

Subsection 6.4.3: Laser Scanning Confocal Microscopy

All samples were imaged on a Zeiss LSM 980 Airyscan at room temperature using a 63x objective with immersion oil. Briefly, 20 μL of sample solution (typically 20 μM protein) was pipetted directly onto a #1.5 coverslip (Fisherbrand, 12544E). Initial focus was achieved by focusing on the edge of the sample droplet under transmission mode before switching to fluorescence and focusing down to the coverslip surface. Structures are plentiful and bright enough to allow live calibration of the Airyscan detector in super resolution mode. The fastest possible image scan time was always used with a minimum pixel scan time of 1.00 μs. All captured and reported images had standard Airyscan processing applied. Captured, unedited images were all analyzed in ImageJ.

Subsection 6.4.4: Circular Dichroism of EPs

Synthetically derived EPs were purchased from Genscript. Dried peptides were dissolved in ice-cold deionized water to a concentration of 500 μM and stored on ice until analyzed. For analysis, peptides were diluted to 50 μM in deionized water in a 1

mM quartz cuvette. Measurements were taken on a Jasco J-810 circular dichroism spectrometer from 180 to 300 nm in 1 nm increments with a scan speed of 100 nm/min in triplicate with moderate smoothing.

Chapter 7: Concluding Remarks and Future Outlooks

Section 7.1: Summary of Work Completed

Subsection 7.1.1: Interrogating the Pdu BMC Using Non-Traditional Approaches

This dissertation sought to explore the structure of the model Pdu BMC from *S. enterica*. The structure of the Pdu BMC has been interrogated since the late 1990's but has largely remained enigmatic. "Structure" in this regard is a broad term which encompasses the composition, abundances, behavior, interactions, and relative positioning of the individual protein elements. The Pdu BMC has remained so mysterious because of its sheer size and complexity compared to other cellular substructures. While its structure has been elusive, its pull has only become more alluring with time. By merely existing, such a complex entity demands to have its secrets divulged - and researchers rise to the challenge. However, the mystery is not simply unraveled for its own sake. The very details that evade our eye are inexorably linked to the future of these promising bioplatforms.

Many questions needed to be answered before BMC biotechnology can come of age. For instance, which side of the protein shell faces the inside? What is the internal substructure like? Why are metabolosomes so heterogeneous? This dissertation begins to answer these questions, and more, by using a suite of non-traditional approaches to examine the structural elements of the model Pdu BMC.

In Chapter 2, we introduce methods to probe the Pdu BMC with fluorescent tags. This probing method yields rapid insight into the relative locations of proteins.

Shell proteins were confirmed to label faster than interior, protected, cargo proteins. Surprisingly, the fluorescent tags could access the luminal space of intact Pdu BMCs - contrary to the traditional thought where BMC shells are rigid barriers. The BMCs were later confirmed to welcome these chemical modifications as they were not perturbed by their presence.

Chapter 3 and 4 took the ability to access luminal information and ran with it. Specifically, we turned to a structural mass spectrometry approach called crosslinking mass spectrometry to covalently couple interacting proteins. This coupling tells one both "who" is interacting and "where" that interaction is with residue-level resolution. The crosslinkers could indeed permeate the Pdu BMC and numerous cargo crosslinks were identified. These cargo crosslinks told a very different story than that in the literature, a story where EPs are more critical for cargo interactions than shell. Specifically, EP-tagged cargo form an intense subnetwork within the luminal space where interactions with the shell are largely excluded. Non-native cargos were found to accommodate similar interactions, thereby confirming the modularity and mode of action for EPs. Finally, the signature complex PduCDE was found to be a hub for several different piggybacking interactions for EP-lacking cargo.

Chapter 4 concerned crosslinks among the shell proteins. The crosslinks between the shell and cargo point towards the convex surface of the shell being luminal. Evidence from hydroxyl-radical footprinting further confirms the guilty verdict. Knowing which side faces which direction allows us to map the relative locations of several disordered shell protein extensions. Of particular interest is the N-terminus of PduB, a shell protein found in several cargo crosslinks, which we find faces luminally

to interact with cargo. PduB is further found to support two different sets of interactions with the hexamers PduA and PduJ which suggests that these homologs have unique roles in the assembly of the shell. Thanks to these findings, and modeling of the facet sizes, we were able to produce a model for the Pdu BMC shell. Importantly, this model may help to explain the physical basis of heterogeneity observed in Pdu BMCs.

Chapter 5 returned to a recurring theme - where relatively large probes can sneak through the protein shell. We hypothesize that this stems from the outer shell potentially being dynamic, flexible, and/or be the result of shell proteins assembling/disassembling actively on the shell. We initially find that cargo proteins can become less accessible when the shell is rigidized with a crosslinker, suggesting that the shell is responsible for the observed leakiness. Exogenously supplied shell protein PduA was also found to bind to the Pdu BMC *in vitro*, a result we speculate may indicate incorporation into the shell itself. We finally turned to confocal microscopy to analyze fluorescent tagged BMCs *in vitro*. By modulating the extent of fixation and equilibration, we could establish the protein elements of the Pdu BMC will exchange and colocalize over time. We ultimately hypothesize that the BMCs are in equilibrium with their surroundings, ultimately shedding tagged protein elements which get incorporated into other BMCs.

If confocal microscopy worked with entities as small as BMCs, we reasoned that it could likewise serve as a unique means to study the higher-order structures produced by shell proteins. Indeed, we were able to visualize an array of structures, such as nanotubes, rosettes, aggregates, and flat sheets, formed by the shell proteins PduABB'J. These structures were found to be dynamic in solution; they would colocalize when

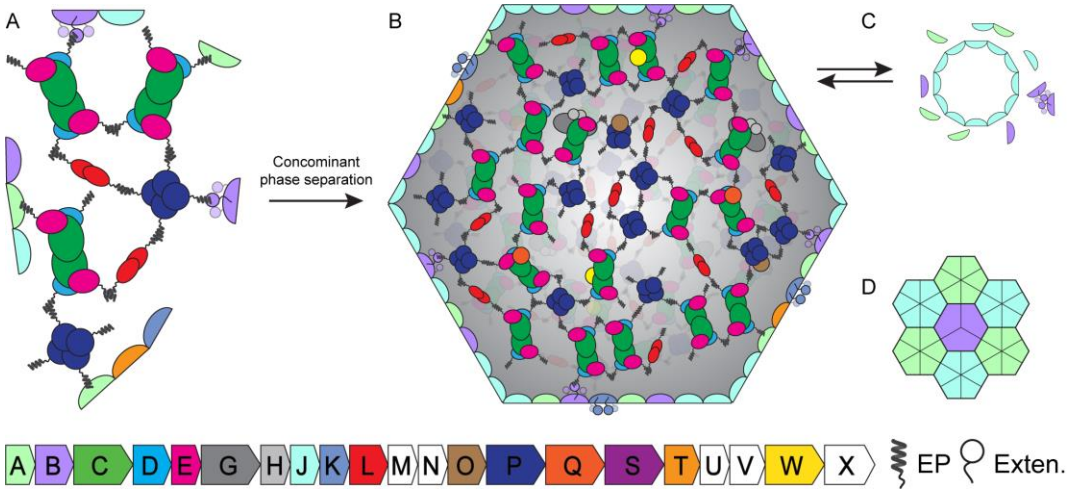


Figure 7.1: An integrated model for the biogenesis, assembly state, and behavior of the Pdu BMC. (A) Our data is consistent with a concomitant biogenesis pathway where shell and cargo proteins aggregate prior to complete shell envelopment. We speculate the initial aggregation is facilitated by a liquid-liquid phase separation mechanism. (B) A model for the fully assembled Pdu BMC. The protein shell has its convex surface facing inwards. Cargo proteins interact sparely with the protein shell, and largely with each other in a phase-separated network that extends throughout the luminal space. These interactions are dictated by encapsulation peptides. (C) The Pdu BMC exists in equilibrium with individual shell elements in solution. These elements can self-assemble into an array of structures like nanotubes. (D) The polyhedral facets of the Pdu BMC are mostly comprised of BMC-T proteins surrounded and interspaced by BMC-H.

equilibrated. As a polyhedral collection of shell proteins, Pdu BMCs themselves would also colocalize to PduA and PduJ structures. TEM confirmed this while also adding that Pdu BMC are not just binding but rather being disrupted by the shell protein superstructures. Further, the N-terminus of PduB is found to enhance its aggregation propensity and greatly influence its ability to bind EPs. PduA was also found to bind EPs while PduJ, in comparison, could do so only weakly. These results dictate that Pdu BMCs are dynamically controlled by a delicate balance of shell components. The shell proteins PduAB, we speculate, may also play a role in the early stages of biogenesis.

Subsection 7.1.2: Interpretations and Implications

The data reported herein allow us to propose an integrated model of the Pdu BMC that may generally apply to other BMC. This model accounts for the biogenesis,

assembly state, and solution behavior of the Pdu BMC (Figure 7.1). We theorize that the Pdu BMC follows a concomitant assembly pathway (Kerfeld *et al.*, 2018), in which shell and cargo components phase separate in tandem prior to complete shell envelopment (Figure 7.1a). This is supported by our data showing EPs interact with the first two proteins in the *pdu* operon, the shell proteins PduAB, but not latter encoded proteins like PduJ despite PduJ being more abundant (Yang *et al.*, 2020). PduAB being responsible for cargo interactions has been previously supported by a variety of *in vivo* knockout data (Fan *et al.*, 2012; Lehman *et al.*, 2017). This is akin to the proposed model for the α-carboxysome (Iancu *et al.*, 2010; Mohajerani *et al.*, 2021) despite being more related to the β-carboxysome (Kerfeld and Melnicki *et al.*, 2016). The differences may be partially explained by the lack of a dedicated scaffold protein in the Pdu BMC which drives the multivalent interactions between the cargo and the protein shell. Instead, the Pdu EPs themselves demonstrate emergent multivalency due to the oligomeric state of EP-tagged proteins, effectively acting as a scaffold. The initial recruitment of PduAB to the condensate helps further traffic other shell and cargo proteins. Eventually the shell proteins hit a critical mass and inflict a predestined size and curvature to the inner cargo. PduA and PduB may be required over PduB' and PduJ to tune the condensates maturation but fulfill similar roles in shell assembly.

Once assembled, the Pdu BMC acts as a shell-stabilized cargo phase separation (Figure 7.1b). Here, EP-tagged proteins make up an inner protein network that extends throughout the luminal space. The shell itself, while acting to stabilize the cargo, is also dynamic and in equilibrium with its environment (Figure 7.1c). Different BMCs may have this equilibrium tuned differently for different reasons (Figure 6.12).

Metabolosomes may represent a more environmentally responsive class of BMC, assembling and disassembling rapidly, while carboxysomes are more static and enduring. This would allow researchers to pick and choose different BMC shell chassis that can suit different functions.

Our crosslinking and footprinting data reveal that shell proteins present their concave surface outwards to interact with the cytosol, such as the case for the iron-sulfur cluster binding site on the C-terminus of PduK (Crowley *et al.*, 2010). This leaves the convex surface as responsible for luminal interactions and represents a designable surface to support new interactions. These data are consistent with models for small, empty BMCs such as the Hoch BMC (Sutter *et al.*, 2017). Our crosslinking data also reveals that the fundamental building blocks of the Pdu BMC where BMC-T are surrounded by and interspaced by BMC-H (Figure 7.1d). Again, this is like other small BMCs that have had their structures reported suggesting that this is a universal theme for their construction. Therefore, BMC shells are a reproducible architecture that is agnostic to luminal composition.

Section 7.2: Future Directions

This dissertation addresses critical gaps in our understanding concerning the construction of BMCs. However, some questions remain open. For example, our reliance on *in vitro* methods means that we were unable to capture cytosolic interactions between the BMC and the rest of the cell. Carboxysomes are understood to be influenced by a complex positioning system (Rilemma *et al.*, 2021), but such is not known about their metabolosome cousins. Understanding how the BMC interacts natively with the cytosol has wide ranging implications in their biogenesis, destruction,

and positioning. Further, this space may represent unrealized bioengineering potential to control these aspects.

Our data and interpretations are wholly consistent with and highly supportive of the role of liquid-liquid phase separation in the biogenesis of the Pdu BMC. Indeed, reading the literature can feel as if the field has been building gradually towards this conclusion for decades. However, the methods we used fundamentally cannot prove this point. Others may seek to explore this (as some recent preprints do, uncited here) and elucidate the roles of EPs, valency, shell proteins, and their termini in this process, if applicable. Such work has the potential to rewrite our understanding of BMCs and help usher in the next generation of synthetic BMC design paradigms.

For its issues, this work also introduces a variety of new themes that we hope others can build from in a constructive way. For instance, this work started with the simple concept of fluorescent tagging BMCs. It is easy to imagine that other, far more inventive, permutations of this concept exist. This can open the doors for treating BMC shells as a derivitizable biomaterial to aid in drug delivery, rigidize the shells, and adapt them for new functions.

The dynamic qualities of BMC systems will undoubtedly be a high-priority research target moving forward. While we used traditional confocal microscopy to explore this, there exists a clear avenue for techniques like super-resolution nanoscopy and fluorescence correlation spectroscopy to exploit. This work will be critical to understand the factors that influence/control BMC shell dynamics and whether the size of BMC shells can be tuned. In this context, work will unquestionably continue concerning the non-native structures generated by BMC shell proteins. Shell protein

superstructures may represent an untapped module for tunable architectures both *in vivo* and *in vitro*.

Appendices

Appendix A: All redundant protein-protein crosslinks identified

Crosslinks are tabulated and organized within what sample and dataset they originated from. The nomenclature C-C, S-C, and S-S refer to cargo-cargo, shell-cargo, and shell-shell crosslinked protein pairs.

DSSO (MS2)

ID	Score	Protein A	Position A	Protein B	Position B	Class
1	30.64	PduL	112	PduC	414	C-C
2	50.4	PduE	94	PduC	414	C-C
3	53.52	PduP	57	PduD	170	C-C
4	32.32	PduD	215	PduC	414	C-C
5	10.87	PduP	319	PduD	170	C-C
6	47.14	PduC	466	PduC	414	C-C
7	42.8	PduC	395	PduC	414	C-C
8	67.03	PduL	112	PduD	170	C-C
9	63.05	PduA	55	PduK	144	S-S
10	32.32	PduL	112	PduP	320	C-C
11	67.76	PduD	215	PduD	170	C-C
12	37.56	PduE	54	PduE	94	C-C
13	46.35	PduC	466	PduD	170	C-C
14	61.26	PduC	11	PduC	414	C-C
15	43.82	PduC	466	PduP	320	C-C
16	39.44	PduD	6	PduC	414	C-C
17	55.68	PduL	112	PduE	94	C-C
18	37.56	PduE	54	PduC	466	C-C
19	38.32	PduD	101	PduC	414	C-C
20	38	PduJ	0	PduK	144	S-S
21	36.87	PduA	0	PduK	144	S-S
22	54.4	PduP	319	PduC	466	C-C
23	38.19	PduD	101	PduO	182	C-C
24	56.09	PduD	6	PduD	170	C-C
25	94.67	PduA	0	PduK	156	S-S
26	110.63	PduD	215	PduL	112	C-C
27	35.11	PduD	101	PduD	170	C-C
28	137.5	PduD	215	PduC	466	C-C
29	10.87	PduD	109	PduD	170	C-C
30	59.8	PduD	101	PduP	57	C-C
31	65.04	PduC	395	PduC	466	C-C
32	153.61	PduE	150	PduE	94	C-C
33	60.4	PduA	86	PduA	90	S-S
34	106.84	PduJ	0	PduK	156	S-S
35	174.64	PduA	55	PduA	0	S-S
36	34.09	PduD	101	PduL	112	C-C

37	115.71	PduA	86	PduA	0	S-S
38	70.86	PduE	0	PduC	414	C-C
39	38.19	PduE	150	PduE	170	C-C
40	62.72	PduD	215	PduE	150	C-C
41	159.14	PduE	0	PduE	94	C-C
42	14.62	PduP	340	PduC	414	C-C
43	45.99	PduS	356	PduS	363	C-C
44	22.36	PduD	101	PduE	170	C-C
45	47.82	PduJ	0	PduD	170	S-C
46	174.85	PduB	138	PduA	55	S-S
47	86.34	PduJ	89	PduA	90	S-S
48	53.71	PduD	6	PduL	12	C-C
49	85.52	PduN	15	PduN	23	S-S
50	66.14	PduD	101	PduC	466	C-C
51	45.16	PduC	148	PduC	414	C-C
52	75.31	PduE	1	PduE	94	C-C
53	53.47	PduE	0	PduD	170	C-C
54	112.75	PduJ	89	PduJ	0	S-S
55	67.06	PduA	0	PduA	90	S-S
56	73.62	PduB	263	PduL	112	S-C
57	34.09	PduA	86	PduJ	0	S-S
58	106.42	PduB	263	PduD	215	S-C
59	75.82	PduC	11	PduD	109	C-C
60	37.35	PduD	0	PduD	0	C-C
61	104.64	PduE	0	PduL	12	C-C
62	41.96	PduD	109	PduE	170	C-C
63	55.6	PduE	0	PduP	320	C-C
64	65.44	PduP	0	PduC	414	C-C
65	34.08	PduP	340	PduP	57	C-C
66	153.52	PduE	0	PduC	466	C-C
67	56.01	PduE	0	PduP	57	C-C
68	38.32	PduE	0	PduC	255	C-C
69	56.55	PduP	340	PduL	112	C-C
70	34.81	PduE	0	PduG	463	C-C
71	64	PduP	0	PduD	170	C-C
72	38.32	PduC	494	PduC	414	C-C
73	126.24	PduE	150	PduE	0	C-C
74	35.49	PduP	340	PduC	466	C-C
75	62.84	PduD	24	PduP	57	C-C
76	53.47	PduJ	54	PduD	170	S-C
77	56.12	PduE	0	PduL	112	C-C
78	183.46	PduB	138	PduB	146	S-S
79	74.42	PduD	215	PduE	0	C-C
80	87.95	PduP	0	PduP	169	C-C

81	88.95	PduC	11	PduE	0	C-C
82	59.8	PduB	263	PduD	101	S-C
83	120.76	PduB	263	PduD	6	S-C
84	114.02	PduP	0	PduP	57	C-C
85	50.97	PduE	150	PduE	1	C-C
86	67.03	PduC	395	PduE	0	C-C
87	62.39	PduE	0	PduD	0	C-C
88	30.64	PduP	0	PduC	255	C-C
89	135.35	PduB	146	PduB	146	S-S
90	37.56	PduE	73	PduE	54	C-C
91	101.21	PduP	0	PduE	94	C-C
92	67.76	PduO	194	PduP	0	C-C
93	42.11	PduC	494	PduL	112	C-C
94	114.89	PduC	11	PduP	0	C-C
95	45.95	PduD	215	PduP	0	C-C
96	33.54	PduE	73	PduC	414	C-C
97	37.35	PduT	1	PduK	144	S-S
98	119.57	PduC	86	PduC	97	C-C
99	215.31	PduJ	0	PduA	0	S-S
100	49.35	PduE	73	PduD	170	C-C
101	42.72	PduC	458	PduD	170	C-C
102	40.43	PduE	133	PduL	112	C-C
103	140.21	PduP	0	PduL	12	C-C
104	83.61	PduP	0	PduC	466	C-C
105	66.99	PduE	133	PduE	94	C-C
106	135.35	PduE	133	PduD	215	C-C
107	58.98	PduE	73	PduE	94	C-C
108	46.02	PduD	101	PduE	0	C-C
109	54.14	PduE	73	PduP	319	C-C
110	74.74	PduE	73	PduP	57	C-C
111	55.57	PduE	73	PduC	11	C-C
112	60.94	PduC	458	PduG	463	C-C
113	64.73	PduJ	0	PduA	72	S-S
114	151.17	PduE	73	PduD	215	C-C
115	94.26	PduE	73	PduL	112	C-C
116	114.72	PduB	177	PduA	0	S-S
117	32.35	PduP	0	PduE	150	C-C
118	83.38	PduT	1	PduT	117	S-S
119	116.35	PduE	73	PduC	466	C-C
120	121.68	PduG	59	PduC	466	C-C
121	56.12	PduD	101	PduP	0	C-C
122	71.03	PduC	458	PduC	466	C-C
123	172.54	PduJ	0	PduJ	1	S-S
124	72.02	PduE	133	PduD	6	C-C

125	46.62	PduE	133	PduE	170	C-C
126	99.01	PduP	340	PduE	0	C-C
127	90.99	PduE	0	PduB	146	S-C
128	35.17	PduP	0	PduD	0	C-C
129	31.93	PduC	458	PduC	97	C-C
130	45.16	PduE	133	PduO	1	C-C
131	235.31	PduE	0	PduE	1	C-C
132	190.73	PduB	263	PduB	1	S-S
133	180.27	PduE	73	PduB	263	S-C
134	111.28	PduE	133	PduE	0	C-C
135	144.29	PduP	340	PduP	0	C-C
136	95.79	PduJ	54	PduJ	0	S-S
137	97.07	PduE	59	PduE	0	C-C
138	201.04	PduP	0	PduE	0	C-C
139	97.46	PduE	73	PduP	340	C-C
140	213.66	PduC	494	PduE	0	C-C
141	160.69	PduE	73	PduE	0	C-C
142	137.5	PduE	73	PduE	1	C-C
143	40.95	PduC	475	PduD	170	C-C
144	126.24	PduC	86	PduE	0	C-C
145	118.01	PduE	133	PduP	0	C-C
146	175.94	PduE	73	PduP	0	C-C
147	215.31	PduC	458	PduE	0	C-C
148	117.4	PduG	59	PduE	0	C-C
149	204.37	PduT	1	PduA	0	S-S
150	35.22	PduC	458	PduP	0	C-C
151	157.54	PduE	73	PduG	59	C-C
152	158.53	PduT	1	PduJ	0	S-S
153	53.47	PduB	1	PduM	0	S-C
154	121.96	PduJ	54	PduB	1	S-S
155	46.02	PduC	475	PduE	0	C-C

DSSO (MS2-MS3)

ID	Score	Protein A	Position A	Protein B	Position B	Class
1	145.92	PduK	156	PduK	144	S-S
2	119.62	PduD	170	PduP	320	C-C
3	281.93	PduE	94	PduD	170	C-C
4	114.86	PduC	395	PduC	414	C-C
5	175.51	PduL	112	PduD	170	C-C
6	268.9	PduA	55	PduK	144	S-S
7	223.07	PduE	54	PduE	94	C-C
8	226.99	PduC	466	PduD	170	C-C
9	189.73	PduC	466	PduG	255	C-C
10	146.01	PduA	86	PduK	144	S-S
11	144.28	PduC	466	PduP	320	C-C

12	171.26	PduE	54	PduL	112	C-C
13	191.33	PduP	319	PduE	94	C-C
14	200.52	PduP	57	PduL	12	C-C
15	179.36	PduK	156	PduA	90	S-S
16	150.21	PduD	101	PduC	414	C-C
17	191.31	PduE	54	PduC	466	C-C
18	279.48	PduL	112	PduE	94	C-C
19	153.77	PduA	1	PduK	156	S-S
20	168.82	PduQ	21	PduD	170	C-C
21	253.31	PduC	466	PduE	94	C-C
22	230.4	PduD	215	PduL	112	C-C
23	33.95	PduD	101	PduD	170	C-C
24	147.1	PduS	356	PduC	414	C-C
25	310.25	PduE	94	PduE	150	C-C
26	163.64	PduE	54	PduD	6	C-C
27	240.85	PduD	215	PduC	466	C-C
28	167.12	PduD	101	PduP	57	C-C
29	101.49	PduJ	89	PduA	86	S-S
30	177.83	PduD	101	PduL	112	C-C
31	178.71	PduD	101	PduE	94	C-C
32	210.98	PduD	101	PduD	215	C-C
33	294.29	PduB	138	PduA	55	C-C
34	228.93	PduQ	21	PduC	466	C-C
35	142.08	PduE	150	PduC	466	C-C
36	333.91	PduN	15	PduN	23	S-S
37	255.13	PduB	263	PduE	94	C-C
38	144.7	PduJ	89	PduJ	1	S-S
39	100.12	PduA	86	PduJ	1	S-S
40	61.7	PduQ	214	PduC	414	C-C
41	71.36	PduE	1	PduP	320	C-C
42	105.17	PduS	356	PduS	93	C-C
43	121.6	PduL	3	PduD	6	C-C
44	221.36	PduP	340	PduC	466	C-C
45	251.81	PduB	138	PduB	146	S-S
46	58.6	PduE	73	PduC	414	C-C
47	101.43	PduB	146	PduB	146	S-S
48	49.9	PduC	494	PduL	112	C-C
49	69.82	PduE	73	PduE	54	C-C
50	67.88	PduE	73	PduD	170	C-C
51	165.34	PduC	458	PduD	170	C-C
52	22.22	PduG	59	PduD	170	C-C
53	149.35	PduE	133	PduL	112	C-C
54	160.89	PduE	133	PduD	215	C-C
55	147.26	PduQ	214	PduQ	199	C-C

56	266.68	PduE	73	PduE	94	C-C
57	241.15	PduC	458	PduG	463	C-C
58	147.58	PduE	73	PduD	215	C-C
59	234.24	PduP	306	PduL	112	C-C
60	177.97	PduC	86	PduC	97	C-C
61	134.59	PduE	73	PduC	466	C-C
62	231.02	PduG	59	PduC	466	C-C
63	72.18	PduC	458	PduC	97	C-C
64	120.6	PduA	1	PduJ	1	S-S
65	130.52	PduJ	54	PduJ	1	S-S
66	169.58	PduB	2	PduB	146	S-S
67	154.82	PduC	475	PduD	170	C-C
68	91.13	PduC	475	PduC	395	C-C
69	231.69	PduG	59	PduE	1	C-C
70	319.81	PduB	263	PduB	2	S-S
71	298.61	PduE	94	PduE	168	C-C
72	218.94	PduC	466	PduC	414	C-C
73	168.9	PduC	466	PduP	169	C-C
74	110.08	PduB	263	PduE	54	S-C

DSSO (MS2-MS2-MS3)

ID	Score	Protein A	Position A	Protein B	Position B	Class
1	168.74	PduK	156	PduK	144	S-S
2	33.54	PduE	94	PduC	414	C-C
3	140.63	PduE	94	PduE	168	C-C
4	98.9	PduD	215	PduC	414	C-C
5	88.02	PduC	395	PduC	414	C-C
6	40.3	PduA	90	PduK	144	S-S
7	36.87	PduJ	89	PduK	144	S-S
8	41.24	PduA	0	PduK	144	S-S
9	56.94	PduD	215	PduE	170	C-C
10	36.87	PduJ	0	PduK	144	S-S
11	140.14	PduL	112	PduE	94	C-C
12	128.81	PduC	466	PduE	94	C-C
13	76.17	PduE	0	PduC	414	C-C
14	58.66	PduD	101	PduD	170	C-C
15	324.89	PduE	150	PduE	94	C-C
16	99.76	PduA	86	PduA	90	S-S
17	37.16	PduJ	0	PduK	156	S-S
18	57.5	PduD	101	PduL	112	C-C
19	168.38	PduA	86	PduA	0	S-S
20	38.19	PduE	150	PduE	170	C-C
21	275.95	PduD	101	PduD	215	C-C
22	55.43	PduJ	89	PduA	90	S-S
23	161.76	PduB	138	PduA	55	S-S

24	160.69	PduE	0	PduE	94	C-C
25	216	PduE	1	PduE	94	C-C
26	214.54	PduC	466	PduD	6	C-C
27	180.49	PduA	55	PduA	0	S-S
28	96.34	PduJ	89	PduJ	0	S-S
29	175.79	PduJ	1	PduJ	89	S-S
30	67.06	PduA	0	PduA	90	S-S
31	56.43	PduC	11	PduD	109	C-C
32	24.47	PduD	109	PduE	170	C-C
33	107.59	PduB	263	PduL	12	S-C
34	153.52	PduE	0	PduC	466	C-C
35	78.62	PduE	0	PduP	57	C-C
36	30.64	PduE	0	PduC	255	C-C
37	87.76	PduE	0	PduL	12	C-C
38	47.82	PduP	0	PduD	170	C-C
39	19.48	PduJ	54	PduD	170	S-C
40	80.03	PduD	215	PduE	0	C-C
41	78.19	PduE	0	PduD	0	C-C
42	95.62	PduP	0	PduP	169	C-C
43	111.17	PduC	11	PduE	0	C-C
44	41.89	PduE	150	PduE	0	C-C
45	38.32	PduE	133	PduC	414	C-C
46	183.46	PduB	138	PduB	146	S-S
47	110.43	PduP	0	PduL	12	C-C
48	94.26	PduC	395	PduE	0	C-C
49	114.02	PduO	194	PduP	0	C-C
50	94.26	PduB	146	PduB	146	S-S
51	294.75	PduE	73	PduE	54	C-C
52	45.75	PduE	73	PduD	170	C-C
53	95.79	PduP	0	PduC	466	C-C
54	269.61	PduE	133	PduL	112	C-C
55	89.37	PduE	133	PduD	215	C-C
56	178.32	PduE	73	PduP	57	C-C
57	112.55	PduQ	214	PduQ	199	C-C
58	223.6	PduA	0	PduJ	0	S-S
59	34.72	PduE	73	PduC	11	C-C
60	70.91	PduE	73	PduE	94	C-C
61	148.21	PduE	73	PduD	215	C-C
62	334.51	PduC	458	PduG	463	C-C
63	348.57	PduE	73	PduL	112	C-C
64	123.96	PduC	86	PduC	97	C-C
65	143.09	PduT	1	PduT	117	S-S
66	172.88	PduE	73	PduC	466	C-C
67	289.34	PduG	59	PduC	466	C-C

68	54.41	PduC	148	PduE	0	C-C
69	71.87	PduE	0	PduB	146	S-C
70	87.43	PduC	458	PduC	97	C-C
71	45.16	PduE	133	PduO	1	C-C
72	160.69	PduE	133	PduE	0	C-C
73	162.97	PduB	263	PduB	1	S-S
74	43.49	PduE	73	PduB	263	S-C
75	146.58	PduP	340	PduP	0	C-C
76	213.66	PduP	0	PduE	0	C-C
77	104.22	PduC	494	PduE	0	C-C
78	215.31	PduE	73	PduE	0	C-C
79	167.35	PduB	2	PduB	146	S-S
80	121.68	PduE	73	PduE	1	C-C
81	97.97	PduC	475	PduD	170	C-C
82	138.66	PduE	133	PduP	0	C-C
83	24.19	PduA	37	PduA	90	S-S
84	215.31	PduC	458	PduE	0	C-C
85	176.98	PduT	1	PduJ	0	S-S
86	30.8	PduG	59	PduE	0	C-C
87	393.63	PduB	263	PduB	2	S-S
88	204.37	PduT	1	PduA	0	S-S
89	196.39	PduE	73	PduP	0	C-C
90	114.02	PduP	31	PduE	0	C-C
91	130.27	PduB	1	PduB	177	S-S
92	195.68	PduB	1	PduB	146	S-S
93	189.14	PduJ	54	PduB	1	S-S
94	201.48	PduB	1	PduB	252	S-S
95	154.09	PduA	0	PduA	0	S-S
96	56.01	PduC	475	PduP	0	C-C
97	97.57	PduC	466	PduG	255	C-C
98	117.52	PduE	54	PduC	466	C-C
99	55.19	PduA	55	PduB	146	S-S

DSBU (MS2)

ID	Score	Protein A	Position A	Protein B	Position B	Domain
1	51	PduC	255	PduC	414	C-C
2	77.53	PduP	169	PduP	320	C-C
3	65.33	PduC	395	PduC	414	C-C
4	65.33	PduL	112	PduD	170	C-C
5	46.07	PduA	90	PduK	144	S-S
6	106.84	PduL	112	PduP	57	C-C
7	107.43	PduL	112	PduE	94	C-C
8	196.39	PduE	150	PduE	94	C-C
9	89.08	PduJ	89	PduA	90	S-S
10	71.79	PduN	15	PduN	23	S-S

11	56.47	PduS	356	PduS	363	C-C
12	262.84	PduB2	138	PduA	55	S-S
13	194.28	PduA	55	PduB2	146	S-S
14	203.1	PduB2	138	PduB2	146	S-S
15	100.48	PduJ	36	PduJ	89	S-S
16	119.89	PduC	86	PduC	97	C-C
17	317.98	PduJ	54	PduA	55	S-S
18	135.55	PduJ	54	PduB2	138	S-S
19	96.37	PduP	306	PduP	320	C-C
20	72.62	PduJ	36	PduA	90	S-S
21	135.36	PduC	475	PduC	466	C-C
22	129.84	PduA	12	PduA	90	S-S
23	85.52	PduK	156	PduK	144	S-S
24	65.44	PduE	94	PduC	414	C-C
25	81.69	PduP	169	PduP	320	C-C
26	81.67	PduD	170	PduC	255	C-C
27	46.35	PduP	57	PduD	170	C-C
28	66.99	PduD	199	PduC	414	C-C
29	41.96	PduC	466	PduC	414	C-C
30	68.6	PduC	395	PduC	414	C-C
31	81.67	PduA	72	PduD	170	S-C
32	61.26	PduC	97	PduC	414	C-C
33	70.86	PduD	6	PduC	414	C-C
34	41.94	PduA	86	PduK	144	S-S
35	131.06	PduL	112	PduE	94	C-C
36	40.91	PduD	215	PduE	170	C-C
37	40.91	PduD	6	PduD	170	C-C
38	151.51	PduD	215	PduL	112	C-C
39	148.28	PduC	466	PduE	94	C-C
40	53.52	PduD	101	PduD	170	C-C
41	195.8	PduE	150	PduE	94	C-C
42	102.62	PduA	86	PduA	90	S-S
43	74.53	PduC	395	PduD	170	C-C
44	57.47	PduD	199	PduD	6	C-C
45	67.06	PduJ	89	PduA	90	S-S
46	224.52	PduB2	138	PduA	55	S-S
47	97.31	PduA	90	PduA	90	S-S
48	54.53	PduP	340	PduP	57	C-C
49	67.03	PduP	340	PduD	215	C-C
50	188.07	PduA	55	PduB2	146	S-S
51	169.41	PduB2	138	PduB2	146	S-S
52	137.55	PduC	86	PduC	97	C-C
53	105.39	PduC	97	PduC	97	C-C
54	162.97	PduJ	54	PduA	55	S-S

55	48.74	PduC	542	PduC	395	C-C
56	72.23	PduG	59	PduC	466	C-C
57	88.82	PduP	306	PduP	320	C-C
58	42.69	PduE	133	PduP	340	C-C
59	92.69	PduC	475	PduC	466	C-C
60	49.35	PduA	12	PduA	90	S-S
61	117.93	PduC	494	PduL	112	C-C
62	140.63	PduE	73	PduD	215	C-C
			BS(PEG)5 (MS2)			
1	176.9	PduB2	138	PduA	55	S-S
2	172.56	PduJ	54	PduA	55	S-S
3	165.59	PduB2	118	PduB2	263	S-S
4	154.46	PduB2	138	PduB2	263	S-S
5	135.5	PduE	133	PduE	0	C-C
6	134.52	PduC	86	PduW	0	C-C
7	132.22	PduE	73	PduE	0	C-C
8	131.13	PduJ	54	PduB2	252	S-S
9	128.97	PduP	90	PduP	340	C-C
10	126.67	PduA	1	PduJ	1	S-S
11	126.24	PduJ	54	PduB	0	S-S

Appendix B: All protein-protein crosslinks identified for GFP-containing Pdu BMCs

The GFP-containing Pdu BMC were crosslinked using the crosslinker DSBU and analyzed in the MS2 mode. Note, crosslinking to the N-terminus to PduP or GFP are considered equivalent in this analysis, since the GFP contained the EP for PduP at that position.

DSBU (MS2) for GFP containing Pdu BMCs

ID	Score	Protein A	Position A	Protein B	Position B	Class
1	204.37	PduA	55	PduA	0	S-S
2	194.49	PduA	1	PduA	0	S-S
3	177.83	PduB	138	PduA	55	S-S
4	176.78	PduC	97	PduC	97	C-C
5	161.96	PduE	1	PduE	94	C-C
6	145.01	PduE	150	PduE	94	C-C
7	142.76	PduP/GFP	1	PduP	57	C-C
8	140.21	PduP/GFP	0	PduL	12	C-C
9	140.1	PduA	55	PduA	1	S-S
10	123.99	PduJ	89	PduJ	1	S-S
11	113.71	PduJ	54	PduA	55	S-S
12	112.75	PduT	1	PduT	117	S-S
13	97.07	PduE	0	PduC	466	C-C
14	88.68	PduJ	89	PduJ	0	S-S
15	88.25	PduB	138	PduB	146	S-S
16	72	GFP	132	GFP	126	C-C

17	71.86	PduC	395	PduC	414	C-C
18	68.42	GFP	132	PduP/GFP	0	C-C
19	67.06	PduA	0	PduA	90	S-S
20	67.03	PduL	112	PduE	94	C-C
21	61.26	PduE	0	PduC	414	C-C
22	53.47	PduP/GFP	0	GFP	126	C-C
23	50.97	PduD	215	PduC	466	C-C
24	44.45	PduD	215	PduE	0	C-C
25	41.98	PduE	150	PduE	0	C-C
26	41.57	PduS	356	PduS	363	C-C
27	41.23	PduP	340	PduP/GFP	0	C-C

Appendix C: All nonredundant protein-protein crosslinks identified

Crosslinks from Appendix A were filtered to remove crosslinks captured in multiple datasets. Only the crosslink with the highest score is kept. This simplified list represents the crosslinks used in Chapters 3 and 4.

ID	Dataset	Score	Protein A	Position A	Protein B	Position B	Class
1	DSSO (MS2)	38	PduJ	0	PduK	144	S-S
2	DSSO (MS2)	94.67	PduA	0	PduK	156	S-S
3	DSSO (MS2)	106.84	PduJ	0	PduK	156	S-S
4	DSSO (MS2)	112.75	PduJ	89	PduJ	0	S-S
5	DSSO (MS2)	67.06	PduA	0	PduA	90	S-S
6	DSSO (MS2)	34.09	PduA	86	PduJ	0	S-S
7	DSSO (MS2)	135.35	PduB	146	PduB	146	S-S
8	DSSO (MS2)	37.35	PduT	1	PduK	144	S-S
9	DSSO (MS2)	64.73	PduJ	0	PduA	72	S-S
10	DSSO (MS2)	114.72	PduB	177	PduA	0	S-S
11	DSSO (MS2)	172.54	PduJ	0	PduJ	1	S-S
12	DSSO (MS2)	190.73	PduB	263	PduB	1	S-S
13	DSSO (MS2)	95.79	PduJ	54	PduJ	0	S-S
14	DSSO (MS2)	204.37	PduT	1	PduA	0	S-S
15	DSSO (MS2)	47.82	PduJ	0	PduD	170	S-C
16	DSSO (MS2)	73.62	PduB	263	PduL	112	S-C
17	DSSO (MS2)	106.42	PduB	263	PduD	215	S-C
18	DSSO (MS2)	53.47	PduJ	54	PduD	170	S-C
19	DSSO (MS2)	59.8	PduB	263	PduD	101	S-C
20	DSSO (MS2)	120.76	PduB	263	PduD	6	S-C
21	DSSO (MS2)	90.99	PduE	0	PduB	146	S-C
22	DSSO (MS2)	180.27	PduE	73	PduB	263	S-C
23	DSSO (MS2)	53.47	PduB	1	PduM	0	S-C
24	DSSO (MS2)	30.64	PduL	112	PduC	414	C-C
25	DSSO (MS2)	53.52	PduP	57	PduD	170	C-C
26	DSSO (MS2)	32.32	PduL	112	PduP	320	C-C
27	DSSO (MS2)	67.76	PduD	215	PduD	170	C-C

28	DSSO (MS2)	61.26	PduC	11	PduC	414	C-C
29	DSSO (MS2)	54.4	PduP	319	PduC	466	C-C
30	DSSO (MS2)	38.19	PduD	101	PduO	182	C-C
31	DSSO (MS2)	65.04	PduC	395	PduC	466	C-C
32	DSSO (MS2)	62.72	PduD	215	PduE	150	C-C
33	DSSO (MS2)	159.14	PduE	0	PduE	94	C-C
34	DSSO (MS2)	53.71	PduD	6	PduL	12	C-C
35	DSSO (MS2)	66.14	PduD	101	PduC	466	C-C
36	DSSO (MS2)	45.16	PduC	148	PduC	414	C-C
37	DSSO (MS2)	53.47	PduE	0	PduD	170	C-C
38	DSSO (MS2)	75.82	PduC	11	PduD	109	C-C
39	DSSO (MS2)	37.35	PduD	0	PduD	0	C-C
40	DSSO (MS2)	104.64	PduE	0	PduL	12	C-C
41	DSSO (MS2)	41.96	PduD	109	PduE	170	C-C
42	DSSO (MS2)	55.6	PduE	0	PduP	320	C-C
43	DSSO (MS2)	65.44	PduP	0	PduC	414	C-C
44	DSSO (MS2)	38.32	PduE	0	PduC	255	C-C
45	DSSO (MS2)	56.55	PduP	340	PduL	112	C-C
46	DSSO (MS2)	34.81	PduE	0	PduG	463	C-C
47	DSSO (MS2)	64	PduP	0	PduD	170	C-C
48	DSSO (MS2)	38.32	PduC	494	PduC	414	C-C
49	DSSO (MS2)	126.24	PduE	150	PduE	0	C-C
50	DSSO (MS2)	62.84	PduD	24	PduP	57	C-C
51	DSSO (MS2)	56.12	PduE	0	PduL	112	C-C
52	DSSO (MS2)	114.02	PduP	0	PduP	57	C-C
53	DSSO (MS2)	50.97	PduE	150	PduE	1	C-C
54	DSSO (MS2)	30.64	PduP	0	PduC	255	C-C
55	DSSO (MS2)	101.21	PduP	0	PduE	94	C-C
56	DSSO (MS2)	114.89	PduC	11	PduP	0	C-C
57	DSSO (MS2)	45.95	PduD	215	PduP	0	C-C
58	DSSO (MS2)	140.21	PduP	0	PduL	12	C-C
59	DSSO (MS2)	66.99	PduE	133	PduE	94	C-C
60	DSSO (MS2)	46.02	PduD	101	PduE	0	C-C
61	DSSO (MS2)	54.14	PduE	73	PduP	319	C-C
62	DSSO (MS2)	55.57	PduE	73	PduC	11	C-C
63	DSSO (MS2)	151.17	PduE	73	PduD	215	C-C
64	DSSO (MS2)	32.35	PduP	0	PduE	150	C-C
65	DSSO (MS2)	56.12	PduD	101	PduP	0	C-C
66	DSSO (MS2)	71.03	PduC	458	PduC	466	C-C
67	DSSO (MS2)	72.02	PduE	133	PduD	6	C-C
68	DSSO (MS2)	46.62	PduE	133	PduE	170	C-C
69	DSSO (MS2)	99.01	PduP	340	PduE	0	C-C
70	DSSO (MS2)	35.17	PduP	0	PduD	0	C-C
71	DSSO (MS2)	235.31	PduE	0	PduE	1	C-C

72	DSSO (MS2)	97.07	PduE	59	PduE	0	C-C
73	DSSO (MS2)	97.46	PduE	73	PduP	340	C-C
74	DSSO (MS2)	213.66	PduC	494	PduE	0	C-C
75	DSSO (MS2)	137.5	PduE	73	PduE	1	C-C
76	DSSO (MS2)	126.24	PduC	86	PduE	0	C-C
77	DSSO (MS2)	117.4	PduG	59	PduE	0	C-C
78	DSSO (MS2)	35.22	PduC	458	PduP	0	C-C
79	DSSO (MS2)	157.54	PduE	73	PduG	59	C-C
80	DSSO (MS2)	46.02	PduC	475	PduE	0	C-C
81	DSSO (MS2-MS3)	268.9	PduA	55	PduK	144	S-S
82	DSSO (MS2-MS3)	146.01	PduA	86	PduK	144	S-S
83	DSSO (MS2-MS3)	179.36	PduK	156	PduA	90	S-S
84	DSSO (MS2-MS3)	153.77	PduA	1	PduK	156	S-S
85	DSSO (MS2-MS3)	101.49	PduJ	89	PduA	86	S-S
86	DSSO (MS2-MS3)	333.91	PduN	15	PduN	23	S-S
87	DSSO (MS2-MS3)	100.12	PduA	86	PduJ	1	S-S
88	DSSO (MS2-MS3)	251.81	PduB	138	PduB	146	S-S
89	DSSO (MS2-MS3)	120.6	PduA	1	PduJ	1	S-S
90	DSSO (MS2-MS3)	130.52	PduJ	54	PduJ	1	S-S
91	DSSO (MS2-MS3)	169.58	PduB	2	PduB	146	S-S
92	DSSO (MS2-MS3)	294.29	PduB	138	PduA	55	S-S
93	DSSO (MS2-MS3)	110.08	PduB	263	PduE	54	S-C
94	DSSO (MS2-MS3)	255.13	PduB	263	PduE	94	S-C
95	DSSO (MS2-MS3)	119.62	PduD	170	PduP	320	C-C
96	DSSO (MS2-MS3)	281.93	PduE	94	PduD	170	C-C
97	DSSO (MS2-MS3)	114.86	PduC	395	PduC	414	C-C
98	DSSO (MS2-MS3)	175.51	PduL	112	PduD	170	C-C
99	DSSO (MS2-MS3)	223.07	PduE	54	PduE	94	C-C
100	DSSO (MS2-MS3)	226.99	PduC	466	PduD	170	C-C
101	DSSO (MS2-MS3)	189.73	PduC	466	PduG	255	C-C
102	DSSO (MS2-MS3)	144.28	PduC	466	PduP	320	C-C
103	DSSO (MS2-MS3)	171.26	PduE	54	PduL	112	C-C
104	DSSO (MS2-MS3)	191.33	PduP	319	PduE	94	C-C
105	DSSO (MS2-MS3)	200.52	PduP	57	PduL	12	C-C
106	DSSO (MS2-MS3)	150.21	PduD	101	PduC	414	C-C
107	DSSO (MS2-MS3)	191.31	PduE	54	PduC	466	C-C
108	DSSO (MS2-MS3)	279.48	PduL	112	PduE	94	C-C
109	DSSO (MS2-MS3)	168.82	PduQ	21	PduD	170	C-C
110	DSSO (MS2-MS3)	253.31	PduC	466	PduE	94	C-C
111	DSSO (MS2-MS3)	230.4	PduD	215	PduL	112	C-C
112	DSSO (MS2-MS3)	147.1	PduS	356	PduC	414	C-C
113	DSSO (MS2-MS3)	163.64	PduE	54	PduD	6	C-C
114	DSSO (MS2-MS3)	240.85	PduD	215	PduC	466	C-C
115	DSSO (MS2-MS3)	167.12	PduD	101	PduP	57	C-C

116	DSSO (MS2-MS3)	177.83	PduD	101	PduL	112	C-C
117	DSSO (MS2-MS3)	178.71	PduD	101	PduE	94	C-C
118	DSSO (MS2-MS3)	228.93	PduQ	21	PduC	466	C-C
119	DSSO (MS2-MS3)	142.08	PduE	150	PduC	466	C-C
120	DSSO (MS2-MS3)	61.7	PduQ	214	PduC	414	C-C
121	DSSO (MS2-MS3)	71.36	PduE	1	PduP	320	C-C
122	DSSO (MS2-MS3)	105.17	PduS	356	PduS	93	C-C
123	DSSO (MS2-MS3)	121.6	PduL	3	PduD	6	C-C
124	DSSO (MS2-MS3)	221.36	PduP	340	PduC	466	C-C
125	DSSO (MS2-MS3)	58.6	PduE	73	PduC	414	C-C
126	DSSO (MS2-MS3)	49.9	PduC	494	PduL	112	C-C
127	DSSO (MS2-MS3)	67.88	PduE	73	PduD	170	C-C
128	DSSO (MS2-MS3)	165.34	PduC	458	PduD	170	C-C
129	DSSO (MS2-MS3)	160.89	PduE	133	PduD	215	C-C
130	DSSO (MS2-MS3)	147.26	PduQ	214	PduQ	199	C-C
131	DSSO (MS2-MS3)	266.68	PduE	73	PduE	94	C-C
132	DSSO (MS2-MS3)	234.24	PduP	306	PduL	112	C-C
133	DSSO (MS2-MS3)	177.97	PduC	86	PduC	97	C-C
134	DSSO (MS2-MS3)	154.82	PduC	475	PduD	170	C-C
135	DSSO (MS2-MS3)	91.13	PduC	475	PduC	395	C-C
136	DSSO (MS2-MS3)	231.69	PduG	59	PduE	1	C-C
137	DSSO (MS2-MS3)	298.61	PduE	94	PduE	168	C-C
138	DSSO (MS2-MS3)	218.94	PduC	466	PduC	414	C-C
139	DSSO (MS2-MS3)	168.9	PduC	466	PduP	169	C-C
140	DSSO (MS2-MS2-MS3)	168.74	PduK	156	PduK	144	S-S
141	DSSO (MS2-MS2-MS3)	36.87	PduJ	89	PduK	144	S-S
142	DSSO (MS2-MS2-MS3)	41.24	PduA	0	PduK	144	S-S
143	DSSO (MS2-MS2-MS3)	168.38	PduA	86	PduA	0	S-S
144	DSSO (MS2-MS2-MS3)	180.49	PduA	55	PduA	0	S-S
145	DSSO (MS2-MS2-MS3)	175.79	PduJ	1	PduJ	89	S-S
146	DSSO (MS2-MS2-MS3)	223.6	PduA	0	PduJ	0	S-S
147	DSSO (MS2-MS2-MS3)	143.09	PduT	1	PduT	117	S-S
148	DSSO (MS2-MS2-MS3)	176.98	PduT	1	PduJ	0	S-S
149	DSSO (MS2-MS2-MS3)	393.63	PduB	263	PduB	2	S-S
150	DSSO (MS2-MS2-MS3)	130.27	PduB	1	PduB	177	S-S
151	DSSO (MS2-MS2-MS3)	195.68	PduB	1	PduB	146	S-S
152	DSSO (MS2-MS2-MS3)	189.14	PduJ	54	PduB	1	S-S
153	DSSO (MS2-MS2-MS3)	201.48	PduB	1	PduB	252	S-S
154	DSSO (MS2-MS2-MS3)	154.09	PduA	0	PduA	0	S-S
155	DSSO (MS2-MS2-MS3)	107.59	PduB	263	PduL	12	S-C
156	DSSO (MS2-MS2-MS3)	98.9	PduD	215	PduC	414	C-C
157	DSSO (MS2-MS2-MS3)	56.94	PduD	215	PduE	170	C-C
158	DSSO (MS2-MS2-MS3)	76.17	PduE	0	PduC	414	C-C
159	DSSO (MS2-MS2-MS3)	58.66	PduD	101	PduD	170	C-C

160	DSSO (MS2-MS2-MS3)	324.89	PduE	150	PduE	94	C-C
161	DSSO (MS2-MS2-MS3)	38.19	PduE	150	PduE	170	C-C
162	DSSO (MS2-MS2-MS3)	275.95	PduD	101	PduD	215	C-C
163	DSSO (MS2-MS2-MS3)	216	PduE	1	PduE	94	C-C
164	DSSO (MS2-MS2-MS3)	214.54	PduC	466	PduD	6	C-C
165	DSSO (MS2-MS2-MS3)	153.52	PduE	0	PduC	466	C-C
166	DSSO (MS2-MS2-MS3)	78.62	PduE	0	PduP	57	C-C
167	DSSO (MS2-MS2-MS3)	80.03	PduD	215	PduE	0	C-C
168	DSSO (MS2-MS2-MS3)	78.19	PduE	0	PduD	0	C-C
169	DSSO (MS2-MS2-MS3)	95.62	PduP	0	PduP	169	C-C
170	DSSO (MS2-MS2-MS3)	111.17	PduC	11	PduE	0	C-C
171	DSSO (MS2-MS2-MS3)	38.32	PduE	133	PduC	414	C-C
172	DSSO (MS2-MS2-MS3)	94.26	PduC	395	PduE	0	C-C
173	DSSO (MS2-MS2-MS3)	114.02	PduO	194	PduP	0	C-C
174	DSSO (MS2-MS2-MS3)	294.75	PduE	73	PduE	54	C-C
175	DSSO (MS2-MS2-MS3)	95.79	PduP	0	PduC	466	C-C
176	DSSO (MS2-MS2-MS3)	269.61	PduE	133	PduL	112	C-C
177	DSSO (MS2-MS2-MS3)	178.32	PduE	73	PduP	57	C-C
178	DSSO (MS2-MS2-MS3)	334.51	PduC	458	PduG	463	C-C
179	DSSO (MS2-MS2-MS3)	348.57	PduE	73	PduL	112	C-C
180	DSSO (MS2-MS2-MS3)	172.88	PduE	73	PduC	466	C-C
181	DSSO (MS2-MS2-MS3)	289.34	PduG	59	PduC	466	C-C
182	DSSO (MS2-MS2-MS3)	54.41	PduC	148	PduE	0	C-C
183	DSSO (MS2-MS2-MS3)	87.43	PduC	458	PduC	97	C-C
184	DSSO (MS2-MS2-MS3)	45.16	PduE	133	PduO	1	C-C
185	DSSO (MS2-MS2-MS3)	160.69	PduE	133	PduE	0	C-C
186	DSSO (MS2-MS2-MS3)	146.58	PduP	340	PduP	0	C-C
187	DSSO (MS2-MS2-MS3)	213.66	PduP	0	PduE	0	C-C
188	DSSO (MS2-MS2-MS3)	215.31	PduE	73	PduE	0	C-C
189	DSSO (MS2-MS2-MS3)	138.66	PduE	133	PduP	0	C-C
190	DSSO (MS2-MS2-MS3)	215.31	PduC	458	PduE	0	C-C
191	DSSO (MS2-MS2-MS3)	196.39	PduE	73	PduP	0	C-C
192	DSSO (MS2-MS2-MS3)	114.02	PduP	31	PduE	0	C-C
193	DSSO (MS2-MS2-MS3)	56.01	PduC	475	PduP	0	C-C
194	DSBU (MS2)	46.07	PduA	90	PduK	144	S-S
195	DSBU (MS2)	89.08	PduJ	89	PduA	90	S-S
196	DSBU (MS2)	194.28	PduA	55	PduB	146	S-S
197	DSBU (MS2)	100.48	PduJ	36	PduJ	89	S-S
198	DSBU (MS2)	317.98	PduJ	54	PduA	55	S-S
199	DSBU (MS2)	135.55	PduJ	54	PduB	138	S-S
200	DSBU (MS2)	72.62	PduJ	36	PduA	90	S-S
201	DSBU (MS2)	129.84	PduA	12	PduA	90	S-S
202	DSBU (MS2)	102.62	PduA	86	PduA	90	S-S
203	DSBU (MS2)	97.31	PduA	90	PduA	90	S-S

204	DSBU (MS2)	81.67	PduA	72	PduD	170	S-C
205	DSBU (MS2)	51	PduC	255	PduC	414	C-C
206	DSBU (MS2)	106.84	PduL	112	PduP	57	C-C
207	DSBU (MS2)	56.47	PduS	356	PduS	363	C-C
208	DSBU (MS2)	96.37	PduP	306	PduP	320	C-C
209	DSBU (MS2)	135.36	PduC	475	PduC	466	C-C
210	DSBU (MS2)	65.44	PduE	94	PduC	414	C-C
211	DSBU (MS2)	81.69	PduP	169	PduP	320	C-C
212	DSBU (MS2)	81.67	PduD	170	PduC	255	C-C
213	DSBU (MS2)	66.99	PduD	199	PduC	414	C-C
214	DSBU (MS2)	61.26	PduC	97	PduC	414	C-C
215	DSBU (MS2)	70.86	PduD	6	PduC	414	C-C
216	DSBU (MS2)	99.23	PduD	6	PduD	170	C-C
217	DSBU (MS2)	74.53	PduC	395	PduD	170	C-C
218	DSBU (MS2)	57.47	PduD	199	PduD	6	C-C
219	DSBU (MS2)	54.53	PduP	340	PduP	57	C-C
220	DSBU (MS2)	67.03	PduP	340	PduD	215	C-C
221	DSBU (MS2)	105.39	PduC	97	PduC	97	C-C
222	DSBU (MS2)	48.74	PduC	542	PduC	395	C-C
223	DSBU (MS2)	42.69	PduE	133	PduP	340	C-C
224	BS(PEG)5	165.59	PduB	118	PduB	263	S-S
225	BS(PEG)5	154.46	PduB	138	PduB	263	S-S
226	BS(PEG)5	131.13	PduJ	54	PduB	252	S-S
227	BS(PEG)5	126.67	PduA	1	PduJ	0	S-S
228	BS(PEG)5	126.24	PduJ	54	PduB	0	S-S
229	BS(PEG)5	134.52	PduC	86	PduW	0	C-C
230	BS(PEG)5	128.97	PduP	90	PduP	340	C-C

Bibliography

Abeles, R. H. and Lee, H. A. (1961) 'An intramolecular oxidation-reduction requiring a cobamide coenzyme', *J Biol Chem*, 236, pp. 2347-50.

Arakawa, T., Kita, Y. and Timasheff, S. N. (2007) 'Protein precipitation and denaturation by dimethyl sulfoxide', *Biophys Chem*, 131(1-3), pp. 62-70.

Asija, K., Sutter, M. and Kerfeld, C. A. (2021) 'A Survey of Bacterial Microcompartment Distribution in the Human Microbiome', *Front Microbiol*, 12, pp. 669024.

Aussignargues, C., Paasch, B. C., Gonzalez-Esquer, R., Erbilgin, O. and Kerfeld, C. A. (2015) 'Bacterial microcompartment assembly: The key role of encapsulation peptides', *Commun Integr Biol*, 8(3), pp. e1039755.

Aussignargues, C., Pandelia, M. E., Sutter, M., Plegaria, J. S., Zarzycki, J., Turmo, A., Huang, J., Ducat, D. C., Hegg, E. L., Gibney, B. R. and Kerfeld, C. A. (2016) 'Structure and Function of a Bacterial Microcompartment Shell Protein Engineered to Bind a [4Fe-4S] Cluster', *J Am Chem Soc*, 138(16), pp. 5262-70.

Axen, S. D., Erbilgin, O. and Kerfeld, C. A. (2014) 'A taxonomy of bacterial microcompartment loci constructed by a novel scoring method', *PLoS Comput Biol*, 10(10), pp. e1003898.

Banani, S. F., Lee, H. O., Hyman, A. A. and Rosen, M. K. (2017) 'Biomolecular condensates: organizers of cellular biochemistry', *Nat Rev Mol Cell Biol*, 18(5), pp. 285-298.

Berry, S., Fischer, J. H., Kruip, J., Hauser, M. and Wildner, G. F. (2005) 'Monitoring cytosolic pH of carboxysome-deficient cells of Synechocystis sp. PCC 6803 using fluorescence analysis', *Plant Biol (Stuttg)*, 7(4), pp. 342-7.

Bhattacharyya, S., Yu, H., Mim, C. and Matouschek, A. (2014) 'Regulated protein turnover: snapshots of the proteasome in action', *Nat Rev Mol Cell Biol*, 15(2), pp. 122-33.

Bobik, T. A., Ailion, M. and Roth, J. R. (1992) 'A single regulatory gene integrates control of vitamin B12 synthesis and propanediol degradation', *J Bacteriol*, 174(7), pp. 2253-66.

Bobik, T. A., Havemann, G. D., Busch, R. J., Williams, D. S. and Aldrich, H. C. (1999) 'The propanediol utilization (pdu) operon of Salmonella enterica serovar Typhimurium LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B(12)-dependent 1, 2-propanediol degradation', *J Bacteriol*, 181(19), pp. 5967-75.

Bobik, T. A., Xu, Y., Jeter, R. M., Otto, K. E. and Roth, J. R. (1997) 'Propanediol utilization genes (pdu) of Salmonella typhimurium: three genes for the propanediol dehydratase', *J Bacteriol*, 179(21), pp. 6633-9.

Bolte, S. and Cordelières, F. P. (2006) 'A guided tour into subcellular colocalization analysis in light microscopy', *J Microsc*, 224(Pt 3), pp. 213-32.

Bonacci, W., Teng, P. K., Afonso, B., Niederholtmeyer, H., Grob, P., Silver, P. A. and Savage, D. F. (2012) 'Modularity of a carbon-fixing protein organelle', *Proc Natl Acad Sci U S A*, 109(2), pp. 478-83.

Brodie, N. I., Popov, K. I., Petrotchenko, E. V., Dokholyan, N. V. and Borchers, C. H. (2017) 'Solving protein structures using short-distance cross-linking constraints as a guide for discrete molecular dynamics simulations', *Sci Adv,* 3(7), pp. e1700479.

Bryant, D. A. and Frigaard, N. U. (2006) 'Prokaryotic photosynthesis and phototrophy illuminated', *Trends Microbiol,* 14(11), pp. 488-96.

Cai, F., Menon, B. B., Cannon, G. C., Curry, K. J., Shively, J. M. and Heinhorst, S. (2009) 'The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO2 leakage barrier', *PLoS One,* 4(10), pp. e7521.

Cai, F., Sutter, M., Cameron, J. C., Stanley, D. N., Kinney, J. N. and Kerfeld, C. A. (2013) 'The structure of CcmP, a tandem bacterial microcompartment domain protein from the β-carboxysome, forms a subcompartment within a microcompartment', *J Biol Chem,* 288(22), pp. 16055-63.

Cameron, J. C., Wilson, S. C., Bernstein, S. L. and Kerfeld, C. A. (2013) 'Biogenesis of a bacterial organelle: the carboxysome assembly pathway', *Cell,* 155(5), pp. 1131-40.

Caspar, D. L. and Klug, A. (1962) 'Physical principles in the construction of regular viruses', *Cold Spring Harb Symp Quant Biol,* 27, pp. 1-24.

Chen, A. H., Robinson-Mosher, A., Savage, D. F., Silver, P. A. and Polka, J. K. (2013) 'The bacterial carbon-fixing organelle is formed by shell envelopment of preassembled cargo', *PLoS One,* 8(9), pp. e76127.

Cheng, S. and Bobik, T. A. (2010) 'Characterization of the PduS cobalamin reductase of Salmonella enterica and its role in the Pdu microcompartment', *J Bacteriol,* 192(19), pp. 5071-80.

Cheng, S., Fan, C., Sinha, S. and Bobik, T. A. (2012) 'The PduQ enzyme is an alcohol dehydrogenase used to recycle NAD+ internally within the Pdu microcompartment of Salmonella enterica', *PLoS One,* 7(10), pp. e47144.

Cheng, S., Liu, Y., Crowley, C. S., Yeates, T. O. and Bobik, T. A. (2008) 'Bacterial microcompartments: their properties and paradoxes', *Bioessays,* 30(11-12), pp. 1084-95.

Chessher, A., Breitling, R. and Takano, E. (2015) 'Bacterial Microcompartments: Biomaterials for Synthetic Biology-Based Compartmentalization Strategies'.

Chowdhury, C. and Bobik, T. A. (2019) 'Engineering the PduT shell protein to modify the permeability of the 1,2-propanediol microcompartment of', *Microbiology (Reading),* 165(12), pp. 1355-1364.

Chowdhury, C., Chun, S., Pang, A., Sawaya, M. R., Sinha, S., Yeates, T. O. and Bobik, T. A. (2015) 'Selective molecular transport through the protein shell of a bacterial microcompartment organelle', *Proc Natl Acad Sci U S A,* 112(10), pp. 2990-5.

Chowdhury, C., Chun, S., Sawaya, M. R., Yeates, T. O. and Bobik, T. A. (2016) 'The function of the PduJ microcompartment shell protein is determined by the genomic position of its encoding gene', *Mol Microbiol,* 101(5), pp. 770-83.

Chowdhury, C., Sinha, S., Chun, S., Yeates, T. O. and Bobik, T. A. (2014) 'Diverse bacterial microcompartment organelles', *Microbiol Mol Biol Rev,* 78(3), pp. 438-68.

Chu, F., Thornton, D. T. and Nguyen, H. T. (2018) 'Chemical cross-linking in the structural analysis of protein assemblies', *Methods,* 144, pp. 53-63.

Craciun, S. and Balskus, E. P. (2012) 'Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme', *Proc Natl Acad Sci U S A,* 109(52), pp. 21307-12.

Crowley, C. S., Cascio, D., Sawaya, M. R., Kopstein, J. S., Bobik, T. A. and Yeates, T. O. (2010) 'Structural insight into the mechanisms of transport across the Salmonella enterica Pdu microcompartment shell', *J Biol Chem,* 285(48), pp. 37838-46.

Crowley, C. S., Sawaya, M. R., Bobik, T. A. and Yeates, T. O. (2008) 'Structure of the PduU shell protein from the Pdu microcompartment of Salmonella', *Structure,* 16(9), pp. 1324-32.

Daniel, R., Bobik, T. A. and Gottschalk, G. (1998) 'Biochemistry of coenzyme B12-dependent glycerol and diol dehydratases and organization of the encoding genes', *FEMS Microbiol Rev,* 22(5), pp. 553-66.

Deutsch, E. W., Bandeira, N., Sharma, V., Perez-Riverol, Y., Carver, J. J., Kundu, D. J., García-Seisdedos, D., Jarnuczak, A. F., Hewapathirana, S., Pullman, B. S., Wertz, J., Sun, Z., Kawano, S., Okuda, S., Watanabe, Y., Hermjakob, H., MacLean, B., MacCoss, M. J., Zhu, Y., Ishihama, Y. and Vizcaíno, J. A. (2020) 'The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics', *Nucleic Acids Res,* 48(D1), pp. D1145-D1152.

Dou, Z., Heinhorst, S., Williams, E. B., Murin, C. D., Shively, J. M. and Cannon, G. C. (2008) 'CO2 fixation kinetics of Halothiobacillus neapolitanus mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO2', *J Biol Chem,* 283(16), pp. 10377-84.

Drews, G. and Niklowitz, W. (1956) '[Cytology of Cyanophycea. II. Centroplasm and granular inclusions of Phormidium uncinatum]', *Archiv fur Mikrobiologie,* 24(2), pp. 174-62.

Dryden, K. A., Crowley, C. S., Tanaka, S., Yeates, T. O. and Yeager, M. (2009) 'Two-dimensional crystals of carboxysome shell proteins recapitulate the hexagonal packing of three-dimensional crystals', *Protein Sci,* 18(12), pp. 2629-35.

Erbilgin, O., Sutter, M. and Kerfeld, C. A. (2016) 'The Structural Basis of Coenzyme A Recycling in a Bacterial Organelle', *PLoS Biol,* 14(3), pp. e1002399.

Erdős, G., Pajkos, M. and Dosztányi, Z. (2021) 'IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation', *Nucleic Acids Res,* 49(W1), pp. W297-W303.

Fan, C. and Bobik, T. A. (2011) 'The N-terminal region of the medium subunit (PduD) packages adenosylcobalamin-dependent diol dehydratase (PduCDE) into the Pdu microcompartment', *J Bacteriol,* 193(20), pp. 5623-8.

Fan, C., Cheng, S., Liu, Y., Escobar, C. M., Crowley, C. S., Jefferson, R. E., Yeates, T. O. and Bobik, T. A. (2010) 'Short N-terminal sequences package proteins into bacterial microcompartments', *Proc Natl Acad Sci U S A,* 107(16), pp. 7509-14.

Fan, C., Cheng, S., Sinha, S. and Bobik, T. A. (2012) 'Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into

bacterial microcompartments', *Proc Natl Acad Sci U S A*, 109(37), pp. 14995-5000.

Fang, Y., Huang, F., Faulkner, M., Jiang, Q., Dykes, G. F., Yang, M. and Liu, L. N. (2018) 'Engineering and Modulating Functional Cyanobacterial CO', *Front Plant Sci*, 9, pp. 739.

Faulkner, M., Rodriguez-Ramos, J., Dykes, G. F., Owen, S. V., Casella, S., Simpson, D. M., Beynon, R. J. and Liu, L. N. (2017) 'Direct characterization of the native structure and mechanics of cyanobacterial carboxysomes', *Nanoscale*, 9(30), pp. 10662-10673.

Faulkner, M., Szabó, I., Weetman, S. L., Sicard, F., Huber, R. G., Bond, P. J., Rosta, E. and Liu, L. N. (2020) 'Molecular simulations unravel the molecular principles that mediate selective permeability of carboxysome shell protein', *Sci Rep*, 10(1), pp. 17501.

Faulkner, M., Zhao, L. S., Barrett, S. and Liu, L. N. (2019) 'Self-Assembly Stability and Variability of Bacterial Microcompartment Shell Proteins in Response to the Environmental Change', *Nanoscale Res Lett*, 14(1), pp. 54.

Ferlez, B., Sutter, M. and Kerfeld, C. A. (2019) 'Glycyl Radical Enzyme-Associated Microcompartments: Redox-Replete Bacterial Organelles', *MBio*, 10(1).

Frank, S., Lawrence, A. D., Prentice, M. B. and Warren, M. J. (2013) 'Bacterial microcompartments moving into a synthetic biological world', *J Biotechnol*, 163(2), pp. 273-9.

Gabashvili, A. N., Chmelyuk, N. S., Efremova, M. V., Malinovskaya, J. A., Semkina, A. S. and Abakumov, M. A. (2020) 'Encapsulins-Bacterial Protein Nanocompartments: Structure, Properties, and Application', *Biomolecules*, 10(6).

Gonzalez-Esquer, C. R., Newnham, S. E. and Kerfeld, C. A. (2016) 'Bacterial microcompartments as metabolic modules for plant synthetic biology', *Plant J*, 87(1), pp. 66-75.

Graf, L., Wu, K. and Wilson, J. W. (2018) 'Transfer and analysis of Salmonella pdu genes in a range of Gram-negative bacteria demonstrate exogenous microcompartment expression across a variety of species', *Microb Biotechnol*, 11(1), pp. 199-210.

Greber, B. J., Sutter, M. and Kerfeld, C. A. (2019) 'The Plasticity of Molecular Interactions Governs Bacterial Microcompartment Shell Assembly', *Structure*, 27(5), pp. 749-763.e4.

Greening, C. and Lithgow, T. (2020) 'Formation and function of bacterial organelles', *Nat Rev Microbiol*, 18(12), pp. 677-689.

Gruber, A. (2019) 'What's in a name? How organelles of endosymbiotic origin can be distinguished from endosymbionts', *Microb Cell*, 6(2), pp. 123-133.

Hagen, A., Sutter, M., Sloan, N. and Kerfeld, C. A. (2018a) 'Programmed loading and rapid purification of engineered bacterial microcompartment shells', *Nat Commun*, 9(1), pp. 2881.

Hagen, A. R., Plegaria, J. S., Sloan, N., Ferlez, B., Aussignargues, C., Burton, R. and Kerfeld, C. A. (2018b) 'In Vitro Assembly of Diverse Bacterial Microcompartment Shell Architectures', *Nano Lett*, 18(11), pp. 7030-7037.

Havemann, G. D. and Bobik, T. A. (2003) 'Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in Salmonella enterica serovar Typhimurium LT2', *J Bacteriol*, 185(17), pp. 5086-95.

Havemann, G. D., Sampson, E. M. and Bobik, T. A. (2002) 'PduA is a shell protein of polyhedral organelles involved in coenzyme B(12)-dependent degradation of 1,2-propanediol in Salmonella enterica serovar typhimurium LT2', *J Bacteriol*, 184(5), pp. 1253-61.

Heldt, D., Frank, S., Seyedarabi, A., Ladikis, D., Parsons, J. B., Warren, M. J. and Pickersgill, R. W. (2009) 'Structure of a trimeric bacterial microcompartment shell protein, EtuB, associated with ethanol utilization in Clostridium kluyveri', *Biochem J*, 423(2), pp. 199-207.

Herring, T. I., Harris, T. N., Chowdhury, C., Mohanty, S. K. and Bobik, T. A. (2018) 'A Bacterial Microcompartment Is Used for Choline Fermentation by Escherichia coli 536', *J Bacteriol*, 200(10).

Hill, N. C., Tay, J. W., Altus, S., Bortz, D. M. and Cameron, J. C. (2020) 'Life cycle of a cyanobacterial carboxysome', *Sci Adv*, 6(19), pp. eaba1269.

Hoyt, E., Cal, P., Oliveira, B. and Bernardes, G. 2019. Contemporary approaches to site-selective protein modification. Nat Rev Chem.

Huang, J., Ferlez, B. H., Young, E. J., Kerfeld, C. A., Kramer, D. M. and Ducat, D. C. (2019) 'Functionalization of Bacterial Microcompartment Shell Proteins With Covalently Attached Heme', *Front Bioeng Biotechnol*, 7, pp. 432.

Huff, J. 2015. The Airyscan detector from ZEISS: confocal imaging with improved signal-to-noise ratio and super-resolution. Nat Methods.

Huseby, D. L. and Roth, J. R. (2013) 'Evidence that a metabolic microcompartment contains and recycles private cofactor pools', *J Bacteriol*, 195(12), pp. 2864-79.

Iancu, C. V., Morris, D. M., Dou, Z., Heinhorst, S., Cannon, G. C. and Jensen, G. J. (2010) 'Organization, structure, and assembly of alpha-carboxysomes determined by electron cryotomography of intact cells', *J Mol Biol*, 396(1), pp. 105-17.

Jackson, M. and Mantsch, H. H. (1991) 'Beware of proteins in DMSO', *Biochim Biophys Acta*, 1078(2), pp. 231-5.

Jakobson, C. M., Kim, E. Y., Slininger, M. F., Chien, A. and Tullman-Ercek, D. (2015) 'Localization of proteins to the 1,2-propanediol utilization microcompartment by non-native signal sequences is mediated by a common hydrophobic motif', *J Biol Chem*, 290(40), pp. 24519-33.

Jakobson, C. M. and Tullman-Ercek, D. (2016) 'Dumpster Diving in the Gut: Bacterial Microcompartments as Part of a Host-Associated Lifestyle', *PLoS Pathog*, 12(5), pp. e1005558.

Jakobson, C. M., Tullman-Ercek, D., Slininger, M. F. and Mangan, N. M. (2017) 'A systems-level model reveals that 1,2-Propanediol utilization microcompartments enhance pathway flux through intermediate sequestration', *PLoS Comput Biol*, 13(5), pp. e1005525.

Jiang, Y. and Kalodimos, C. G. (2017) 'NMR Studies of Large Proteins', *J Mol Biol*, 429(17), pp. 2667-2676.

Jorda, J., Leibly, D. J., Thompson, M. C. and Yeates, T. O. (2016) 'Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein', *Chem Commun (Camb),* 52(28), pp. 5041-4.

Jorda, J., Liu, Y., Bobik, T. A. and Yeates, T. O. (2015) 'Exploring bacterial organelle interactomes: a model of the protein-protein interaction network in the Pdu microcompartment', *PLoS Comput Biol,* 11(2), pp. e1004067.

Juodeikis, R., Lee, M. J., Mayer, M., Mantell, J., Brown, I. R., Verkade, P., Woolfson, D. N., Prentice, M. B., Frank, S. and Warren, M. J. (2020) 'Effect of metabolosome encapsulation peptides on enzyme activity, coaggregation, incorporation, and bacterial microcompartment formation', *Microbiologyopen,* 9(5), pp. e1010.

Kahn, J. D., Lemke, E. A. and Pappu, R. V. (2021) 'Faces, facets, and functions of biomolecular condensates driven by multivalent proteins and nucleic acids', *Biophys J,* 120(7), pp. E1-E4.

Kalnins, G., Cesle, E. E., Jansons, J., Liepins, J., Filimonenko, A. and Tars, K. (2020) 'Encapsulation mechanisms and structural studies of GRM2 bacterial microcompartment particles', *Nat Commun,* 11(1), pp. 388.

Kaval, K. G. and Garsin, D. A. (2018) 'Ethanolamine Utilization in Bacteria', *MBio,* 9(1).

Keeling, P. J. and Archibald, J. M. (2008) 'Organelle evolution: what's in a name?', *Curr Biol,* 18(8), pp. R345-7.

Kennedy, N. W., Hershewe, J. M., Nichols, T. M., Roth, E. W., Wilke, C. D., Mills, C. E., Jewett, M. C. and Tullman-Ercek, D. (2020) 'Apparent size and morphology of bacterial microcompartments varies with technique', *PLoS One,* 15(3), pp. e0226395.

Kennedy, N. W., Ikonomova, S. P., Slininger Lee, M., Raeder, H. W. and Tullman-Ercek, D. (2021) 'Self-assembling Shell Proteins PduA and PduJ have Essential and Redundant Roles in Bacterial Microcompartment Assembly', *J Mol Biol,* 433(2), pp. 166721.

Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F. and Sutter, M. (2018) 'Bacterial microcompartments', *Nat Rev Microbiol,* 16(5), pp. 277-290.

Kerfeld, C. A. and Melnicki, M. R. (2016) 'Assembly, function and evolution of cyanobacterial carboxysomes', *Curr Opin Plant Biol,* 31, pp. 66-75.

Kerfeld, C. A., Sawaya, M. R., Tanaka, S., Nguyen, C. V., Phillips, M., Beeby, M. and Yeates, T. O. (2005) 'Protein structures forming the shell of primitive bacterial organelles', *Science,* 309(5736), pp. 936-8.

Kim, E. Y., Slininger, M. F. and Tullman-Ercek, D. (2014) 'The effects of time, temperature, and pH on the stability of PDU bacterial microcompartments', *Protein Sci,* 23(10), pp. 1434-41.

Kim, E. Y. and Tullman-Ercek, D. (2014) 'A rapid flow cytometry assay for the relative quantification of protein encapsulation into bacterial microcompartments', *Biotechnol J,* 9(3), pp. 348-54.

Kirst, H. and Kerfeld, C. A. (2019) 'Bacterial microcompartments: catalysis-enhancing metabolic modules for next generation metabolic and biomedical engineering', *BMC Biol,* 17(1), pp. 79.

Kirst, H. and Kerfeld, C.A. (2021) ‘Clues to the function of bacterial microcompartments from ancillary genes’ *Biochem Soc Trans.,* 49(3), pp. 1085-98

Klein, M. G., Zwart, P., Bagby, S. C., Cai, F., Chisholm, S. W., Heinhorst, S., Cannon, G. C. and Kerfeld, C. A. (2009) 'Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport', *J Mol Biol,* 392(2), pp. 319-33.

Koppenol, W. H. (1993) 'The centennial of the Fenton reaction', *Free Radic Biol Med,* 15(6), pp. 645-51.

Krupovic, M. and Koonin, E. V. (2017) 'Cellular origin of the viral capsid-like bacterial microcompartments', *Biol Direct,* 12(1), pp. 25.

Kumar, G. and Sinha, S. (2021) 'Biophysical approaches to understand and re-purpose bacterial microcompartments', *Curr Opin Microbiol,* 63, pp. 43-51.

Lassila, J. K., Bernstein, S. L., Kinney, J. N., Axen, S. D. and Kerfeld, C. A. (2014) 'Assembly of robust bacterial microcompartment shells using building blocks from an organelle of unknown function', *J Mol Biol,* 426(11), pp. 2217-28.

Lawrence, A. D., Frank, S., Newnham, S., Lee, M. J., Brown, I. R., Xue, W. F., Rowe, M. L., Mulvihill, D. P., Prentice, M. B., Howard, M. J. and Warren, M. J. (2014) 'Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor', *ACS Synth Biol,* 3(7), pp. 454-465.

Leal, N. A., Havemann, G. D. and Bobik, T. A. (2003) '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,* 180(5), pp. 353-61.

Lee, M. J., Brown, I. R., Juodeikis, R., Frank, S. and Warren, M. J. (2016) 'Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in Escherichia coli', *Metab Eng,* 36, pp. 48-56.

Lee, M. J., Mantell, J., Brown, I. R., Fletcher, J. M., Verkade, P., Pickersgill, R. W., Woolfson, D. N., Frank, S. and Warren, M. J. (2018) 'De novo targeting to the cytoplasmic and luminal side of bacterial microcompartments', *Nat Commun,* 9(1), pp. 3413.

Lee, M. J., Palmer, D. J. and Warren, M. J. (2019) 'Biotechnological Advances in Bacterial Microcompartment Technology', *Trends Biotechnol,* 37(3), pp. 325-336.

Lehman, B. P., Chowdhury, C. and Bobik, T. A. (2017) 'The N Terminus of the PduB Protein Binds the Protein Shell of the Pdu Microcompartment to Its Enzymatic Core', *J Bacteriol,* 199(8).

Li, T., Jiang, Q., Huang, J., Aitchison, C. M., Huang, F., Yang, M., Dykes, G. F., He, H. L., Wang, Q., Sprick, R. S., Cooper, A. I. and Liu, L. N. (2020) 'Reprogramming bacterial protein organelles as a nanoreactor for hydrogen production', *Nat Commun,* 11(1), pp. 5448.

Li, Y., Kennedy, N. W., Li, S., Mills, C. E., Tullman-Ercek, D. and Olvera de la Cruz, M. (2021) 'Computational and Experimental Approaches to Controlling Bacterial Microcompartment Assembly', *ACS Cent Sci,* 7(4), pp. 658-670.

Liu, L.-N., Yang, M., Sun, Y. and Yang, J. (2021) ‘Protein stoichiometry, structural plasticity and regulation of bacterial microcompartments’, *Curr. Opin. Microbiol.*, 63, pp. 131-41.

Liu, Y., Jorda, J., Yeates, T. O. and Bobik, T. A. (2015) 'The PduL Phosphotransacylase Is Used To Recycle Coenzyme A within the Pdu Microcompartment', *J Bacteriol*, 197(14), pp. 2392-9.

Lundgren, B. R., Sarwar, Z., Pinto, A., Ganley, J. G. and Nomura, C. T. (2016) 'Ethanolamine Catabolism in Pseudomonas aeruginosa PAO1 Is Regulated by the Enhancer-Binding Protein EatR (PA4021) and the Alternative Sigma Factor RpoN', *J Bacteriol*, 198(17), pp. 2318-29.

Lundin, A. P., Stewart, K. L., Stewart, A. M., Herring, T. I., Chowdhury, C. and Bobik, T. A. (2020) 'Genetic Characterization of a Glycyl Radical Microcompartment Used for 1,2-Propanediol Fermentation by Uropathogenic Escherichia coli CFT073', *J Bacteriol*, 202(9).

Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W. and Huber, R. (1995) 'Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution', *Science*, 268(5210), pp. 533-9.

Mahinthichaichan, P., Morris, D. M., Wang, Y., Jensen, G. J. and Tajkhorshid, E. (2018) 'Selective Permeability of Carboxysome Shell Pores to Anionic Molecules', *J Phys Chem B*, 122(39), pp. 9110-9118.

Mallette, E. and Kimber, M. S. (2017) 'A Complete Structural Inventory of the Mycobacterial Microcompartment Shell Proteins Constrains Models of Global Architecture and Transport', *J Biol Chem*, 292(4), pp. 1197-1210.

Mallette, E. and Kimber, M. S. (2018) 'Structure and Kinetics of the S-(+)-1-Amino-2-propanol Dehydrogenase from the RMM Microcompartment of Mycobacterium smegmatis', *Biochemistry*, 57(26), pp. 3780-3789.

Martin, W. F., Garg, S. and Zimorski, V. (2015) 'Endosymbiotic theories for eukaryote origin', *Philos Trans R Soc Lond B Biol Sci*, 370(1678), pp. 20140330.

Masuda, J., Shibata, N., Morimoto, Y., Toraya, T. and Yasuoka, N. (2000) 'How a protein generates a catalytic radical from coenzyme B(12): X-ray structure of a diol-dehydratase-adeninylpentylcobalamin complex', *Structure*, 8(7), pp. 775-88.

Masuda, J., Yamaguchi, T., Tobimatsu, T., Toraya, T., Suto, K., Shibata, N., Morimoto, Y., Higuchi, Y. and Yasuoka, N. (1999) 'Crystallization and preliminary x-ray study of two crystal forms of Klebsiella oxytoca diol dehydratase-cyanocobalamin complex', *Acta Crystallogr D Biol Crystallogr*, 55(Pt 4), pp. 907-9.

Mattson, G., Conklin, E., Desai, S., Nielander, G., Savage, M. D. and Morgensen, S. (1993) 'A practical approach to crosslinking', *Mol Biol Rep*, 17(3), pp. 167-83.

Matzinger, M. and Mechtler, K. (2021) 'Cleavable Cross-Linkers and Mass Spectrometry for the Ultimate Task of Profiling Protein-Protein Interaction Networks', *J Proteome Res*, 20(1), pp. 78-93.

Mayer, M. J., Juodeikis, R., Brown, I. R., Frank, S., Palmer, D. J., Deery, E., Beal, D. M., Xue, W. F. and Warren, M. J. (2016) 'Effect of bio-engineering on size, shape, composition and rigidity of bacterial microcompartments', *Sci Rep*, 6, pp. 36899.

McKenzie-Coe, A., Montes, N. S. and Jones, L. M. (2021) 'Hydroxyl Radical Protein Footprinting: A Mass Spectrometry-Based Structural Method for Studying the Higher Order Structure of Proteins', *Chem Rev*.

Mintseris, J. and Gygi, S. P. (2020) 'High-density chemical cross-linking for modeling protein interactions', *Proc Natl Acad Sci U S A*, 117(1), pp. 93-102.

Mohajerani, F., Sayer, E., Neil, C., Inlow, K. and Hagan, M. F. (2021) 'Mechanisms of Scaffold-Mediated Microcompartment Assembly and Size Control', *ACS Nano*, 15(3), pp. 4197-4212.

Mori, K., Hieda, N., Yamanishi, M., Shibata, N. and Toraya, T. (2005) 'Crystallization and preliminary X-ray analysis of molecular chaperone-like diol dehydratase-reactivating factor in ADP-bound and nucleotide-free forms', *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 61(Pt 6), pp. 603-5.

Moses, J. E. and Moorhouse, A. D. (2007) 'The growing applications of click chemistry', *Chem Soc Rev*, 36(8), pp. 1249-62.

Murat, D., Byrne, M. and Komeili, A. (2010) 'Cell biology of prokaryotic organelles', *Cold Spring Harb Perspect Biol*, 2(10), pp. a000422.

Mädler, S., Bich, C., Touboul, D. and Zenobi, R. (2009) 'Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities', *J Mass Spectrom*, 44(5), pp. 694-706.

Nichols, R. J., Cassidy-Amstutz, C., Chaijarasphong, T. and Savage, D. F. (2017) 'Encapsulins: molecular biology of the shell', *Crit Rev Biochem Mol Biol*, 52(5), pp. 583-594.

Nichols, R. J., LaFrance, B., Phillips, N. R., Radford, D. R., Oltrogge, L. M., Valentin-Alvarado, L. E., Bischoff, A. J., Nogales, E. and Savage, D. F. (2021) 'Discovery and characterization of a novel family of prokaryotic nanocompartments involved in sulfur metabolism', *Elife*, 10.

Ninfa, A. J. and Jiang, P. (2005) 'PII signal transduction proteins: sensors of alpha-ketoglutarate that regulate nitrogen metabolism', *Curr Opin Microbiol*, 8(2), pp. 168-73.

Noël, C., Cai, F. aöd Kerfeld, C. (2016) 'Purification and Characterization of Protein Nanotubes Assembled from a Single Bacterial Microcompartment Shell Subunit', *Adv. Mater. Interfaces*, 3, pp. 1500295.

O'Reilly, F. J. and Rappsilber, J. (2018) 'Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology', *Nat Struct Mol Biol*, 25(11), pp. 1000-1008.

Ochoa, J. M. and Yeates, T. O. (2021) 'Recent structural insights into bacterial microcompartment shells', *Curr Opin Microbiol*, 62, pp. 51-60.

Oltrogge, L. M., Chaijarasphong, T., Chen, A. W., Bolin, E. R., Marqusee, S. and Savage, D. F. (2020) 'Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation', *Nat Struct Mol Biol*, 27(3), pp. 281-287.

Palacios, S., Starai, V.J., Escalante-Semerena, J.C. (2003) 'Propionyl coenzyme A is a common intermediate in the 1,2-propanediol and propionate catabolic pathways needed for expressed of the prpBCDE operon during growth of Salmonella enterica on 1,2-propanediol', *J Bacteriol*., 185(9), pp. 2802-10.

Pang, A., Frank, S., Brown, I., Warren, M. J. and Pickersgill, R. W. (2014) 'Structural insights into higher order assembly and function of the bacterial microcompartment protein PduA', *J Biol Chem*, 289(32), pp. 22377-84.

Pang, A., Liang, M., Prentice, M. B. and Pickersgill, R. W. (2012) 'Substrate channels revealed in the trimeric Lactobacillus reuteri bacterial microcompartment shell protein PduB', *Acta Crystallogr D Biol Crystallogr*, 68(Pt 12), pp. 1642-52.

Pang, A., Warren, M. J. and Pickersgill, R. W. (2011) 'Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe-4S cluster-binding site', *Acta Crystallogr D Biol Crystallogr*, 67(Pt 2), pp. 91-6.

Papapostolou, D. and Howorka, S. (2009) 'Engineering and exploiting protein assemblies in synthetic biology', *Mol Biosyst*, 5(7), pp. 723-32.

Parfrey, L. W., Lahr, D. J., Knoll, A. H. and Katz, L. A. (2011) 'Estimating the timing of early eukaryotic diversification with multigene molecular clocks', *Proc Natl Acad Sci U S A*, 108(33), pp. 13624-9.

Parry, M. A., Keys, A. J., Madgwick, P. J., Carmo-Silva, A. E. and Andralojc, P. J. (2008) 'Rubisco regulation: a role for inhibitors', *J Exp Bot*, 59(7), pp. 1569-80.

Parsons, J. B., Dinesh, S. D., Deery, E., Leech, H. K., Brindley, A. A., Heldt, D., Frank, S., Smales, C. M., Lünsdorf, H., Rambach, A., Gass, M. H., Bleloch, A., McClean, K. J., Munro, A. W., Rigby, S. E., Warren, M. J. and Prentice, M. B. (2008) 'Biochemical and structural insights into bacterial organelle form and biogenesis', *J Biol Chem*, 283(21), pp. 14366-75.

Parsons, J. B., Frank, S., Bhella, D., Liang, M., Prentice, M. B., Mulvihill, D. P. and Warren, M. J. (2010) 'Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement', *Mol Cell*, 38(2), pp. 305-15.

Perez-Riverol Y., Csordas A., Bai J., Bernal-Llinares M., Hewapathirana S., Kundu D.J., Inuganti A., Griss J., Mayer G., Eisenacher M., Pérez E., Uszkoreit J., Pfeuffer J., Sachsenberg T., Yilmaz S., Tiwary S., Cox J., Audain E., Walzer M., Jarnuczak A.F., Ternent T., Brazma A., Vizcaíno J.A. (2019) ‘The PRIDE database and related tools and resources in 2019: improving support for quantification data’, *Nucleic Acids Res.*, 47(D1), pp. D442-D450.

Petit, E., LaTouf, W. G., Coppi, M. V., Warnick, T. A., Currie, D., Romashko, I., Deshpande, S., Haas, K., Alvelo-Maurosa, J. G., Wardman, C., Schnell, D. J., Leschine, S. B. and Blanchard, J. L. (2013) 'Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by Clostridium phytofermentans', *PLoS One*, 8(1), pp. e54337.

Pitts, A. C., Tuck, L. R., Faulds-Pain, A., Lewis, R. J. and Marles-Wright, J. (2012) 'Structural insight into the Clostridium difficile ethanolamine utilisation microcompartment', *PLoS One*, 7(10), pp. e48360.

Planamente, S. and Frank, S. (2019) 'Bio-engineering of bacterial microcompartments: a mini review', *Biochem Soc Trans*, 47(3), pp. 765-777.

Pokhrel, A., Kang, S. Y. and Schmidt-Dannert, C. (2021) 'Ethanolamine bacterial microcompartments: from structure, function studies to bioengineering applications', *Curr Opin Microbiol*, 62, pp. 28-37.

Polka, J. K., Hays, S. G. and Silver, P. A. (2016) 'Building Spatial Synthetic Biology with Compartments, Scaffolds, and Communities', *Cold Spring Harb Perspect Biol,* 8(8).

Prasad, B. V. and Schmid, M. F. (2012) 'Principles of virus structural organization', *Adv Exp Med Biol,* 726, pp. 17-47.

Prentice, M. B. (2021) 'Bacterial microcompartments and their role in pathogenicity', *Curr Opin Microbiol,* 63, pp. 19-28.

Rae, B. D., Long, B. M., Badger, M. R. and Price, G. D. (2013) 'Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO2 fixation in cyanobacteria and some proteobacteria', *Microbiol Mol Biol Rev,* 77(3), pp. 357-79.

Ravcheev, D. A., Moussu, L., Smajic, S. and Thiele, I. (2019) 'Comparative Genomic Analysis Reveals Novel Microcompartment-Associated Metabolic Pathways in the Human Gut Microbiome', *Front Genet,* 10.

Riback, J. A., Zhu, L., Ferrolino, M. C., Tolbert, M., Mitrea, D. M., Sanders, D. W., Wei, M. T., Kriwacki, R. W. and Brangwynne, C. P. (2020) 'Composition-dependent thermodynamics of intracellular phase separation', *Nature,* 581(7807), pp. 209-214.

Rillema, R., Hoang, Y., MacCready, J. S. and Vecchiarelli, A. G. (2021) 'Carboxysome Mispositioning Alters Growth, Morphology, and Rubisco Level of the Cyanobacterium Synechococcus elongatus PCC 7942', *mBio,* 12(4), pp. e0269620.

Ris, H. and Singh, R. N. (1961) 'Electron microscope studies on blue-green algae', *J Biophys Biochem Cytol,* 9, pp. 63-80.

Rodriguez-Ramos, J., Faulkner, M. and Liu, L. N. (2018) 'Nanoscale Visualization of Bacterial Microcompartments Using Atomic Force Microscopy', *Methods Mol Biol,* 1814, pp. 373-383.

Saier, M. H. and Bogdanov, M. V. (2013) 'Membranous organelles in bacteria', *J Mol Microbiol Biotechnol,* 23(1-2), pp. 5-12.

Sampson, E. M. and Bobik, T. A. (2008) 'Microcompartments for B12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate', *J Bacteriol,* 190(8), pp. 2966-71.

Santiago, C. P., Quick, L. N. and Wilson, J. W. (2011) 'Self-transmissible IncP R995 plasmids with alternative markers and utility for Flp/FRT cloning strategies', *J Microbiol Biotechnol,* 21(11), pp. 1123-6.

Sauer, P., Dominguez-Martin, M., Kirst, H., Sutter, M., Bina, D., Greber, B., Nogales, E., Polívka, T. and Kerfeld, C. 2021. Structures of the Cyanobacterial Phycobilisome. bioRxiv.

Schmid, M. F., Paredes, A. M., Khant, H. A., Soyer, F., Aldrich, H. C., Chiu, W. and Shively, J. M. (2006) 'Structure of Halothiobacillus neapolitanus carboxysomes by cryo-electron tomography', *J Mol Biol,* 364(3), pp. 526-35.

Shibata, N., Masuda, J., Tobinmatsu, T., Toraya, T., Suto, K., Morimoto, Y., Yasuoka, N. (1999) 'A new mode of B12 binding and the direct participation of a potassium ion in enzyme catalysis: X-ray structure of diol dehydratase', *Structure*, 7(8), pp. 997-1008.

Shibata, N., Mori, K., Hieda, N., Higuchi, Y., Yamanishi, M. and Toraya, T. (2005) 'Release of a damaged cofactor from a coenzyme B12-dependent enzyme: X-ray structures of diol dehydratase-reactivating factor', *Structure,* 13(12), pp. 1745-54.

Shibata, N., Tamagaki, H., Hieda, N., Akita, K., Komori, H., Shomura, Y., Terawaki, S., Mori, K., Yasuoka, N., Higuchi, Y. and Toraya, T. (2010) 'Crystal structures of ethanolamine ammonia-lyase complexed with coenzyme B12 analogs and substrates', *J Biol Chem,* 285(34), pp. 26484-93.

Shih, Y. L. and Rothfield, L. (2006) 'The bacterial cytoskeleton', *Microbiol Mol Biol Rev,* 70(3), pp. 729-54.

Shively, J. (1974) 'Inclusion bodies of prokaryotes' *Annu. Rev. Microbiol*., pp. 167-185.

Shively, J. M., Ball, F., Brown, D. H. and Saunders, R. E. (1973) 'Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of Thiobacillus neapolitanus', *Science,* 182(4112), pp. 584-6.

Sinha, S., Cheng, S., Fan, C. and Bobik, T. A. (2012) 'The PduM protein is a structural component of the microcompartments involved in coenzyme B(12)-dependent 1,2-propanediol degradation by Salmonella enterica', *J Bacteriol,* 194(8), pp. 1912-8.

Sinz, A. (2017) 'Divide and conquer: cleavable cross-linkers to study protein conformation and protein-protein interactions', *Anal Bioanal Chem,* 409(1), pp. 33-44.

Spicer, C. D. and Davis, B. G. (2014) 'Selective chemical protein modification', *Nat Commun,* 5, pp. 4740.

Stewart, A. M., Stewart, K. L., Yeates, T. O. and Bobik, T. A. (2021) 'Advances in the World of Bacterial Microcompartments', *Trends Biochem Sci,* 46(5), pp. 406-416.

Sturms, R., Streauslin, N. A., Cheng, S. and Bobik, T. A. (2015) 'In Salmonella enterica, Ethanolamine Utilization Is Repressed by 1,2-Propanediol To Prevent Detrimental Mixing of Components of Two Different Bacterial Microcompartments', *J Bacteriol,* 197(14), pp. 2412-21.

Sutter, M., Boehringer, D., Gutmann, S., Günther, S., Prangishvili, D., Loessner, M. J., Stetter, K. O., Weber-Ban, E. and Ban, N. (2008) 'Structural basis of enzyme encapsulation into a bacterial nanocompartment', *Nat Struct Mol Biol,* 15(9), pp. 939-47.

Sutter, M., Faulkner, M., Aussignargues, C., Paasch, B. C., Barrett, S., Kerfeld, C. A. and Liu, L. N. (2016) 'Visualization of Bacterial Microcompartment Facet Assembly Using High-Speed Atomic Force Microscopy', *Nano Lett,* 16(3), pp. 1590-5.

Sutter, M., Greber, B., Aussignargues, C. and Kerfeld, C. A. (2017) 'Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell', *Science,* 356(6344), pp. 1293-1297.

Sutter, M., Laughlin, T. G., Sloan, N. B., Serwas, D., Davies, K. M. and Kerfeld, C. A. (2019a) 'Structure of a Synthetic beta-Carboxysome Shell', *Plant Physiol,* 181(3), pp. 1050-1058.

Sutter, M., McGuire, S., Ferlez, B. and Kerfeld, C. A. (2019b) 'Structural Characterization of a Synthetic Tandem-Domain Bacterial Microcompartment Shell Protein Capable of Forming Icosahedral Shell Assemblies', *ACS Synth Biol*, 8(4), pp. 668-674.

Sutter, M., Melnicki, M. R., Schulz, F., Woyke, T. and Kerfeld, C. A. (2021) 'A catalog of the diversity and ubiquity of bacterial microcompartments', *Nat Commun*, 12(1), pp. 3809.

Takenoya, M., Nikolakakis, K. and Sagermann, M. (2010) 'Crystallographic insights into the pore structures and mechanisms of the EutL and EutM shell proteins of the ethanolamine-utilizing microcompartment of Escherichia coli', *J Bacteriol*, 192(22), pp. 6056-63.

Tan, Y. Q., Ali, S., Xue, B., Teo, W. Z., Ling, L. H., Go, M. K., Lv, H., Robinson, R. C., Narita, A. and Yew, W. S. (2021) 'Structure of a Minimal α-Carboxysome-Derived Shell and Its Utility in Enzyme Stabilization', *Biomacromolecules*, 22(10), pp. 4095-4109.

Tanaka, S., Kerfeld, C. A., Sawaya, M. R., Cai, F., Heinhorst, S., Cannon, G. C. and Yeates, T. O. (2008) 'Atomic-level models of the bacterial carboxysome shell', *Science*, 319(5866), pp. 1083-6.

Thompson, M. C., Crowley, C. S., Kopstein, J., Bobik, T. A. and Yeates, T. O. (2014) 'Structure of a bacterial microcompartment shell protein bound to a cobalamin cofactor', *Acta Crystallogr F Struct Biol Commun*, 70(Pt 12), pp. 1584-90.

Tjernberg, A., Markova, N., Griffiths, W. J. and Hallén, D. (2006) 'DMSO-related effects in protein characterization', *J Biomol Screen*, 11(2), pp. 131-7.

Toraya, T. and Mori, K. (1999) 'A reactivating factor for coenzyme B12-dependent diol dehydratase', *J Biol Chem*, 274(6), pp. 3372-7.

Toseland, C. P. (2013) 'Fluorescent labeling and modification of proteins', *J Chem Biol*, 6(3), pp. 85-95.

Tsai, Y., Sawaya, M. R., Cannon, G. C., Cai, F., Williams, E. B., Heinhorst, S., Kerfeld, C. A. and Yeates, T. O. (2007) 'Structural analysis of CsoS1A and the protein shell of the Halothiobacillus neapolitanus carboxysome', *PLoS Biol*, 5(6), pp. e144.

Uddin, I., Frank, S., Warren, M. J. and Pickersgill, R. W. (2018) 'A Generic Self-Assembly Process in Microcompartments and Synthetic Protein Nanotubes', *Small*, 14(19), pp. e1704020.

van Niftrik, L. A., Fuerst, J. A., Sinninghe Damsté, J. S., Kuenen, J. G., Jetten, M. S. and Strous, M. (2004) 'The anammoxosome: an intracytoplasmic compartment in anammox bacteria', *FEMS Microbiol Lett*, 233(1), pp. 7-13.

van Steensel, B., van Binnendijk, E. P., Hornsby, C. D., van der Voort, H. T., Krozowski, Z. S., de Kloet, E. R. and van Driel, R. (1996) 'Partial colocalization of glucocorticoid and mineralocorticoid receptors in discrete compartments in nuclei of rat hippocampus neurons', *J Cell Sci*, 109 (Pt 4), pp. 787-92.

Vreven, T., Schweppe, D. K., Chavez, J. D., Weisbrod, C. R., Shibata, S., Zheng, C., Bruce, J. E. and Weng, Z. (2018) 'Integrating Cross-Linking Experiments with Ab Initio Protein-Protein Docking', *J Mol Biol*, 430(12), pp. 1814-1828.

Wade, Y., Daniel, R. A. and Leak, D. J. (2019) 'Heterologous Microcompartment Assembly in *Bacillaceae*: Establishing the Components Necessary for Scaffold Formation', *ACS Synth Biol,* 8(7), pp. 1642-1654.

Wagner, H. J., Capitain, C. C., Richter, K., Nessling, M. and Mampel, J. (2017) 'Engineering bacterial microcompartments with heterologous enzyme cargos', *Eng Life Sci,* 17(1), pp. 36-46.

Wang, H., Yan, X., Aigner, H., Bracher, A., Nguyen, N. D., Hee, W. Y., Long, B. M., Price, G. D., Hartl, F. U. and Hayer-Hartl, M. (2019) 'Rubisco condensate formation by CcmM in β-carboxysome biogenesis', *Nature,* 566(7742), pp. 131-135.

Wheatley, N. M., Gidaniyan, S. D., Liu, Y., Cascio, D. and Yeates, T. O. (2013) 'Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins', *Protein Sci,* 22(5), pp. 660-5.

Wilson, J. W. (2021) 'Manipulating microcompartment operons to study mechanism and function', *Curr Opin Microbiol,* 60, pp. 66-72.

Yang, M., Simpson, D. M., Wenner, N., Brownridge, P., Harman, V. M., Hinton, J. C. D., Beynon, R. J. and Liu, L. N. (2020) 'Decoding the stoichiometric composition and organisation of bacterial metabolosomes', *Nat Commun,* 11(1), pp. 1976.

Yeates, T. O., Crowley, C. S. and Tanaka, S. (2010) 'Bacterial microcompartment organelles: protein shell structure and evolution', *Annu Rev Biophys,* 39, pp. 185-205.

Yeates, T. O., Kerfeld, C. A., Heinhorst, S., Cannon, G. C. and Shively, J. M. (2008) 'Protein-based organelles in bacteria: carboxysomes and related microcompartments', *Nat Rev Microbiol,* 6(9), pp. 681-91.

Yeates, T. O., Thompson, M. C. and Bobik, T. A. (2011) 'The protein shells of bacterial microcompartment organelles', *Curr Opin Struct Biol,* 21(2), pp. 223-31.

Yuan, S., Wang, J., Zhu, D., Wang, N., Gao, Q., Chen, W., Tang, H., Zhang, X., Liu, H., Rao, Z. and Wang, X. (2018) 'Cryo-EM structure of a herpesvirus capsid at 3.1 Å', *Science,* 360(6384).

Yung, M. C., Bourguet, F. A., Carpenter, T. S. and Coleman, M. A. (2017) 'Re-directing bacterial microcompartment systems to enhance recombinant expression of lysis protein E from bacteriophage ϕX174 in Escherichia coli', *Microb Cell Fact,* 16(1), pp. 71.

Zakeri, B., Fierer, J. O., Celik, E., Chittock, E. C., Schwarz-Linek, U., Moy, V. T. and Howarth, M. (2012) 'Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin', *Proc Natl Acad Sci U S A,* 109(12), pp. E690-7.

Zang, K., Wang, H., Hartl, F. U. and Hayer-Hartl, M. (2021) 'Scaffolding protein CcmM directs multiprotein phase separation in β-carboxysome biogenesis', *Nat Struct Mol Biol,* 28(11), pp. 909-922.

Zarzycki, J., Sutter, M., Cortina, N. S., Erb, T. J. and Kerfeld, C. A. (2017) 'In Vitro Characterization and Concerted Function of Three Core Enzymes of a Glycyl Radical Enzyme - Associated Bacterial Microcompartment', *Sci Rep,* 7, pp. 42757.

Zhao, Y. Y., Jiang, Y. L., Chen, Y., Zhou, C. Z. and Li, Q. (2019) 'Crystal structure of pentameric shell protein CsoS4B of Halothiobacillus neapolitanus α-carboxysome', *Biochem Biophys Res Commun,* 515(3), pp. 510-515.
Zimorski, V., Ku, C., Martin, W. F. and Gould, S. B. (2014) 'Endosymbiotic theory for organelle origins', *Curr Opin Microbiol,* 22, pp. 38-48.