Connecting a Bacterial Organelle to Its Positioning System

by

Joseph L. Basalla

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular, Cellular, and Developmental Biology)
in the University of Michigan
2024

Doctoral Committee:

Assistant Professor Anthony G. Vecchiarelli, Chair
Professor James C. A. Bardwell
Associate Professor Allen P. Liu
Professor Lyle A. Simmons

Joseph L. Basalla

jbasalla@umich.edu

ORCID iD: 0009-0006-5444-344X

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Anthony Vecchiarelli. Every aspect of my professional skillset has benefitted from Anthony's guidance over the years. I have become a better writer, a more critical thinker, a more effective communicator, and a more thoughtful mentor all due to my time spent in his lab. And above all else, I am grateful for his enthusiasm. I would not have half the academic success that I have now if not for Anthony's contagious energy and excitement for potential future directions. Thank you, Anthony!

I would also like to thank my committee members, Dr. James Bardwell, Dr. Lyle Simmons, and Dr. Allen Liu. My committee meetings have always had a good balance of enthusiasm and constructive criticism. Both of my first author papers as well as this dissertation are stronger because of our meetings and discussions. I have gained a lot of perspective from our meetings that will continue to influence how I view my own work, and I thank you all for that!

I would also like to thank Dr. Matthew Chapman and Dr. Hans Wildschutte. Matt helped recruit me to Michigan and was constantly supportive during my time here. Thank you, Matt! And Hans helped foster my interest in research and develop the skills that got me accepted into graduate school at Michigan. Thank you, Hans!

All the postdocs, graduate students, and undergraduates that I met while in Anthony's lab have been an important part of my graduate experience, and I am thankful to them all. But I would especially like to thank the cohort of people in Anthony's lab when I joined. Anne, Lisa,

and Josh were all great scientists who I benefited from being around, but also great people who I enjoyed being around. This group will always be an important part of my memories of graduate school, and I am happy to have worked with you all.

And lastly, all the friends I met while I was here as well as my family. I am incredibly lucky to have a network of care and support. Thank you all!

Table of Contents

List of Abbreviations

Δ	Deletion
ATP	Adenosine-5'-Triphosphate
BMC	Bacterial microcompartment
CC	Coiled-coil
CCM	Carbon concentrating mechanism
CD	Circular dichroism
CTD	C-terminal domain
C-terminal	Carboxy terminal
DLS	Dynamic light scattering
DNA	Deoxyribonucleic acid
FRAP	Fluorescence-recovery after photobleaching
Hn	*Halothiobacillus neapolitanus* c2
IDR	Intrinsically disordered region
IPTG	Isopropyl-beta-D-thiogalactopyranoside
I-TASSER	Iterative Threading ASSEmbly Refinement
LLPS	Liquid-liquid phase separation
McdA	Maintenance of carboxysome distribution protein A
McdB	Maintenance of carboxysome distribution protein B
mNG	monomericNeonGreen
mTQ	monomericTurquoise2
NMR	Nuclear magnetic resonance
NTD	N-terminal domain
N-terminal	Amino terminal
PCC	Pearson's correlation coefficient
PDB	Protein databank
PEG	Polyethylene glycol
PONDR	Predictor of Natural Disorder Regions
PSCP	Phase separation coupled to percolation
Rubisco	Ribulose-1,5-bisphosphate carboxylase/oxygenase
Se	*Synechococcus elongatus* PCC 7942
SEC-MALS	Size Exclusion Chromatography–Multiple Angle Light Scattering

List of Figures

Abstract

In bacteria, protein-based organelles called bacterial microcompartments (BMCs) are a widespread strategy for subcellular organization. BMCs encapsulate enzymes within a selectively permeable protein shell to drive unfavorable reactions and sequester intermediates. Through this mechanism, BMCs regulate a diversity of metabolisms across bacteria, including human pathogens. The study of BMC biology is therefore important for understanding aspects of bacterial cell biology and metabolism, its consequences to human health, and the development of *in vivo* encapsulation biotechnologies.

The model BMC is the carboxysome, which helps drive carbon-fixation in cyanobacteria and some chemoautotrophs. Carboxysomes are a paradigm for understanding fundamental aspects of BMC biology, including their spatial organization in the cell. Carboxysome spatial organization results in their uniform distribution in the cell, and disruptions to this organization cause carboxysome aggregation, decreased carbon-fixation, and slower cell growth. Thus, understanding how cells spatially distribute carboxysomes is necessary for our understanding and application of functional and efficient BMCs.

Our lab recently identified the two-protein system which spatially organizes carboxysomes, named the maintenance of carboxysome distribution (Mcd) system, consisting of the proteins McdA and McdB. McdA is the positioning ATPase that drives active carboxysome distribution, but does not interact directly with carboxysomes. Instead, McdB associates with carboxysomes, acting as an adaptor to link carboxysomes to the positioning ATPase, McdA. We

now know that McdAB systems are widespread in BMC-containing bacteria, yet how different McdB proteins associate with their respective BMCs remains to be determined. McdB thus represents a novel, widespread, but unstudied class of proteins.

To address this gap, my thesis work began with a biochemical characterization of McdB proteins from several carboxysome-containing bacteria. Intriguingly, all purified McdB proteins formed condensates *in vitro*. Condensates are the result of molecules undergoing a density transition to form two coexisting phases: a dense, solvent-poor condensate phase and a dilute soluble phase. Condensate formation has now been implicated in a diversity of biological processes across eukaryotes and prokaryotes. My thesis work addresses two important questions critical to our understanding of the mechanisms governing the spatial organization of carboxysomes, and BMCs in general: (1) What is the functional consequence of McdB condensate formation in the spatial organization of carboxysomes and (2) What are the molecular features of McdB proteins that specify their interactions with a specific BMC-type.

To answer these questions, I first dissected the condensate formation and oligomerization activities of McdB, and provided evidence suggesting that McdB condensation plays a role in its association with carboxysomes. Second, I identified C-terminal motifs containing an invariant tryptophan necessary for McdB proteins to associate with their respective carboxysomes. Substituting this tryptophan with other aromatic residues reveals a gradient of carboxysome colocalization by McdB, and a corresponding gradient in carboxysome positioning activity *in vivo*. Intriguingly, these activity gradients correlated with the ability of McdB to form condensates *in vitro*. Together, my thesis reveals a common mechanism underlying adaptor protein binding for carboxysomes, and possibly for other BMCs that use McdAB-like systems for their active positioning. My findings advance our understanding of subcellular organization

in bacteria and, more specifically, the mechanisms governing the spatial regulation of protein-based organelles across the bacterial world.

Chapter 1

Positioning the Model Bacterial Organelle, the Carboxysome

1.1 Outline of the Dissertation

My research dissertation lies at the intersection of three subfields: (1) spatial regulation systems in bacteria, (2) bacterial organelles, and (3) biomolecular condensates. In the following two chapters, I provide brief introductions to each of these fields, highlighting specific considerations that impact the generalizability and choice of direction for my main data chapters. In Chapter 1, I introduce the protein of interest for my studies, named the maintenance of carboxysome distribution protein B (McdB). I describe our recent identification of this protein, its widespread prevalence in bacteria, and its relationship to organelle trafficking systems in bacteria. I also note an intriguing property of McdB proteins that I determined early on, their ability to form liquid-like condensates *in vitro*. I then provide background for biomolecular condensates in Chapter 2, and specifically detail my choice of terminology and methodology for studying McdB function through the lens of condensate formation. I briefly summarize the numerous processes in bacteria whereby functional condensates have been implicated, and how these studies motivated my investigation into the functionality of McdB condensate formation in organelle trafficking. My thesis research on McdB biochemistry, condensate formation, and mechanism of function is then detailed in Chapters 3 and 4. This research has contributed to each of the subfields described above, and provides a groundwork for future studies discussed in Chapter 5.

1.2 The ParA/MinD-family of ATPases positions structures in bacteria

The ParA/MinD-family of positioning ATPases is responsible for the spatial regulation of several genetic- and protein-based cargos across the bacterial world [1, 2, 3, 4]. The two ATPases for which the family is named, ParA and MinD, are the best studied. ParA ATPases are known to partition plasmids and segregate chromosomes prior to cell division, thus ensuring faithful inheritance. MinD, on the other hand, is involved in aligning the divisome at midcell thus preventing asymmetric cell division and the formation of mini-cells. Less studied is the growing list of ParA/MinD-family ATPases, widespread across prokaryotes, involved in spatially regulating diverse protein-based complexes and organelles, such as flagella [5], chemotaxis clusters [6], conjugation machinery [7], and bacterial microcompartments (BMCs) [8, 9] (Figure 1.1A).

Spatial regulation by the ParA/MinD family of positioning ATPases typically relies on two proteins: an ATPase that provides active positioning and an adaptor protein that provides specificity for the cargo [1, 2, 4]. It is becoming apparent that this family of ATPases, along with their adaptor proteins, follow a general mechanism despite positioning such disparate cellular cargos. For example, consider the mechanism by which the ATPase ParA works to segregate chromosomes during cell division [1, 2]. Upon ATP-binding, ParA forms an ATP-sandwich homodimer (Figure 1.1B). ParA dimers bind the nucleoid via non-specific DNA interactions. Dimerization also forms a binding site for the adaptor protein, ParB, which can stimulate ParA ATPase activity. ParB additionally binds to a specific DNA binding site, called parS, to form a massive complex on the chromosome near the origin of replication (Figure 1.1C). This complex locally stimulates ParA ATPase activity and nucleoid release, which generates ParA gradients on

the nucleoid (Figure 1.1C). Segregation ensues as sister chromosomes chase and release

nucleoid-bound ParA in opposite directions. Without ParABS, DNA is asymmetrically inherited,

resulting in anucleate and polyploid cells, and reduced cell fitness or death. This general

mechanism is summarized in Figure 1.1D.

In the mechanism outlined in Figure 1.1D, the adaptor protein plays two critical roles: (1)

it interacts with the ATPase and stimulates ATP hydrolysis to drive the reaction forward, and (2)

it binds to the cargo to link the reaction to cargo positioning. Data across the field supports the

idea that nearly all known adaptor proteins interact with their ATPases via a positively charged

and disordered N-terminus [10, 11]. The rest of the adaptor protein is dedicated to cargo

association. This means that, while adaptor proteins interact with their ATPases via a similar

mechanism, they recognize and bind to their cargos via distinct mechanisms. Determining the

features that specify the interactions between different adaptor proteins and their cargos has

therefore been critical for advancing our understanding of spatial regulation in bacteria [10]. One

of the main goals of my research was to identify the specificity determinants by which a newly

identified class of adaptor proteins associates with their cargos, bacterial microcompartments.
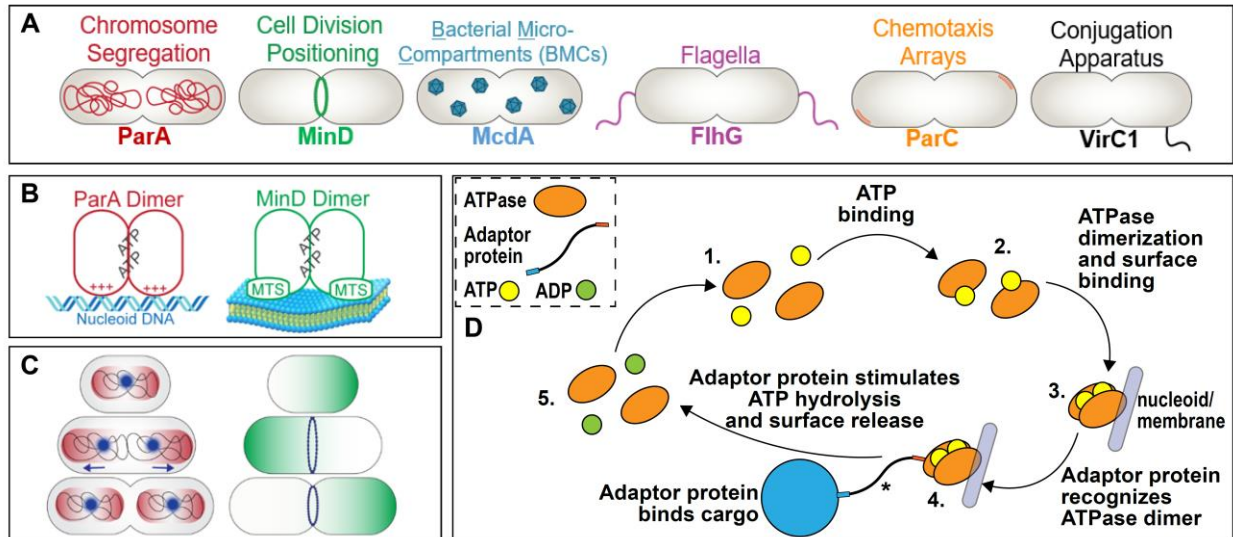
Figure 1.1. Adaptor proteins play critical roles in the spatial regulation of cargos by ParA/MinD-family ATPases. (**A**) A variety of cellular cargos (*top*) are spatially regulated via ParA/MinD-family ATPases (*bottom*). (**B**) All known ParA/MinD-family ATPases form ATP-dependent homodimers. In the dimer form, the ATPase nonspecifically binds a surface which it uses to position the associated cargos. For example, ParA dimers bind nucleoid DNA, and MinD dimers bind the membrane. (**C**) Interactions between the ParA/MinD-family ATPases and their adaptor proteins result in ATPase gradients that help drive positioning. (*left*) The adaptor protein ParB complexes with *parS* sites (blue dot) and stimulates the local release of ParA (red) from the nucleoid. Resulting ParA gradients segregate replicated chromosomes. (*right*) MinD oscillations arise from interactions with the adaptors MinC and MinE. Oscillations align the divisome at midcell. (**D**) Schematic of ParA/MinD-family ATPase cycle with major steps numbered 1-5. *Note the adaptor protein plays 2 key roles: i) stimulating ATP hydrolysis to continuously drive the positioning reaction, and ii) linking the reaction cycle to a cargo to drive its positioning. Much of the work described in this dissertation has focused on dissecting the latter for a newly identified adaptor protein.

1.3 Bacterial microcompartments (BMCs) are widespread protein-based organelles

1.3.1 BMCs are nanoscale compartments that contain enzymes encapsulated within a protein shell

In bacteria, protein-based organelles called bacterial microcompartments (BMCs) are a widespread strategy for subcellular compartmentalization [12]. A recent study found 68 unique BMC types across 45 bacterial phyla and ~20% of all sequenced bacterial genomes, showing

BMCs are prevalent across the bacterial kingdom. [13]. In general, BMCs function to encapsulate a set of enzymes within a selectively-permeable protein shell (Figure 1.2A). Encapsulation facilitates the co-concentration of enzymes with their substrates to drive unfavorable reactions as well as sequester toxic intermediates from diffusing in the cytoplasm [12].

While the enzyme sets encapsulated within different BMCs are diverse, the outer shells are comprised of a set of proteins with high sequence and structural conservation [12, 13]. Shell proteins fall into one of three oligomeric categories: hexamers (BMC-H), pentamers (BMC-P), or trimers (BMC-T) (Figure 1.2B) [12, 13]. Notably, the structures of these shell proteins give rise to selectively permeable central pores where metabolites can cross into and out of the BMC (Figure 1.2B) [12, 13]. It has been proposed that the structural and sequence variability at the pore regions of different BMCs dictates substrate preference and selective permeability for different metabolites [13]. Therefore, aside from the pore regions, BMC shells have a high degree of homology with one another [12, 13] Here, I focus on investigating the mechanism by which an adaptor protein interacts with BMCs to drive their spatial regulation. We use a model BMC called the carboxysome to investigate this interaction. As described below, we have evidence that this adaptor protein interacts with the carboxysome shell to drive its spatial regulation. Knowing that BMC shells have a high degree of homology, I expect that my investigations into interactions at the carboxysome shell will be generalizable to BMCs at large.
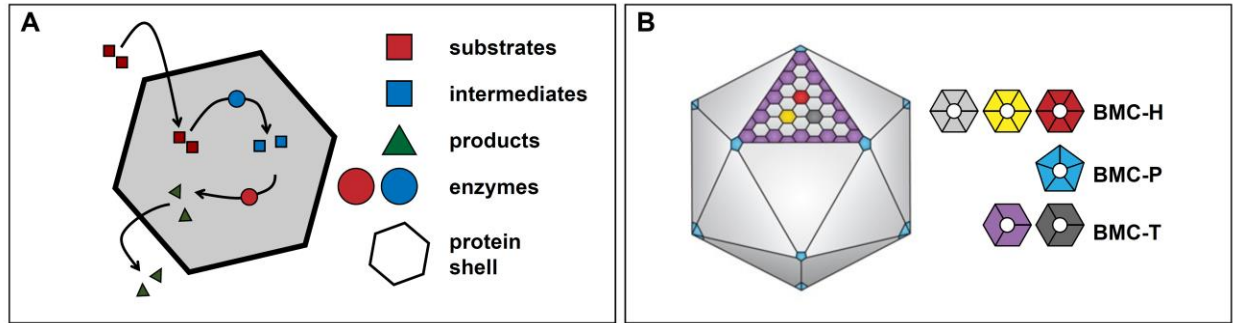
Figure 1.2. Overview of bacterial microcompartments (BMCs). (**A**) Schematic of a BMC. A specific set of enzymes is encapsulated within a protein shell, which is selectively permeable to substrates and products, but not intermediates. (**B**). Graphic of a BMC shell with once facet of the shell showing details of the component proteins. The shells of all known BMCs are made from the same groups of protein homologs, which can be classified as BMC shell hexamers (BMC-H), pentamers (BMC-P), or trimers (BMC-T). Classifications of shell components is colored to relate to the graphic. Note the central pore shown for each of the shell proteins, which can mediate metabolite flux. Graphic was adapted from MacCready, Basalla, and Vecchiarelli, *Mol Bio Evo* (2020) [14].

1.3.2 Carboxysomes are the model BMC

Out of the 68 identified BMC types, the carboxysome is by far the best studied [15]. Carboxysomes are found within many autotrophic bacteria, and function to encapsulate the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) with its substrate CO_2 to drive carbon fixation. Due to their prevalence and effect on Rubisco efficiency, carboxysomes are estimated to account for 35% of all global carbon-fixation [16, 17]. Studies on fundamental aspects of carboxysomes, such as their structure, assembly, shell permeability, and spatial regulation, have served as the basis for our general understanding of BMC biology [15]. Furthermore, carboxysomes are also of interest in synthetic biology for their potential to confer efficient carbon-fixation to cells, leading to efforts to express functional carboxysomes in both bacteria and plants [18, 19]. Therefore, efforts to understand carboxysomes are important for

deepening our understanding of BMCs in general as well as advancing carbon-capture technologies.

Carboxysomes have two subtypes, α and ß [20]. ß-carboxysomes are found in ß-cyanobacteria and α-carboxysomes are found in numerous phylogenetically distinct groups, including α-cyanobacteria and several types of proteobacteria [13]. While functionally equivalent, α- and ß-carboxysomes are thought to be the product of convergent evolution, where phylogenetic trees of BMC-P shell proteins suggest that α-carboxysomes are more closely related to other BMC types than to ß-carboxysomes [13, 20]. α- and ß-carboxysomes also have some important differences in their assembly pathways, enzyme subtypes, and modes of regulation [15, 20]. Therefore, α- and ß-carboxysomes represent evolutionarily distinct BMC types, making comparative studies between the two useful for generalizing observations to other BMCs.

At the beginning of my graduate studies, the ParA/Mind-family ATPase and adaptor protein responsible for spatially regulating ß-carboxysomes were first identified. Whether this system existed for α-carboxysomes and other BMC types was of immediate interest. Below, I outline the data we collected to identify and characterize this positioning system, and determine its prevalence in BMC-containing bacteria. These studies were important for setting the stage for my papers outlined in Chapters 3 and 4, focused on investigating the mechanism of function for the adaptor protein McdB.

1.4 Maintenance of Carboxysome Distribution systems spatially regulate both α- and ß-carboxysomes

1.4.1 ß-carboxysomes are spatially regulated by a ParA/MinD-family ATPase and its adaptor protein

Our current understanding of ß-carboxysome structure, function, and regulation has largely been through studies in the model rod-shaped cyanobacterium *Synechococcus elongatus* PCC 7942 (hereafter "Se") [15]. In Se, carboxysomes are uniformly distributed across the cell length, which ensures that daughter cells inherit an equal complement of this vital carbon-fixing organelle following cell division [12, 15, 21]. The mechanism by which ß-carboxysomes are spatially regulated, and whether this mechanism can be generalized to the spatial regulation of α-carboxysomes and other BMC types, is a recent area of research [21]. The Vecchiarelli lab recently found that, in Se, the spatial regulation of ß-carboxysomes is dependent upon a two-protein system, which involves a ParA/MinD-family ATPase we named the maintenance of carboxysome distribution protein A, or McdA, and its adaptor protein, McdB [8]. Both McdA and McdB were determined to be necessary for the distribution and functionality of ß-carboxysomes, where deletion of either protein caused carboxysome aggregation, decreased carbon fixation, and caused slower cell growth [22]. Similar to other ParA/MinD-family ATPases, McdA does not interact directly with its cargo. Instead, McdB localizes to carboxysomes to drive their spatial regulation, thus serving as the adaptor protein linking McdA to the carboxysome cargo [8].

Several pieces of data were collected to outline the general mechanism by which McdA and McdB function together to spatially regulate carboxysomes. First, McdA was determined to non-specifically bind DNA when in its ATP-bound form [8]. Additionally, a bacterial two-hybrid

analysis was used to determine that McdB interacts with itself, with McdA, and with several different carboxysome shell proteins [8]. McdB was also determined to stimulate McdA ATPase activity and drive the release of McdA from DNA *in vitro* [8]. Together, this data led to the proposed mechanism by which McdA and McdB function to spatially regulate carboxysomes, summarized in Figure 1.3. Briefly, McdB-bound carboxysomes locally stimulate the release of McdA, generating dynamic gradients on the nucleoid that drive the equidistant positioning of carboxysomes across the length of the nucleoid [8]. The findings from this paper therefore revealed a novel ParA/MinD-family ATPase (McdA) and its adaptor protein (McdB) that together are responsible for spatially regulating ß-carboxysomes. Furthermore, the newly discovered McdB represents a protein that can interact with several shell proteins from the ß-carboxysome, which is a previously unexplored phenomenon with potential implications for our general understanding of BMCs. We were therefore interested in determining if McdAB-like systems existed for α-carboxysomes, and how the respective McdB proteins localize to carboxysomes.
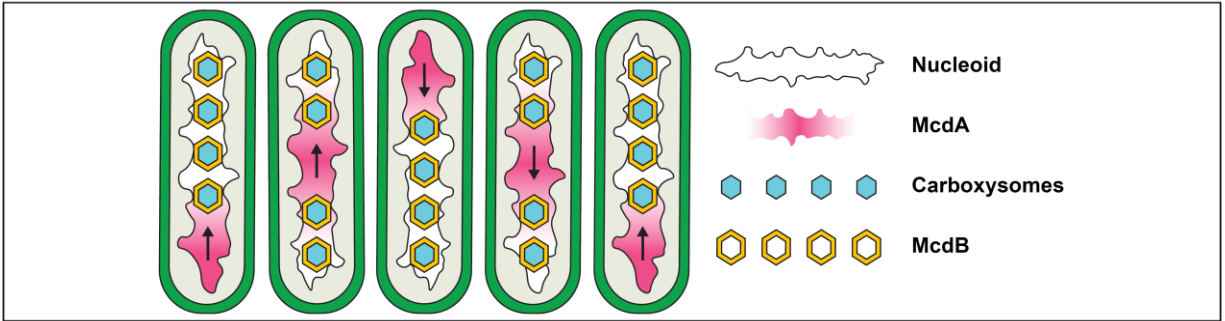
Figure 1.3. General mechanism of carboxysome distribution via McdA and McdB.
McdB-bound carboxysomes stimulate the ATPase activity of nucleoid-bound McdA dimers, resulting in the local release of McdA from the nucleoid. McdB-bound carboxysomes then move toward higher concentrations of McdA, where they further stimulate McdA activity and release. This interplay between nucleoid-bound McdA and carboxysome-bound McdB generates dynamic McdA gradients that drive the uniform spacing of carboxysomes via a Brownian-ratchet mechanism [8]. Graphic adapted from Joshua S. MacCready.

1.4.2 McdA and McdB proteins are widespread in BMC-containing bacteria, and function to position α-carboxysomes

To identify *mcdA*- and *mcdB*-like genes in the genomes of BMC-containing bacteria, we took a gene neighborhood-based approach [14, 23]. ParA/MinD-family ATPases have a high degree of homology with one another, but adaptor proteins can be highly diverse since the bind a diverse set of cargos. Furthermore, it is known that ParA/MinD-family ATPases are typically encoded in regions of the chromosome near the genes encoding the cargos they position [4]. Therefore, we first identified *mcdA*-like genes neighboring operons that encode ß-carboxysomes [14] and α-carboxysomes [23]. We then identified open reading frames within the same operon as the *mcdA*-like genes, and designated these as putative *mcdB*-like genes. Recall that adaptor proteins typically interact with their cognate ParA/MinD-family ATPase via positively charged residues in their N-termini [11]. We therefore verified that the putative *mcdB*-like genes contained an enrichment of positive charges in their N-termini. We extended this approach to

bacteria with BMCs other than carboxysomes and identified hundreds of *mcdA*- and *mcdB*-like genes near operons encoding various other BMC types [23]. From this analysis, we were able to show that putative McdAB systems are widespread across BMC-containing bacteria.

To validate these findings, we experimentally verified the function of the putative McdA- and McdB-like proteins in the proteobacterium *Halothiobacillus neapolitanus* c2 (hereafter "Hn"), which is a model for the study of α-carboxysomes [24]. Here, we found that the proteins in Hn are functional homologs to those found in Se, with McdA being a ParA/MinD-family ATPase and McdB serving as the adaptor protein [23]. Although these proteins are functional homologs, the McdA proteins that position α-carboxysomes (α-McdA) have key differences in sequence from the McdA proteins that position ß-carboxysomes (ß-McdA) [23]. The α- and ß-McdB proteins also have little sequence conservation. The findings suggest that α- and ß-McdAB systems have distinct evolutionary lineages, similar to α- and ß-carboxysomes having distinct evolutionary lineages [20].

Through these studies, we identified a ParA/MinD-family ATPase (McdA) and its adaptor protein (McdB) that are responsible for spatially regulating both α- and ß-carboxysomes, and potentially many other BMCs. However, although functionally equivalent, key differences in sequence and phylogeny suggest that α-McdAB may have a different underlying mechanism than ß-McdAB. This would not be entirely surprising, since α-carboxysomes and ß-carboxysomes are also phylogenetically distinct, and differ in fundamental processes such as assembly [20]. Therefore, early in my graduate studies, an open and influential question was whether α-McdB and ß-McdB proteins bind to their carboxysomes via a shared or distinct mechanism. This was the aim of my research described in Chapter 4.

1.5 McdB and other carboxysome-related proteins form condensates *in vitro*

Lastly, after determining that McdB proteins were widespread among BMC-containing bacteria, I purified several McdB proteins across several diverse bacterial species. Intriguingly, many formed liquid-like condensates *in vitro*, including those from Se and Hn [14, 23]. Notably, around this time, the carboxysome core enzyme Rubisco from both Se and Hn was also found to form condensates *in vitro* along with other carboxysome core components [25, 26]. Together, this suggested that condensate formation may play a key role in the assembly pathway of α- and ß-carboxysomes, and potentially be a key factor in BMC assembly in general. Many in the field then began thinking of how protein condensate formation might influence the structure, assembly, and spatial organization of carboxysomes and BMCs [27, 28]. Some additional motivations for me at the beginning of my dissertation research were therefore to determine how McdB proteins formed condensates *in vitro*, whether we could make mutations to specifically alter this process, and whether condensate formation contributed to McdB association with carboxysomes *in vivo.* This was the aim of my research described in Chapter 3. In the following chapter, I outline some of the key concepts, terminologies, methodologies, and findings needed to appreciate my data described in Chapter 3.

1.6 References

1.   J. Lutkenhaus, The ParA/MinD family puts things in their place. *Trends Microbiol.* **20**, 411–418 (2012).
2.   A. G. Vecchiarelli, K. Mizuuchi, B. E. Funnell, Surfing biological surfaces: exploiting the nucleoid for partition and transport in bacteria. *Mol. Microbiol.* **86**, 513–523 (2012).
3.   E. V. Koonin, A Superfamily of ATPases with Diverse Functions Containing Either Classical or Deviant ATP-binding Motif. *J. Mol. Biol.* **229**, 1165–1174 (1993).
4.   L. T. Pulianmackal, *et al.*, Multiple ParA/MinD ATPases coordinate the positioning of disparate cargos in a bacterial cell. *Nat. Commun.* **14**, 1–15 (2023).

5. J. S. Schuhmacher, *et al.*, MinD-like ATPase FlhG effects location and number of bacterial flagella during C-ring assembly. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3092–3097 (2015).

6. S. Ringgaard, K. Schirner, B. M. Davis, M. K. Waldor, A family of ParA-like ATPases promotes cell pole maturation by facilitating polar localization of chemotaxis proteins. *Genes Dev.* **25**, 1544–1555 (2011).

7. K. Atmakuri, E. Cascales, O. T. Burton, L. M. Banta, P. J. Christie, Agrobacterium ParA/MinD-like VirC1 spatially coordinates early conjugative DNA transfer reactions. *EMBO J.* **26**, 2540–2551 (2007).

8. J. S. MacCready, *et al.*, Protein gradients on the nucleoid position the carbon-fixing organelles of cyanobacteria. *eLife* **7** (2018).

9. P. Hakim, Y. Hoang, A. G. Vecchiarelli, Dissection of the ATPase active site of McdA reveals the sequential steps essential for carboxysome distribution. *Mol. Biol. Cell* **32** (2021).

10. B. E. Funnell, ParB partition proteins: Complex formation and spreading at bacterial and plasmid centromeres. *Front. Mol. Biosci.* **3**, 44 (2016).

11. D. Schumacher, A. Harms, S. Bergeler, E. Frey, L. Søgaard-Andersen, PomX, a ParA/MinD ATPase activating protein, is a triple regulator of cell division in Myxococcus xanthus. *eLife* **10** (2021).

12. C. A. Kerfeld, C. Aussignargues, J. Zarzycki, F. Cai, M. Sutter, Bacterial microcompartments. *Nat. Rev. Microbiol.* **16**, 277–290 (2018).

13. M. Sutter, M. R. Melnicki, F. Schulz, T. Woyke, C. A. Kerfeld, A catalog of the diversity and ubiquity of bacterial microcompartments. *Nat. Commun.* **12**, 1–12 (2021).

14. J. S. MacCready, J. L. Basalla, A. G. Vecchiarelli, J. Echave, Origin and Evolution of Carboxysome Positioning Systems in Cyanobacteria. *Mol. Biol. Evol.* **37**, 1434–1451 (2020).

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

16. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, E. Stackebrandt, The Prokaryotes (Springer New York, 2006).

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

18. A. I. Flamholz, *et al.*, Functional reconstitution of a bacterial co2 concentrating mechanism in E. coli. *Elife* **9**, 1–57 (2020).

19. B. M. Long, *et al.*, Carboxysome encapsulation of the CO2-fixing enzyme Rubisco in tobacco chloroplasts. *Nat. Commun.* **9**, 1–14 (2018).

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

21. D. F. Savage, B. Afonso, A. H. Chen, P. A. Silver, Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science (80-.).* **327**, 1258–1261 (2010).

22. R. Rillema, Y. Hoang, J. S. MacCready, A. G. Vecchiarelli, Carboxysome mispositioning alters growth, morphology, and Rubisco level of the cyanobacterium Synechococcus elongatus PCC 7942. *MBio* **12** (2021).

23. J. S. MacCready, L. Tran, J. L. Basalla, P. Hakim, A. G. Vecchiarelli, The McdAB system positions α-carboxysomes in proteobacteria. *Mol. Microbiol.* **116**, 277–297 (2021).

24. Q. Zhu, *et al.*, Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nat. Commun.* **10**, 1–14 (2019).

25. H. Wang, *et al.*, Rubisco condensate formation by CcmM in β-carboxysome biogenesis. *Nature* **566**, 131–135 (2019).

26. L. M. Oltrogge, *et al.*, Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation. *Nat. Struct. Mol. Biol.* **27**, 281–287 (2020).

27. B. M. Long, B. Förster, S. B. Pulsford, G. D. Price, M. R. Badger, Rubisco proton production can drive the elevation of CO2 within condensates and carboxysomes. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021).

28. L. A. Metskas, *et al.*, Rubisco forms a lattice inside alpha-carboxysomes. *Nat. Commun.* **13**, 1–9 (2022).

Chapter 2

Biomolecular Condensate Formation as an Organizing Principle in Bacteria

2.1 Biomolecules can form liquid-like condensates

2.1.1 Definition of biomolecular condensates and condensate formation

Certain biomolecules, such as proteins and nucleic acids, have been shown to form what are now termed "condensates" *in vitro*. These condensates often appear as micron-sized droplets (Figure 2.1A) that can exhibit several liquid-like behaviors, such as fusion and relaxation into spheres [1, 2]. Initial biochemical characterizations of the liquid-like nature of these structures [3, 4, 5] ultimately led to many different names, including droplets, coacervates, and condensates. It is now agreed that each of these terms comes with their own set of mechanistic implications, and so it is important to be careful when choosing terminology. Here, we refer to the structures that form *in vitro* as condensates, with the following definition put forth by leaders in the field: condensates are entities not bound by a membrane that concentrate specific types of biomolecules which are often non-stoichiometric assemblies of multiple proteins and/or nucleic acids [1, 6, 7].
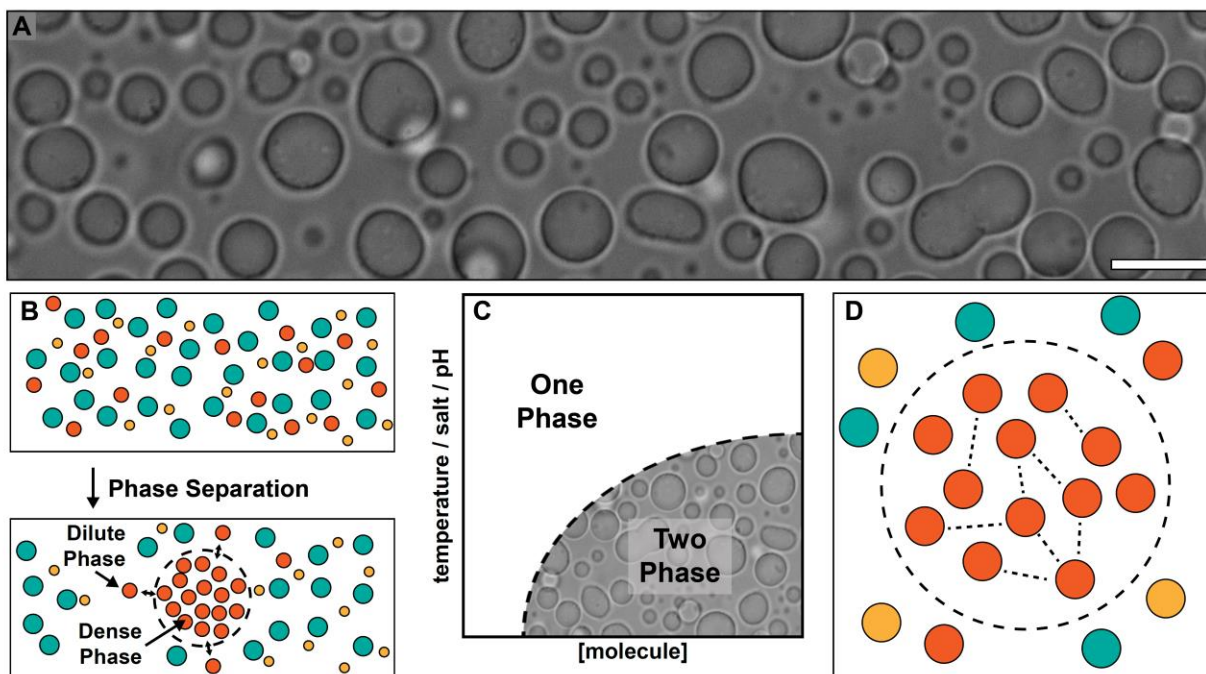
Figure 2.1. Biomolecular condensates and processes underlying their formation. (**A**) Representative image of biomolecular condensates using widefield microscopy. Condensates display liquid-like behaviors such as fusions and relaxation into spheres. Scale bar: 10 µm. (**B**) Schematic representing the process of phase separation. Each colored dot represents a different component in a mixture (*top*). After a specific component undergoes phase separation (*red dots*), this component adopts two coexisting phases, named dense and dilute (*bottom*). The dense phase, enclosed in the dotted line, is enriched in the molecule, but individual molecules can exchange between phases. (**C**) A schematic of a phase diagram. Solution conditions such as temperature, salt levels, pH, and concentration of the molecule of interest can determine if a molecule will undergo phase separation. Under a specific set of conditions, here within the dotted line, separation into two phases is promoted. (**D**) A schematic representation of percolation within a condensate. Here, molecules within the dense phase have undergone percolation to form a network of interactions represented by the dotted lines. This process can change the material properties of condensates over time.

Importantly, the term "condensates" is intentionally meant to not assign any specific mechanism to their formation. This came after many groups were quick to assign phase separation as the sole mechanism underlying condensate formation [8]. However, for reasons described in more detail below, it is now clear that phase separation alone is often not sufficient to describe the formation and physical nature of biomolecular condensates [7]. Therefore,

throughout this dissertation, we often use the term "condensate formation" instead of simply "phase separation" to reference the many mechanisms, including but not limited to phase separation, that can underlie the formation of condensates *in vitro* and *in vivo*.

2.1.2 Biomolecular condensates can form via several mechanisms

Phase separation is a physical process that has been studied for decades [9, 10]. The basic principle is that certain molecules that are homogeneous in a fluid can undergo a density transition to form two coexisting phases: a dense phase that is enriched in the molecule and a dilute phase that has low density [7, 11] (Figure 2.1B). In a given mixture, phase separation will occur under a specific set of conditions which include parameters such as temperature, pH, ionic strength, and concentration of the molecule [8] (Figure 2.1C). An important driving force for phase separation is therefore the interaction of the molecule with its surrounding solvent, where unfavorable interactions can drive the molecule to separate out into a dense, solvent-poor phase [7].

Unlike simple homopolymers, biopolymers like proteins and nucleic acids are complex heteropolymers that can have multiple inter- and intra-molecular interactions with varying strengths and specificities. Therefore, liquid-liquid phase separation (LLPS) driven by unfavorable solvent interactions is often not sufficient to model the formation of biomolecular condensates. Instead, it is more appropriate to think of biomolecular condensate formation as occurring via what is termed "complex coacervation", which involves weak, multivalent interactions between molecules as well as solvent-driven phase separation [12]. Additionally, once condensates are formed, biomolecules can continue to develop intermolecular interactions within condensates over time that can change the material properties of the condensates [13, 14].

17

This means that, although biomolecular condensates may be initially liquid-like in nature, they can often "mature" into more solid-like entities due to the development of internal interaction networks [13, 14] (Figure 2.1D). This process of network formation is referred to as percolation [7].

All of the above mechanisms can simultaneously influence the nature and formation of biomolecular condensates [7]. As described in the following section, it is important to have a general appreciation of these physical processes to determine what techniques are useful for investigating and perturbing biomolecular condensates, including techniques used in this dissertation. Additionally, the descriptions of the physical nature of condensates can help emphasize the advantages they provide cells, allowing for dynamic assemblages of specific molecular components.

2.1.3 An understanding of the processes of condensate formation can direct their investigation

A basic appreciation of the physical processes underlying biomolecular condensate formation has been key to directing the field as to which techniques and observations are useful for determining the roles of condensate formation in biology [8]. For instance, an understanding that these condensates form in part by phase separation reveals that the condensates seen via microscopy (Figure 2.1A) represent only the dense phase of two coexisting phases (Figure 2.1B). This means that individual molecules can fluctuate between the dense and dilute phases, but an equilibrium will be reached between the amount in the two phases. Therefore, centrifugation assays are often used to determine the ratios of molecules between the dense phase condensates and dilute, soluble phases to quantify the ratio of the two phases at a given condition [8]. Furthermore, as mentioned, some condensates can undergo percolation over time to mature from

18

being liquid-like to more solid-like. This is often monitored using fluorescence-recovery after photobleaching (FRAP) to determine how dynamic the molecules within the condensate are, and how these dynamics change over time [7, 8]. These strategies are then often used on mutant variants of proteins to quantify the degree to which mutations specifically affect condensate formation and maturation, and then investigate these effects in the cell. These techniques are used in the literature and in subsequent chapters of this dissertation to quantitatively investigate the biological implications of condensate formation.

2.2 Biomolecular condensates influence cellular biology as "membraneless organelles"

Over the past decade, numerous examples of how biomolecular condensates play roles in diverse cellular processes have been described. Notable examples of eukaryotic structures that behave as condensates include cytoplasmic bodies such as P granules [3, 4] and stress granules [13, 15] as well as nuclear structures such as nucleoli [3, 16] and transcription factor hubs [17, 18]. In addition to condensate-like structures, examples have been described in which the processes of condensate formation are themselves functionally important more so than the structure [19]. In these examples, the rapid and conditional changes in density of specific molecules can provide a signal to alter cellular behaviors at large [19, 20, 21, 22]. Therefore, both the products and the process of biomolecular condensate formation have been shown to play critical roles in cellular biology. Often, these condensate-based structures *in vivo* are referred to as "membraneless organelles" [23]. This is due to these condensate-based structures acting to confine specific cellular processes in time and space like traditional lipid-bound organelles, but lacking a membrane. There is now a growing list of evidence that in bacteria, which typically

lack traditional lipid-bound organelles altogether, membraneless organelles and the processes of condensate formation play important roles in spatiotemporal organization of the cell.

2.3 The roles of condensate formation in bacteria

Traditional views of bacterial cells as having little to no subcellular structure and organization are outdated. Some examples of organized structures involved in key bacterial processes include mRNA processing bodies, heterochromatin, biofilm formation, and BMCs. Recent evidence suggests condensate formation is a key factor underlying these and other bacterial processes [24]. Notably, several adaptor proteins to ParA/MinD-family ATPases have also been proposed to form condensates *in vitro* and *in vivo* [25, 26, 27]. Thus, condensate formation has widespread functional implications in bacterial cell biology. Below I describe my specific contributions in studies implicating condensate formation in regulating mobile genetic elements in bacteria and in the formation of bacterial biofilms. I then summarize how condensate formation influences adaptor protein function, and the implications this may have for McdB function in carboxysome organization in the cell.

2.3.1 Condensate formation in regulating mobile genetic elements in bacteria

One of the most widespread and potentially oldest biomolecules is polyphosphate, or polyP [28]. PolyP is a polymer of inorganic phosphate, and can function as a storage molecule for phosphate as well as play a wide variety of functional roles in all kingdoms of life [29]. Recently, in bacteria, polyP has been implicated in playing a key role in regulating mobile genetic elements (MGEs) [30]. MGEs are sections of bacterial chromosomes, such as prophage

20

DNA, that can mobilize upon stress and cause gross reorganization and damage to the chromosome [31]. A recent study showed that deletion of polyP kinase (PPK), which is responsible for synthesizing polyP, resulted in an increase in the amount of MGEs expressed in *Escherichia coli* cells [30]. A follow-up study showed that this is due to the interactions polyP has with the DNA binding protein Hfq [31]. It is now appreciated that bacteria contain several different nucleoid-associated proteins, including Hfq, that can occupy large segments of the chromosome to silence genes and MGEs, similar to eukaryotic heterochromatin [32]. These studies showed that loss of polyP through deletion of PPK altered the nucleoid occupancy pattern of Hfq, which resulted in an increased expression of MGEs [30, 31]. For this report, I reconstituted this system *in vitro* to show that purified Hfq can form condensates, which recruit and enrich both polyP and DNA [33]. Furthermore, using FRAP across several conditions, I found that the addition of polyP and DNA into the Hfq condensates altered the material properties of the condensates [33]. Together, our observations suggest that Hfq condensate formation may contribute to its interactions with polyP and DNA *in vivo* in a way that allows polyP to mediate Hfq-DNA binding and MGE expression.

2.3.2 Condensate formation in bacterial biofilms

One of the major components of bacterial biofilms are proteinaceous curli fibers [34]. These curli fibers are considered "functional amyloids", in that the major protein components of curli form amyloids similar to those associated with diseases like Parkinson's, but curli provide advantages through processes such as biofilm formation, cell-cell adhesion, and cell-host interaction [34, 35]. An understanding of the molecular processes underlying curli formation has

therefore helped our understanding of essential bacterial processes as well as the general principles of amyloid formation and structure.

The major protein component of curli fibers is curli-specific gene A (CsgA) [34]. CsgA is secreted into the extracellular matrix where amylogenesis is nucleated by the minor curli component CsgB [36]. On their own, curli fibers formed by CsgA and CsgB will not localize to the cell surface, but instead require the protein CsgF for localization [37]. How CsgF helps localize and nucleate curli fibers at cell surface has not been understood. However, a recent paper from the Chapman group has provided evidence suggesting this cell surface localization is influenced by CsgF condensate formation [38]. In this study, I showed that CsgF can form condensates *in vitro* that can concentrate the nucleator protein CsgB and stimulate its amylogenesis [38]. Furthermore, I showed these condensates form more readily at high salt, implicating an importance for hydrophobic residues [38]. Mutating key aromatic residues in CsgF decreased its ability to form condensates *in vitro*, and these same mutations decreased its ability to localize curli to the cell surface *in vivo* [38]. Together, these data suggest that CsgF condensates help concentrate and localize CsgB to the cell surface to initiate its amylogenesis and therefore nucleate CsgA fibers at the cell surface.

2.3.3 *The adaptor proteins associated with ParA/MinD positioning ATPases can form condensates with functional implications*

Recent evidence has implicated condensate formation to play important roles in mediating the interactions between some adaptor proteins with their cargos. For example, ParA ATPases are responsible for the spatial regulation of chromosomes and plasmids that are bound by the adaptor protein ParB [39, 40, 41]. The exact nature of the interaction between ParB and

DNA remains a vibrant area of research [42], but recent reports show that ParB-DNA complexes behave as dynamic, liquid-like condensates both *in vitro* [25] and *in vivo* [26], implementing condensate formation as an underlying assembly mechanism. Another example is the co-complex of adaptor proteins, PomX and PomY, for the ParA/MinD ATPase called PomZ. PomXYZ is responsible for spatially regulating division sites in some bacteria [43]. A recent study has shown that PomY forms condensates that nucleate GTP-dependent FtsZ polymerization, suggesting a novel mechanism for positioning cell division [27]. It is intriguing to speculate why such disparate adaptor proteins could use condensate formation as an underlying mechanism in localizing to such disparate cargos. The ability to undergo changes in density at specific locations and times within the cell may confer an advantage to these highly dynamic spatial regulation systems, and thus may also play a role for McdB function in carboxysome organization in the cell.

2.4 Motivations

In Chapter 1, I described the identification of a novel and widespread two-protein system composed of a ParA/MinD-family ATPase (McdA) and its adaptor protein (McdB), which are responsible for spatially regulating α- and ß-carboxysomes, and potentially other BMCs. How McdB interacts with carboxysomes, and whether the mode of interaction is the same for α- and ß-carboxysomes is of great interest for the fields of BMC biology and was one of the main motivations of the work described below.

Additionally, I determined that McdB proteins robustly form condensates *in vitro*. This was especially exciting given recent data suggesting that condensate formation played a functional role in carboxysome assembly and in the subcellular localization of different adaptor

proteins associated with other ParA/MinD ATPases. Therefore, I was highly motivated to investigate if McdB condensate formation contributed to its carboxysome positioning function and localization in the cell.

In Chapter 2, I discuss biomolecular condensates and how the field is generally interested in the biochemistry underlying condensate formation for diverse biomolecules. Understanding both the unique and shared biochemical features across different condensates can help deepen our understanding of how these condensates form and how we can control them to engineer cells [44, 45]. Thus, I was also motivated to dissect the biochemistry underlying McdB condensate formation, and determine if and how we could tune this process both *in vitro* and *in vivo*. From these motivations, I produced two first-author publications in which I determine the biochemistry of McdB condensate formation and how to control it, as well as make a significant first step in determining the mechanism by which McdB proteins interact with both α- and ß-carboxysomes.

2.5 References

1. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
2. A. S. Lyon, W. B. Peeples, M. K. Rosen, A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.* **22**, 215–235 (2020).
3. C. P. Brangwynne, T. J. Mitchison, A. A. Hyman, Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 4334–4339 (2011).
4. C. P. Brangwynne, *et al.*, Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science (80-.).* **324**, 1729–1732 (2009).
5. S. Elbaum-Garfinkle, *et al.*, The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7189–7194 (2015).
6. J. M. Choi, A. S. Holehouse, R. V. Pappu, Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu. Rev. Biophys.* **49**, 107–133 (2020).
7. T. Mittag, R. V. Pappu, A conceptual framework for understanding phase separation and addressing open questions and challenges. *Mol. Cell* **82**, 2201–2214 (2022).

8. S. Alberti, A. Gladfelter, T. Mittag, Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).
9. P. J. Flory, Thermodynamics of High Polymer Solutions. *J. Chem. Phys.* **10**, 51–61 (1942).
10. M. L. Huggins, Solutions of Long Chain Compounds. *J. Chem. Phys.* **9**, 440–440 (1941).
11. J. S. Rowlinson, Translation of J. D. van der Waals' "The thermodynamik theory of capillarity under the hypothesis of a continuous variation of density." *J. Stat. Phys.* **20**, 197–200 (1979).
12. C. W. Pak, *et al.*, Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol. Cell* **63**, 72–85 (2016).
13. A. Molliex, *et al.*, Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **163**, 123–133 (2015).
14. A. Patel, *et al.*, A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066–1077 (2015).
15. P. Yang, *et al.*, G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress Granules. *Cell* **181**, 325-345.e28 (2020).
16. D. L. J. Lafontaine, J. A. Riback, R. Bascetin, C. P. Brangwynne, The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol.* **22**, 165–182 (2020).
17. X. Liu, *et al.*, Mitotic Implantation of the Transcription Factor Prospero via Phase Separation Drives Terminal Neuronal Differentiation. *Dev. Cell* **52**, 277-293.e8 (2020).
18. A. Boija, *et al.*, Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855.e16 (2018).
19. H. Yoo, C. Triandafillou, D. A. Drummond, Cellular sensing by phase separation: Using the process, not just the products. *J. Biol. Chem.* **294**, 7151–7159 (2019).
20. J. A. Riback, *et al.*, Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* **168**, 1028-1040.e19 (2017).
21. T. M. Franzmann, *et al.*, Phase separation of a yeast prion protein promotes cellular fitness. *Science (80-.).* **359** (2018).
22. M. Du, Z. J. Chen, DNA-induced liquid phase condensation of cGAS activates innate immune signaling. *Science (80-.).* **361**, 704–709 (2018).
23. T. Hirose, K. Ninomiya, S. Nakagawa, T. Yamazaki, A guide to membraneless organelles and their various roles in gene regulation. *Nat. Rev. Mol. Cell Biol.* **24**, 288–304 (2022).
24. C. A. Azaldegui, A. G. Vecchiarelli, J. S. Biteen, The emergence of phase separation as an organizing principle in bacteria. *Biophys. J.* **120**, 1123–1138 (2021).
25. M. R. W. Brown, A. Kornberg, Inorganic polyphosphate in the origin and survival of species. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16085–16087 (2004).
26. N. N. Rao, M. R. Gómez-García, A. Kornberg, Inorganic Polyphosphate: Essential for Growth and Survival. *Annu. Rev. Biochem.* **78**, 605–647 (2009).
27. F. Beaufay, *et al.*, Polyphosphate functions in vivo as an iron chelator and fenton reaction inhibitor. *MBio* **11**, 1–14 (2020).
28. A. Du Toit, Phage induction in different contexts. *Nat. Rev. Microbiol.* **17**, 126–127 (2019).
29. P. L. Freddolino, H. M. Amemiya, T. J. Goss, S. Tavazoie, Dynamic landscape of protein occupancy across the Escherichia coli chromosome. *PLOS Biol.* **19**, e3001306 (2021).
30. F. Beaufay, *et al.*, Polyphosphate drives bacterial heterochromatin formation. *Sci. Adv.* **7** (2021).

31. M. M. Barnhart, M. R. Chapman, Curli Biogenesis and Function. *Annu. Rev. Microbiol.* **60**, 131–147 (2006).

32. A. Balistreri, E. Goetzler, M. Chapman, Functional Amyloids Are the Rule Rather Than the Exception in Cellular Biology. *Microorganisms* **8** (2020).

33. Z. Bian, S. Normark, Nucleator function of CsgB for the assembly of adhesive surface organelles in Escherichia coli. *EMBO J.* **16**, 5827–5836 (1997).

34. A. A. Nenninger, L. S. Robinson, S. J. Hultgren, Localized and efficient curli nucleation requires the chaperone-like amyloid assembly protein CsgF. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 900–905 (2009).

35. H. M. Swasthi, J. L. Basalla, C. E. Dudley, A. G. Vecchiarelli, M. R. Chapman, Cell surface-localized CsgF condensate is a gatekeeper in bacterial curli subunit secretion. *Nat. Commun.* **14**, 1–13 (2023).

36. J. Lutkenhaus, The ParA/MinD family puts things in their place. *Trends Microbiol.* **20**, 411–418 (2012).

37. A. G. Vecchiarelli, K. Mizuuchi, B. E. Funnell, Surfing biological surfaces: exploiting the nucleoid for partition and transport in bacteria. *Mol. Microbiol.* **86**, 513–523 (2012).

38. B. E. Funnell, ParB partition proteins: Complex formation and spreading at bacterial and plasmid centromeres. *Front. Mol. Biosci.* **3**, 44 (2016).

39. R. E. Debaugny, *et al.*, A conserved mechanism drives partition complex assembly on bacterial chromosomes and plasmids. *Mol. Syst. Biol.* **14**, 1–15 (2018).

40. L. Babl, *et al.*, CTP-controlled liquid–liquid phase separation of ParB. *J. Mol. Biol.* **434**, 167401 (2022).

41. B. Guilhas, *et al.*, ATP-Driven Separation of Liquid Phase Condensates in Bacteria. *Mol. Cell* **79**, 293-303.e4 (2020).

42. D. Schumacher, *et al.*, The PomXYZ Proteins Self-Organize on the Bacterial Nucleoid to Stimulate Cell Division. *Dev. Cell* **41**, 299-314.e13 (2017).

43. B. Ramm, *et al.*, Biomolecular condensate drives polymerization and bundling of the bacterial tubulin FtsZ to regulate cell division. *Nat. Commun. 2023 141* **14**, 1–24 (2023).

44. Y. Dai, L. You, A. Chilkoti, Engineering synthetic biomolecular condensates. *Nat. Rev. Bioeng.* **1**, 466–480 (2023).

45. C. D. Reinkemeier, E. A. Lemke, Synthetic biomolecular condensates to engineer eukaryotic cells. *Curr. Opin. Chem. Biol.* **64**, 174–181 (2021).

Chapter 3

Dissecting the Condensate Formation and Oligomerization Activities of the Carboxysome Positioning Protein McdB[1]

3.1 Abstract

Across bacteria, protein-based organelles called bacterial microcompartments (BMCs) encapsulate key enzymes to regulate their activities. The model BMC is the carboxysome that encapsulates enzymes for CO_2 fixation to increase efficiency and is found in many autotrophic bacteria, such as cyanobacteria. Despite their importance in the global carbon cycle, little is known about how carboxysomes are spatially regulated. We recently identified the two-factor system required for the maintenance of carboxysome distribution (McdAB). McdA drives the equal spacing of carboxysomes via interactions with McdB, which associates with carboxysomes. McdA is a ParA/MinD ATPase, a protein family well-studied in positioning diverse cellular structures in bacteria. However, the adaptor proteins like McdB that connect these ATPases to their cargos are extremely diverse. In fact, McdB represents a completely unstudied class of proteins. Despite the diversity, many adaptor proteins undergo phase

---

separation, but functional roles remain unclear. Here, we define the domain architecture of McdB from the model cyanobacterium *Synechococcus elongatus* PCC 7942, and dissect its mode of biomolecular condensate formation. We identify an N-terminal intrinsically disordered region (IDR) that modulates condensate solubility, a central coiled-coil dimerizing domain that drives condensate formation, and a C-terminal domain that trimerizes McdB dimers and provides increased valency for condensate formation. We then identify critical basic residues in the IDR, which we mutate to glutamines to solubilize condensates. Finally, we find that a condensate-defective mutant of McdB has altered association with carboxysomes and influences carboxysome enzyme content. The results have broad implications for understanding spatial organization of BMCs and the molecular grammar of protein condensates.

3.2 Introduction

Compartmentalization is a fundamental feature by which cells regulate metabolism. Although bacteria lack extensive lipid-membrane systems, recent reports have shown that proteinaceous bacterial microcompartments (BMCs) are a widespread strategy for compartmentalization in bacteria [1, 2]. Briefly, BMCs are nanoscale reaction centers where key enzymes are encapsulated within a selectively-permeable protein shell. The best studied BMC is the carboxysome, found within cyanobacteria and other autotrophic bacteria [1, 3]. Carboxysomes encapsulate the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) with its substrate CO_2 to significantly increase the efficiency of carbon fixation. Carboxysomes serve as a paradigm for understanding BMC homeostasis, including assembly, maintenance, permeability, and spatial regulation [1, 2, 3]. Furthermore, to engineer efficient carbon-fixing organisms, efforts to express functional carboxysomes in heterologous hosts are

ongoing [4, 5].

An important aspect of BMC homeostasis is spatial regulation [6]. We recently identified the two-protein system responsible for spatially regulating carboxysomes, which we named the maintenance of carboxysome distribution (McdAB) system [7, 8, 9]. McdA and McdB function to prevent carboxysome aggregation, thereby ensuring optimal function and equal inheritance upon cell division [7, 9]. Briefly, McdA is an ATPase that forms dynamic gradients on the nucleoid in response to an adaptor protein, McdB, which associates with carboxysomes [7, 9]. The interplay between McdA gradients on the nucleoid and McdB-bound carboxysomes result in the equal spacing of carboxysomes down the cell length of rod-shaped bacteria. This mode of spatial regulation by McdA is typical for the widespread and well-studied ParA/MinD family of positioning ATPases, of which McdA is a member. ParA/MinD ATPases spatially organize an array of genetic- and protein-based cargos in the cell, including plasmids, chromosomes, the divisome, flagella, and other mesoscale complexes [10, 11]. While ParA/MinD ATPases are highly similar in sequence and structure, the adaptor proteins that act as adaptors and link the ATPases to their respective cargo are highly diverse, largely due to adaptors providing cargo specificity. Indeed, McdB represents an entirely new class of adaptor proteins, and it is therefore unknown how McdB interacts with itself, McdA, and carboxysomes to confer specificity, or how these interactions are regulated. Bioinformatic analyses show that McdAB systems also exist for several other BMC types [9]. Therefore, an understanding of the biochemical properties of McdB, and how these properties influence its behavior *in vivo,* are important next steps to advancing our knowledge on the spatial regulation of carboxysomes and BMCs in general.

From our initial studies in the model cyanobacterium *Synechococcus elongatus* PCC 7942 (*Se*), we found that McdB self-associates *in vitro* to form both a stable hexamer [9] and

liquid-like condensates [8]. However, the domain architecture of McdB and the regions required for its oligomerization and condensate formation are unknown. Our understanding of how proteins form biomolecular condensates has rapidly developed over the past decade. Briefly, biomolecular condensates are the result of molecules having demixed out of solution to form a dense, solvent-poor phase that exists in equilibrium with the soluble phase [14, 15, 16]. This process occurs under a specific set of conditions where protein-protein interactions are more favorable than protein-solvent interactions [15]. A biochemical understanding of how proteins form condensates *in vitro* has led to a deeper understanding of how this process facilitates subcellular organization in both eukaryotic and prokaryotic cells [17, 18]. Furthermore, characterizing the underlying chemistries for diverse biomolecular condensates has led to the development of these condensates as synthetic tools to engineer cytoplasmic organization [19, 20, 21, 22]. Thus, a major focus of this report is to characterize the biochemistry of McdB, including its condensate formation, and link these properties to the spatial regulation of carboxysomes *in vivo*.

Here we define a domain architecture of *Se* McdB, identify the domains contributing to oligomerization and condensate formation, and discover a potential interplay between these two modes of self-association. We then create a series of point mutations that allow us to fine-tune the solubility of McdB condensates both *in vitro* and *in vivo* without affecting McdB structure or oligomerization. Finally, we use this mutation set to identify *in vivo* phenotypes that relate specifically to the ability of McdB to form condensates and associate with carboxysomes. The findings have implications for the use of carboxysomes in synthetic biology approaches, designing biomolecular condensates, and general BMC biology.

3.3 Results

3.3.1 Structural predictions generate a low confidence α-helical model for Se McdB

We first set out to determine the *Se* McdB crystal structure. However, McdB displayed robust phase separation across a range of buffer conditions, making crystal trials thus far unsuccessful (Figure 3.1). We next turned to I-TASSER (Iterative Threading ASSEmbly Refinement) [24, 25] to generate structural models, which predicted the McdB secondary structure to be predominantly α-helical but with a disordered N-terminus (Figure 3.2A). I-TASSER also generates full-length atomic models of the target sequence that are consistent with its secondary structure predictions via multiple sequence alignments using top matches from the protein databank (PDB). The top three models were once again almost entirely α-helical, with the top model also showing a disordered N-terminus (Figure 3.2B). But ultimately, the top 10 PDB matches identified by I-TASSER aligned poorly with McdB, with each alignment showing low sequence identity (< 20% on average) and low-quality scores (Z-scores < 1 on average) (Figure 3.2C). As a result, the top three final models generated by I-TASSER all have poor confidence scores (Figure 3.2B). These findings are not surprising in context with our previous bioinformatic analyses showing that cyanobacterial McdBs are highly dissimilar to other characterized proteins at the sequence level, potentially related to the high disorder content of McdBs [8, 26]. Together, these data provide low-confidence structural predictions for *Se* McdB. We therefore set out to validate these predictions with empirical approaches.

31
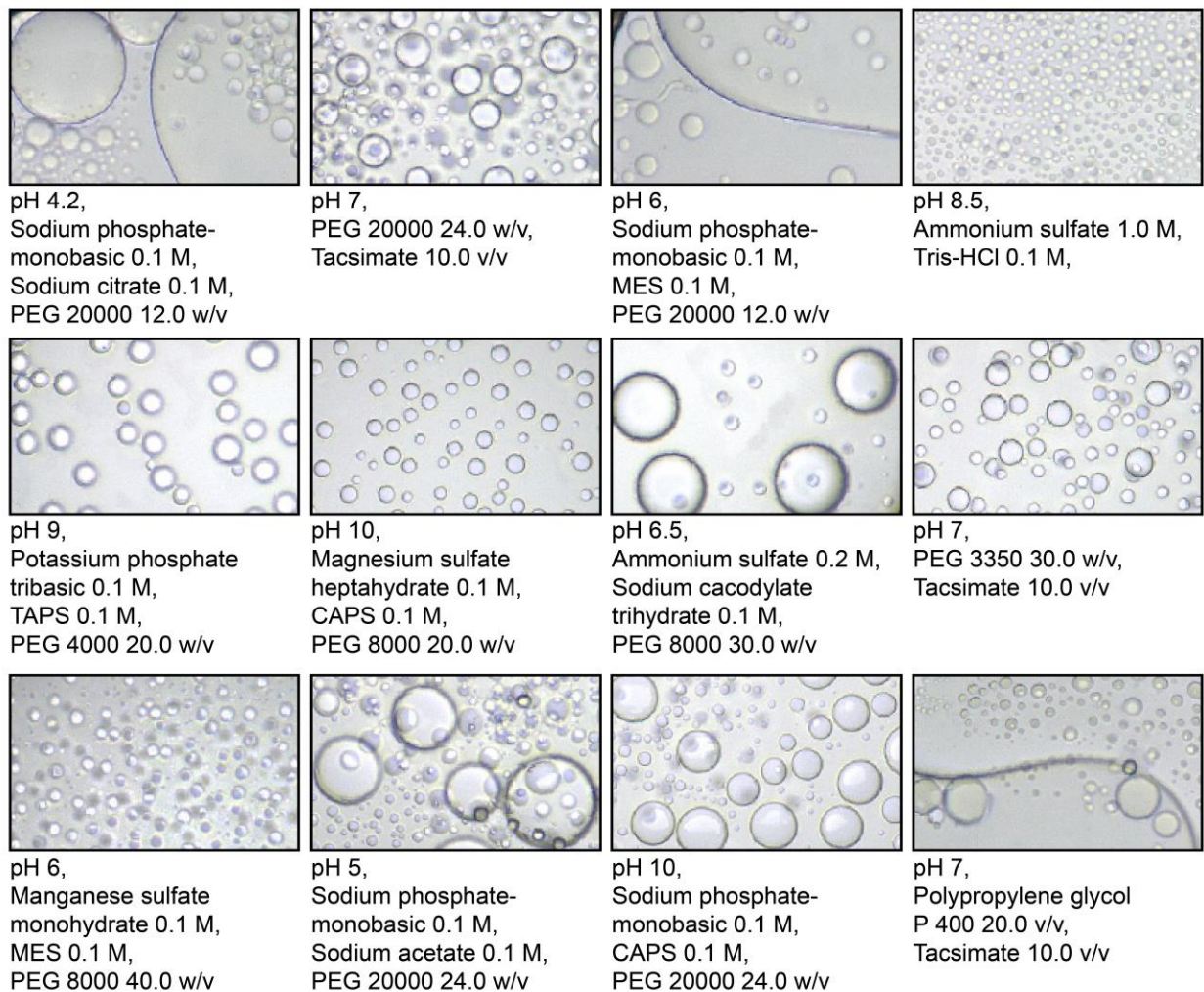
Figure 3.1. Phase separation of *Se* McdB across a range of buffer conditions during crystal screens. Images taken during buffer screens for crystallography. McdB at 10 mg/mL in (50 mM KCl; 10 mM CAPS pH 10.2) was diluted into the buffers indicated below each image. All images shown are at the same final concentration and magnification. Images were taken after 24-36 hours post dilution.

32

A

```
                    20                40                60
                    |                 |                 |
Seq.  MTDAFDRLKKRSRTPIAREGSLTTGPELSDRPLQLLPREFETFCDRYAVHAGDVIEAALDLVLLDPDLQQRLLQRL
Pred. CCCHHHHHHHCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCHHHHHHHHHH
Conf. 9607788886248862014776768565767166623889988898876101499997667610877999999999

          80                100               120               140
          |                 |                 |                 |
Seq.  RQGNGSDRVWLGTACPRSWQQQLQQQAQDQGLSEADLLQEAIAQRLDLVLGQTTLREEVTLLRQELDQLKRKLHGW
Pred. HHCCCCCCSSSCCCCCHHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCC
Conf. 7078976333447596668999999999876865888999999999889871137999999999999999874289
   H:Helix; S:Strand; C:Coil
```

B

1

2

3

C-score = -2.82   C-score = -2.45   C-score = -3.31

C-score range: [-5, 2] with 2 being most confident

N-terminus
C-terminus

C

Rank	PDB Hit	ID 1	ID 2	Z-score
1	4mh6A	0.14	0.20	0.74
2	4wheA	0.21	0.21	0.91
3	7mrwB	0.10	0.16	0.58
4	1aepA	0.18	0.22	0.61
5	5n9j	0.46	0.12	0.74
6	2eh3A	0.13	0.19	0.43
7	5cwjA	0.08	0.19	0.85
8	2wvf	0.11	0.18	1.71
9	6h9lA	0.15	0.18	0.69
10	2pffA	0.18	0.28	0.78
Average:		0.17	0.19	0.80

Figure 3.2: I-TASSER predictions for *Se* McdB. (**A**) McdB amino acid sequence and associated secondary structure predictions by I-TASSER. Each residue has a confidence score that ranges from 0 (least confident) to 9 (most confident). (**B**) Top three final models generated by I-TASSER. Each model is given a C-score that ranges from [-5, 2] with -5 being the least confident and 2 being the most. N-termini are colored blue and C-termini are colored red. (**C**) A table listing the top 10 PDB templates identified by I-TASSER, which were used for generating the models in panel B. ID1 is the percent sequence identity of the templates in the threading-aligned region with the query sequence. ID2 is the percent sequence identity of the whole template chains with the query sequence. Z-scores are a normalized score of the threading alignments. Alignments with a Z-score > 1 equates to good alignment. Overall, I-TASSER was unsuccessful in predicting a structure for McdB.

3.3.2 Defining a tripartite domain architecture for Se McdB

We used circular dichroism (CD) to characterize the secondary structure of *Se* McdB. The spectrum showed a characteristic α-helical signature that remained stable even after incubation at 80°C (Figure 3.3A), indicative of a stabilized coiled-coil [27]. This is consistent

33

with the helical predictions from I-TASSER, and with our previous bioinformatics data that predicted coiled-coil domains to be conserved across all cyanobacterial McdB homologs [8].

We next sought to empirically identify folded domains using limited proteolysis [28, 29, 30] (Figure 3.3B). Trypsin cuts at arginines and lysines, which are frequent throughout *Se* McdB - the largest fragment between any two basic residues is ~3 kDa (Figure 3.3C). Therefore, any stably folded regions that are protected from trypsin would be resolved via this approach. The digestion yielded three major bands of varying stabilities, which we labeled A, B, and C (Figure 3.3B). Band C was most stable, representing an ~11 kDa fragment that remained undigested for 12 hours. This strong protection against trypsin is consistent with our CD data, which showed high resilience to heat denaturation (see Figure 3.3A).

We next subjected bands A, B, and C to N-terminal sequencing to determine the location of these stably folded regions in McdB. All three bands had the same N-terminal sequence starting at E19 (Figure 3.3C). Therefore, the first 18 amino acids at the N-terminus were digested within the first minute to produce band A, and further digestion progressed slowly from the C-terminus to produce bands B and C (Figure 3.3B, C). By combining the N-terminal sequencing results (Figure 3.3C), the molecular weights of the three protected regions (Figure 3.3B), the locations of all arginines and lysines (Figure 3.3C), and the predicted disorder via PONDR VLXT [31] (Figure 3.3D), we developed a model for the domain architecture of *Se* McdB that was consistent with I-TASSER predictions (Figure 3.3D).
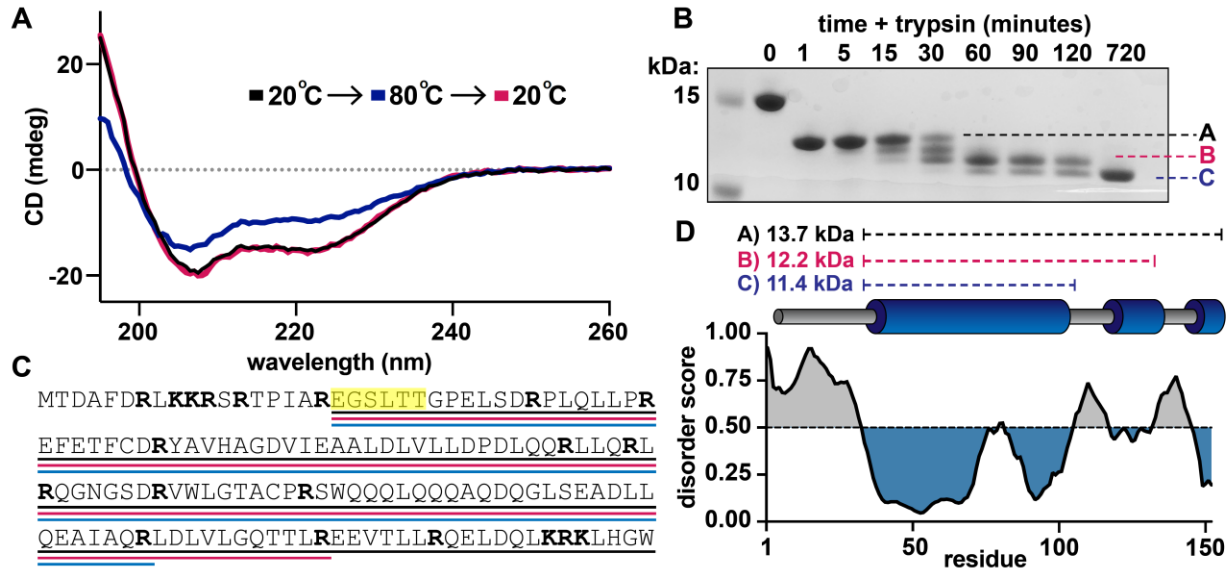
Figure 3.3: **Defining a domain architecture of *Se* McdB**. (**A**) Circular Dichroism (CD) spectra of McdB at 20°C (black), 80°C (blue), and then returned to 20°C (magenta). Spectra show α-helical structure resilient to heat denaturation. (**B**) SDS-PAGE analysis of trypsin-digested McdB sampled over digestion time. Bands labeled A, B, and C were isolated and N-terminally sequenced. (**C**) Amino acid sequence of McdB with basic residues (Lys-K and Arg-R) in bold. Regions corresponding to bands A, B, and C from panel B are underlined in black, magenta, and blue, respectively. Amino acids determined through N-terminal sequencing of bands A, B, and C are highlighted yellow. (**D**) Structural model of *Se* McdB. Regions corresponding to bands A, B, and C are indicated with predicted MWs (*top*). Predicted secondary structure of McdB (*middle*) aligned with a Predictor of Natural Disorder Regions (PONDR) plot using the VLXT algorithm (*bottom*) with disordered regions colored grey and predicted α-helical domains in blue.

3.3.3 Se McdB forms a trimer-of-dimers hexamer

From our structural model, we defined three major domains of *Se* McdB: (1) an intrinsically disordered region (IDR) at the N-terminus, (2) a highly stable central coiled-coil (CC), and (3) a C-terminal domain (CTD) with two short helical regions. We used this model to design a series of truncation mutants, including each of these domains alone as well as the CC domain with either the N-terminal IDR or the CTD (Figure 3.4A). CD spectra of these

truncations showed that the N-terminus was indeed disordered on its own, and both the CC domain and CTD maintained α-helical signatures (Figure 3.4B, Figure 3.5A).

Having previously shown that full-length *Se* McdB forms a hexamer in solution [9], we used these truncations to determine which domains contributed to oligomerization. We first ran size-exclusion chromatography (SEC) with each McdB truncation, and used the full-length protein as a reference for where the hexamer elutes. Although the molecular weight of each monomeric truncation is within ~5 kDa of one another (Figure 3.5B), we found that only the CC+CTD construct eluted at a volume similar to the full-length hexamer (Figure 3.5C). Furthermore, the CC domain alone or with the IDR, appeared to elute between the expected monomer and hexamer peaks, suggesting that the CC domain with or without the IDR forms an oligomeric species that is smaller than a hexamer.

To further resolve the oligomeric states of these McdB truncations, we performed size-exclusion chromatography coupled to multi-angled light scattering (SEC-MALS) on each of the constructs that appeared to oligomerize during SEC (Figure 3.4C). We found that the CC+CTD truncation was indeed hexameric, while the CC domain, with or without the IDR, formed a dimer (Figure 3.4D). These data suggest that the CC domain of McdB contains a dimerization interface, while the CTD subsequently allows for trimerization of dimers. Although we were unable to generate an atomic level structure of McdB nor determine the orientation of monomers within the hexamer, we have identified two key oligomerizing domains and conclude that full-length *Se* McdB forms a hexamer as a trimer-of-dimers.
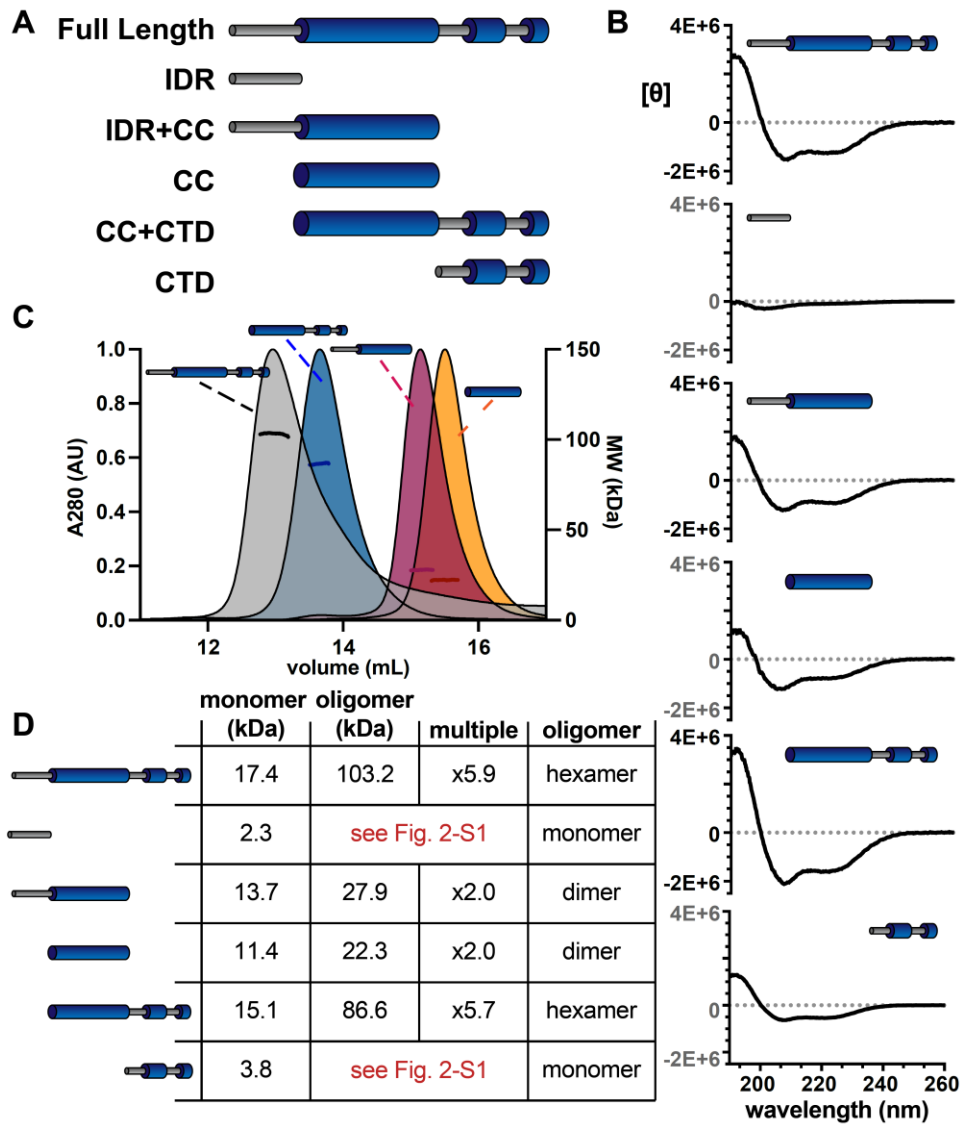
Figure 3.4: **The α-helical domains of McdB form a trimer-of-dimers hexamer**. (**A**) Illustration of McdB truncations generated based on the predicted domain structure. (**B**) CD spectra normalized by MW for the indicated McdB truncations. Spectra show α-helical content for all truncations, except for the disordered N-terminal fragment. (**C**) Size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) for full-length McdB and truncation mutants that showed oligomerization activity (**see Figure 3.5**). (**D**) Summary of the SEC-MALS data from (C).
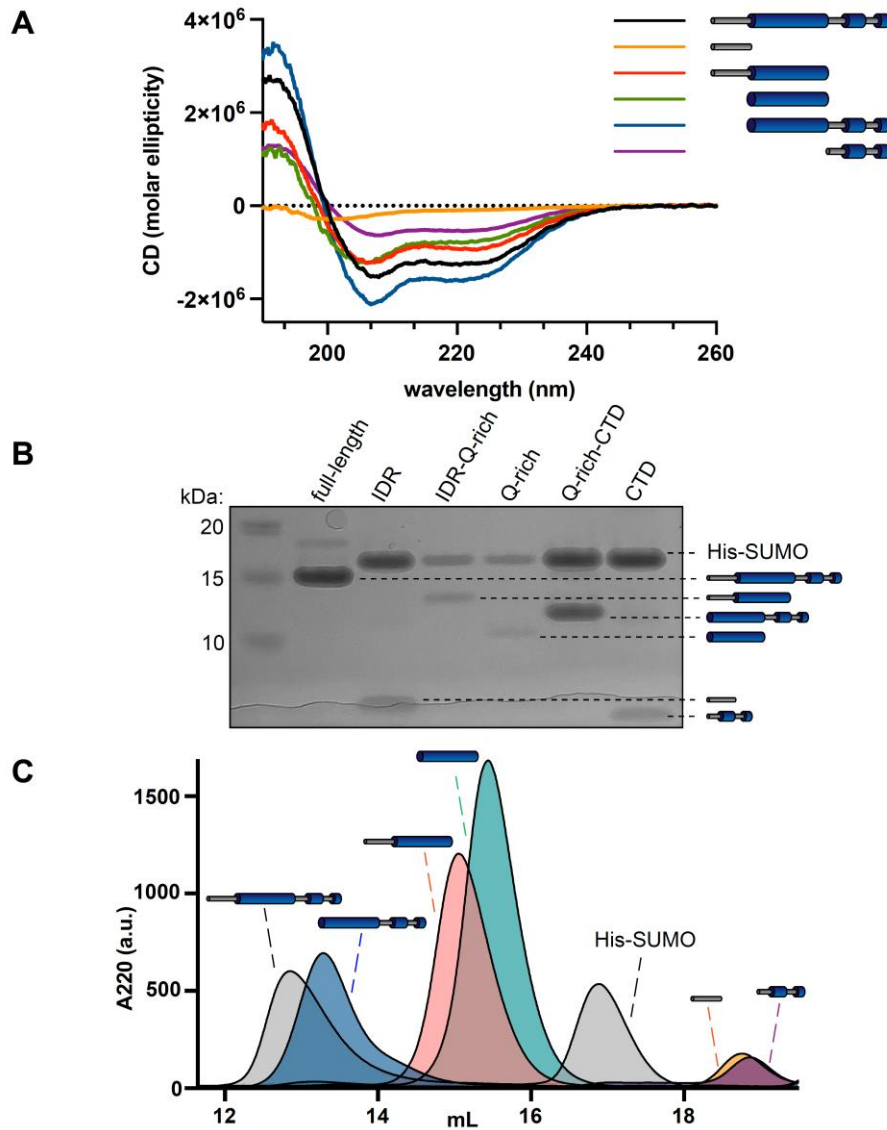
Figure 3.5: **McdB truncations have unique secondary structures and display different oligomeric states.** (**A**) CD spectra of full-length McdB and truncations. Curves from **Figure 2B** are overlayed onto a single graph. (**B**) SDS-PAGE analysis shows that full-length McdB and all truncation mutants run at a lower molecular weight compared to the His-SUMO solubility tag. (**C**) Size exclusion chromatography (SEC) showed that full-length McdB and the CC+CTD domain have similar elution profiles, suggesting similar oligomeric forms. The CC domain with and without IDR also elute similarly but after full-length and before His-SUMO, suggesting an intermediate oligomer. The IDR and CTD mutants eluted after the His-SUMO tag, showing they remain monomeric.

38

3.3.4 Se McdB forms condensates via pH-dependent phase separation coupled to percolation

Recent reports on the formation of protein condensates have begun to unveil an interplay between phase separation and network formation or "percolation" [32, 16]. These reports have shown that, instead of forming strictly through liquid-liquid phase separation, many proteins undergo phase separation coupled to percolation (PSCP) to form condensates [32, 16]. The definitions of phase separation, percolation, and PSCP are rigorous and nuanced. But broadly speaking, phase separation can often involve incompatibilities in solubility to drive transitions in density, while percolation concerns multivalent interactions to form dense networks [16]. For protein systems undergoing PSCP, different regions of the protein can facilitate solubility than the regions that facilitate multivalent networking [16]. We therefore set out to identify if full-length McdB showed signs of PSCP to guide our investigation on how the different domains of McdB affect condensate formation and stability.

Evidence for PSCP has recently been shown by studying the time-dependent viscoelastic nature of condensates [33, 34, 16] and the formation of networks at subsaturating concentrations [32, 16]. We therefore first used fluorescence recovery after photobleaching (FRAP) to assess the viscoelastic nature of McdB condensates [15]. Newly formed condensates recovered within minutes (Figure 3.6A), and readily fused and relaxed into spheres within seconds (Figure 3.6B), reminiscent of liquid-like fluid droplets. But after 18 hrs, recovery was significantly slower (Figure 3.6A) and these 'mature' droplets no longer fused. The time-dependent changes in material properties of McdB condensates is a signature of PSCP [16].

Next, we determined a saturation concentration (c_{sat}) for McdB condensate formation in our standard buffer conditions (100 mM KCl, 20 mM HEPES, pH 7.2). Condensates were observed at or above 2 µM, suggesting a c_{sat} between 1-2 µM (Figure 3.6C). We then used

dynamic light scattering (DLS) to investigate whether McdB formed network-like species at concentrations below c_{sat} as done previously for other proteins [32]. At 0.25 µM (~1/10 the observed c_{sat}) in our standard buffer (pH 7.2), McdB displayed a heterogeneous size distribution of defined species spanning 100 – 1000 nm in diameter (Figure 3.6D); significantly larger than a monodispersed hexamer [35]. Even after high-speed centrifugation, McdB in the supernatant remained mainly as mesoscale clusters on the order of 100 nm, suggesting the formation of McdB clusters at pH 7.2 (Figure 3.6D).

Our previous work has shown that McdB condensates are solubilized at higher pH values [8], therefore we set out to determine if McdB clusters remained in solution at higher pH. Interestingly, the clusters were pH-dependent. At pH 8.2, the largest clusters (> 100 nm) were lost, but smaller clusters (~ 60 nm) remained. At pH 10.2, a single homogeneous species remained at ~ 10 nm (Figure 3.6D), which is consistent with the hydrodynamic diameter of a monodispersed McdB hexamer [35]. Together, the data show that McdB can form mesoscale clusters at sub-saturation concentrations and that McdB condensates show time-dependent changes to viscoelasticity, suggesting McdB forms condensates via pH-dependent PSCP.
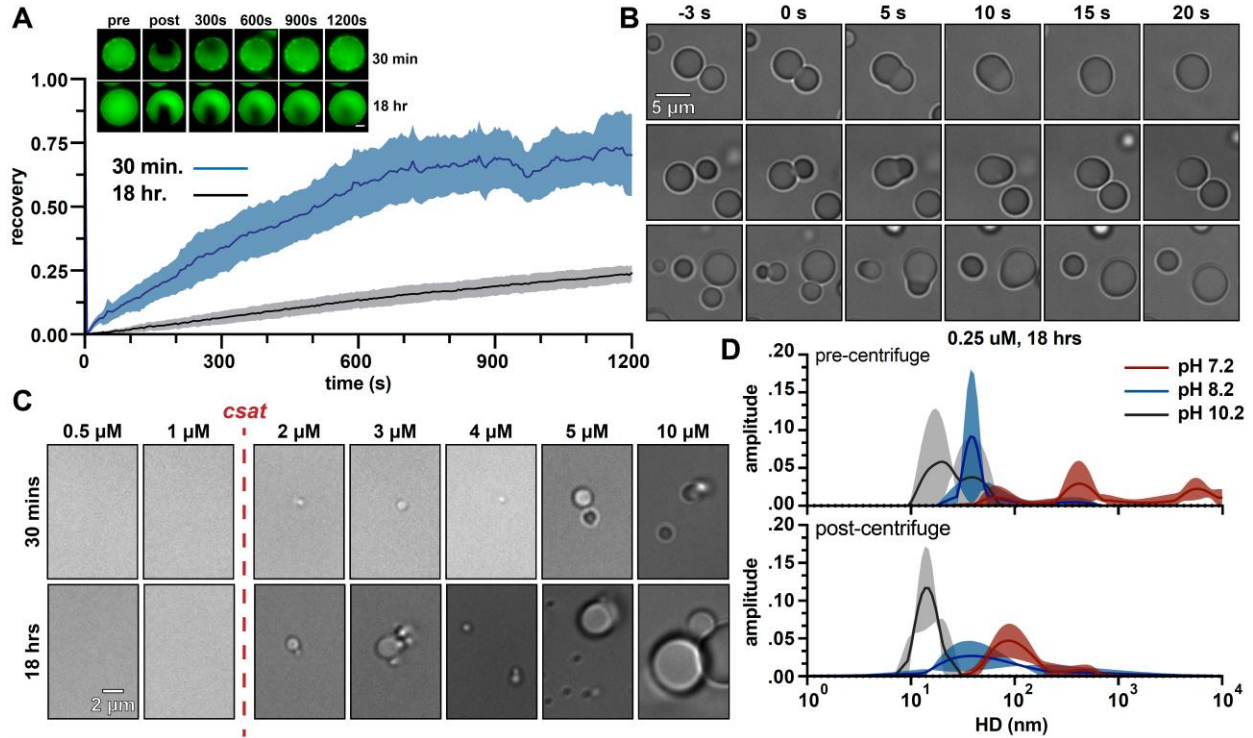
Figure 3.6: **McdB from *Se* forms liquid-like condensates via pH dependent phase separation coupled to percolation (PSCP).** (**A**) Fluorescence-Recovery After Photobleaching (FRAP) of McdB condensates at the indicated time points. Means and SD from n=8 condensates are shown. Representative fluorescence microscopy images for condensates incubated at 30 mins. or 18 hr. are shown (*inlet, scale bar = 2 μm*) (**B**) Representative DIC microscopy timeseries showing newly formed McdB condensates fusing and relaxing into spheres on the order of seconds. Scale bar applies to all images. (**C**) Representative DIC microscopy images at the indicated protein concentrations. McdB condensates were seen at and above concentrations of 2 μM, suggesting a saturation concentration (c_{sat}) between 1 - 2 μM. Scale bar applies to all images. (**D**) Dynamic Light Scattering (DLS) of McdB at a concentration ~1/10 the c_{sat} determined from (C) and at increasing pH values as indicated. Samples were analyzed both before (*top*) and after (*below*) a 5 min. spin at 20,000 x g. Larger "networks" are seen forming at lower pHs, even below the observed csat.

3.3.5 The CC domain of Se McdB is necessary and sufficient for condensate formation

We next used the truncations to determine how each domain of *Se* McdB affects

condensate formation and stability. Interestingly, no McdB truncations formed condensates

under the buffer conditions that sufficed for full-length McdB (Figure 3.7A). This finding suggests that no single domain of McdB is sufficient for full-length level condensate formation. Rather, all domains must influence McdB condensates to some extent.

When we added a crowding agent (10% PEG) to increase the local protein concentration, both the IDR and CTD alone were unable to form condensates, even at concentrations up to 4 mM (Figure 3.7B). However, all truncations containing the CC domain formed condensates. In fact, the CC domain alone was necessary and sufficient for forming condensates, albeit at much higher concentrations than full-length McdB. McdB condensates formed and fused similarly in the presence of other crowding agents, showing these activities were not PEG specific (Figure 3.8). It should be noted that we did not observe condensates forming by any truncation in the absence of a crowding agent, leaving open the possibility that these agents are functioning as multivalent co-assemblers and not simply crowding agents. Still, the data show that the CC domain is necessary for condensate formation and thus implicates this domain as the driver of McdB condensate formation.
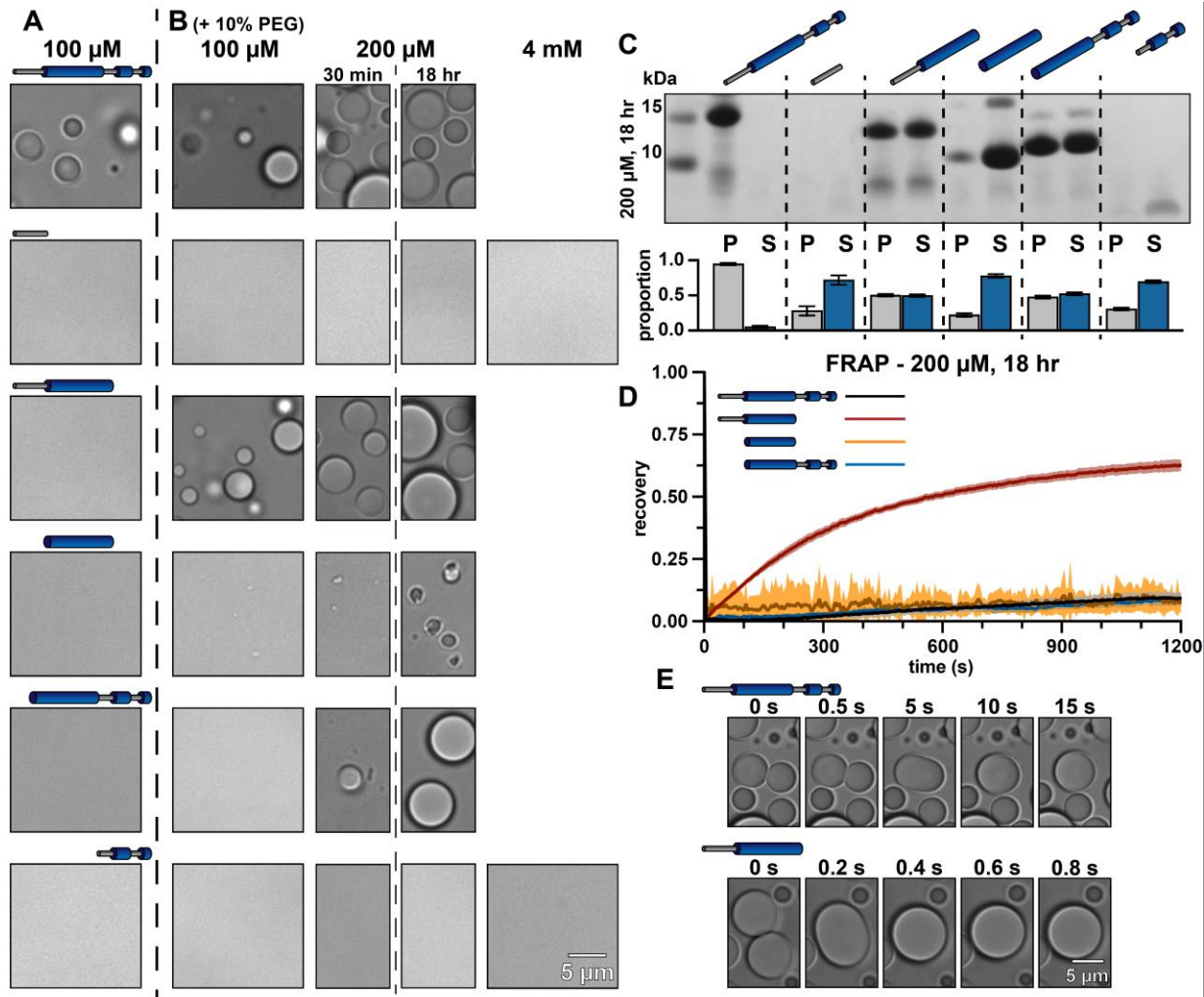
Figure 3.7: Truncations provide insight into the mechanisms of McdB condensate formation and stabilization. (**A**) Representative DIC microscopy images of full-length and truncation mutants of McdB at 100 µM in 150 mM KCl and 20 mM HEPES pH 7.2. (**B**) As in (A), but with increasing protein concentration as indicated and with the addition of 10% PEG-8000. Scale bar applies to all images. All domains are required for FL level condensate formation (**C**) Condensates at 200 µM after 18 hrs were pelleted (P) and run on an SDS-PAGE gel along with the associated supernatant (S) (*top*). P and S band intensities were then quantified (*bottom*). Mean and SD from 3 replicates are shown. (**D**) FRAP of condensates at the indicated condition reveal an increase in dynamics when the N-term IDR is present without the CTD. (**E**) Condensates containing the N-term IDR fuse orders of magnitude more quickly in the absence of the CTD, suggesting a stabilizing interaction between the two termini.

43
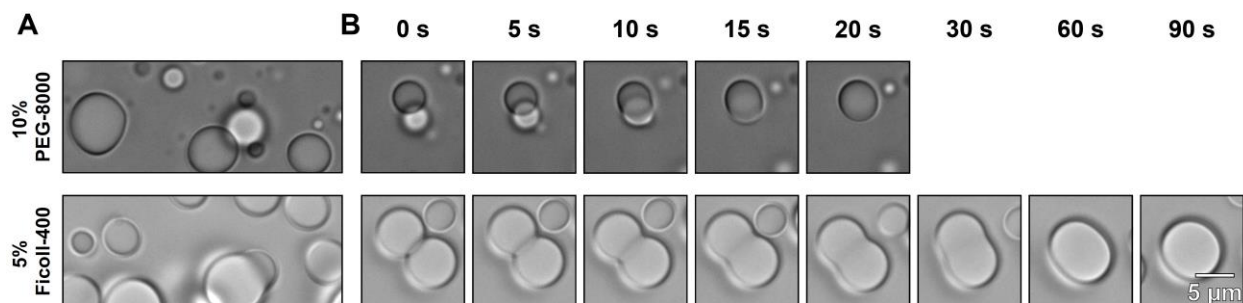
Figure 3.8: **McdB forms liquid-like condensates in both Ficoll and PEG.** (**A**) Representative DIC microscopy images for WT McdB at 100 µM in 100 mM KCl, 20 mM HEPES pH 7.2, and the addition of the indicated crowding agent. (**B**) Time course of images from (A) show that condensates fuse and relax into spheres on similar timescales, regardless of the crowding agent used.

3.3.6 The IDR and CTD domains are modulators of McdB condensate formation

Although the IDR and CTD were not required for the CC domain to form condensates, fusing either of these domains back onto the CC increased condensate formation and size (Figure 3.7B). By using centrifugation to quantify the amount of protein in the dense versus light phases [15], we found that the addition of either the IDR or CTD onto the CC domain comparably increased condensate formation (Figure 3.7C). However, by performing FRAP on 'mature' condensates (incubated for 18 hrs), we found that the IDR+CC condensates recovered much faster than all other constructs, including full-length McdB (Figure 3.7D). Moreover, newly formed IDR+CC condensates fused and relaxed into spheres an order of magnitude faster than newly formed full-length McdB condensates (Figure 3.7E).

Together, we draw the following conclusions: (*i*) The CC domain is necessary and sufficient for condensate formation, although to a lesser extent than full-length McdB; (*ii*) The IDR increases solvent interactions, thus affecting phase separation of the CC domain, but seemingly not percolation. This is supported by the fact that the IDR+CC construct has increased

44

condensate formation compared to the CC alone. However, IDR+CC condensates do not mature, lacking the change in viscoelasticity seen for full-length McdB (see Figure 3.6A); (*iii*) the CTD increases multivalent interactions to support condensate formation via PSCP. This occurs ostensibly by increasing oligomerization, which in turn increases valency [36, 37], and by the CTD itself providing network-forming contacts within condensates; and lastly (*iv*) the fact that full-length McdB, which contains the IDR and CTD, does not show the same fluid-like behavior as the IDR+CC (no CTD), suggests that the CTD may interact with the IDR within condensates formed by full-length McdB. Using this information, we next sought out residues that affect condensate formation, but not McdB structure or its ability to form a hexamer.

3.3.7 Net charge of the IDR modulates McdB condensate solubility

To determine which types of residues influence condensate formation, we first performed turbidity assays across a range of protein concentrations, salt concentrations, and pH as previously described [38, 15]. Over all McdB concentrations, the phase diagrams showed decreased turbidity at higher KCl concentrations (Figure 3.9), implicating electrostatic interactions. We also found that turbidity decreased at higher pH (Figure 3.9), suggesting that positively charged residues are important in the solubilization of condensates. We used centrifugation to quantify the amount of McdB in the dense- versus light-phases across KCl and pH titrations while keeping McdB concentration constant. Again, we found a clear increase in the soluble fraction and decreases in condensate size and number as both KCl (Figure 3.10A) or pH was increased (Figure 3.10B). The data reveal a critical role for positively charged residues in McdB condensate stability.
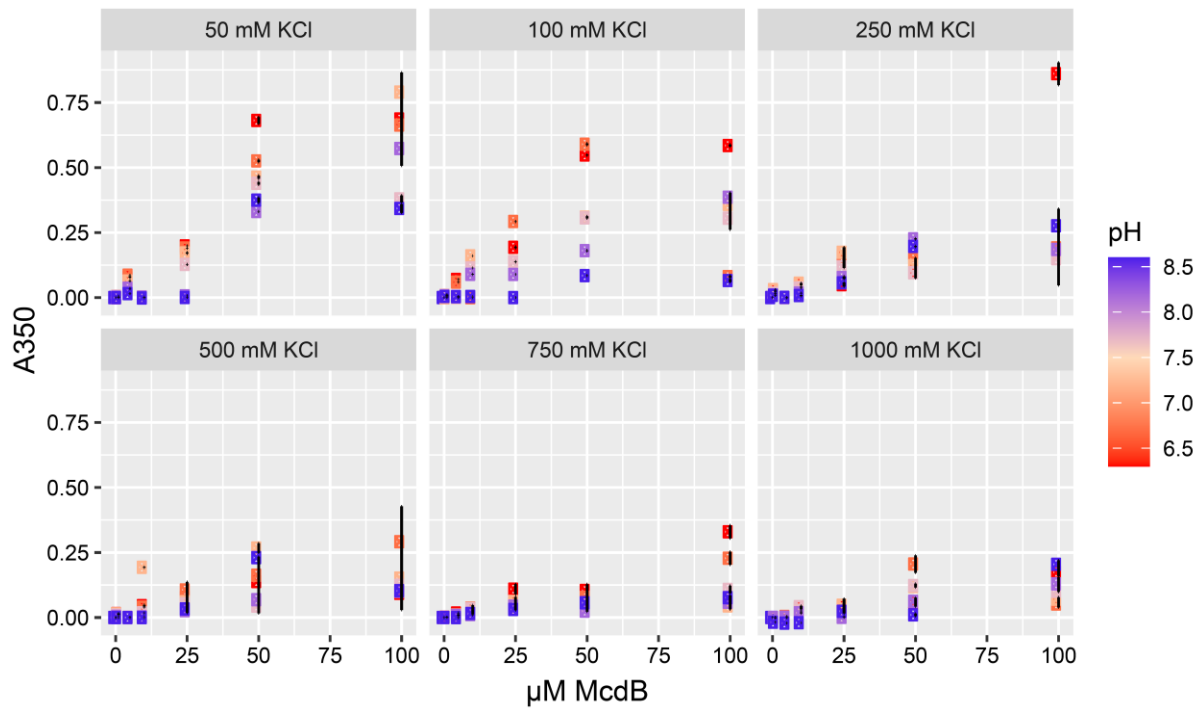
Figure 3.9: **Multi-dimensional phase diagrams for *Se* McdB.** Turbidity-based phase diagrams for McdB across varying protein concentration, KCl concentration, and pH. Data points represent the mean and error bars represent SD from at least three technical replicates. Turbidity monitored at A = 350 nm.

A

50 mM KCl | 100 | 250 | 500 | 750 | 1000

10 µm

15 kDa P S P S P S P S P S P S

proportion
1.0
0.5
0.0

50   100   250   500   750   1000

B

pH 6.3 | 6.6 | 7.2 | 7.7 | 8.2 | 8.6

10 µm

15 kDa P S P S P S P S P S P S

proportion
1.0
0.5
0.0

6.3   6.6   7.2   7.7   8.2   8.6

C

Domain

Name	Net Charge	Domain Sequence	Domain Model
IDR wt	+3	MTDAFDRLKKRSRTPIAREG	
IDR 3Q left	0	MTDAFDQLQQRSRTPIAREG	
IDR 3Q right	0	MTDAFDRLKKQSQTPIAQEG	
IDR all	-3	MTDAFDQLQQQSQTPIAQEG	
CTD wt	0	EEVTLLRQELDQLKRKLHW	
CTD 2Q left	+2	QQVTLLRQELDQLKRKLHW	
CTD 2Q right	+2	EEVTLLRQQLQQLKRKLHW	
CTD all	+4	QQVTLLRQQLQQLKRKLHW	

D

IDR all   WT   CTD all

A280 (AU)
1.0
0.8
0.6
0.4
0.2
0.0

MW (kDa)
150
100
50
0

12   14   16
volume (mL)

	monomer (kDa)	oligomer (kDa)	multiple	oligomer
WT	17.4	96.4	x5.5	hexamer
IDR all	17.3	96.1	x5.6	hexamer
CTD all	17.4	31.8	x1.8	dimer

E

10 µm

kDa: P S P S P S P S
20
15

proportion
1.0
0.5
0.0

47

Figure 3.10 (*previous page*): McdB condensates can be solubilized by mutating basic residues in the N-terminal IDR without affecting McdB structure. (**A**) Representative DIC microscopy images of 50 µM McdB in 20 mM HEPES pH 7.2 and increasing KCl concentration (*top*). Scale bar applies to all images. McdB condensates were pelleted (P) and run on an SDS-PAGE gel along with the associated supernatant (S) (*middle*). P and S band intensities were then quantified (*bottom*). Mean and SD from 3 replicates are shown (**B**) As in (A), except salt was held constant at 100 mM KCl and the pH was increased as indicated. (A) and (B) implicate stabilizing basic residues **(see Figure 3.11)** (**C**) Table showing the net charge and amino acid sequence of wild-type McdB compared to the glutamine (Q) - substitution mutants in both the N-term IDR and CTD. Acidic and basic residues in the IDR are colored red and blue, respectively. Q-substitutions are bolded. Graphical models of the McdB variants are also provided. (**D**) SEC-MALS of WT McdB compared to the full Q-substitution mutants from both the N- and C-termini. (*Below*) Table summarizing the SEC-MALS data, showing that mutations to the IDR does not affect oligomerization, while mutations to the CTD destabilize the trimer-of-dimers hexamer **(see Figure 3.12)**. (**E**) Representative DIC microscopy images for WT and IDR Q-substitution mutants of McdB at 100 µM in 150 mM KCl and 20 mM HEPES pH 7.2 (*top*). Scale bar applies to all images. McdB condensates were pelleted (P) and run on an SDS-PAGE gel along with the associated supernatant (S) (*middle*). P and S band intensities were then quantified (*bottom*). Mean and SD of 3 replicates are shown.

Together with our previous data showing that both the IDR and CTD modulate condensation, we focused on the basic residues within these two domains. By making a series of alanine substitutions (Figure 3.11A), we found that removing positive charge in the IDR, but not the CTD, caused a loss of condensates (Figure 3.11B). However, we also found that substituting charged residues for a more hydrophobic residue like alanine caused protein aggregation (Figure 3.11B) as found for other proteins [39]. Therefore, going forward, we transitioned to substituting these charged residues with polar glutamines (Figure 3.10C), to specifically affect charge and not hydrophilicity.

Data from this report and our previous study [8] suggest a potential electrostatic interaction between the positively-charged IDR and negatively-charged residues of the CTD. We therefore created another series of substitutions where we changed either positive charge in the IDR or negative charge in the CTD to glutamines. Before assessing condensate formation, we

first performed SEC-MALS to verify these mutations had no major impact on McdB structure.

Substituting all six basic residues to glutamines in the IDR had no effect on McdB

hexamerization (Figure 3.10D). On the other hand, only two substitutions in the CTD were

enough to partially destabilize the hexamer (Figure 3.12), and four substitutions produced mainly

McdB dimers (Figure 3.10D). As a result, we were unable to parse out the different roles of the

CTD in McdB oligomerization versus potential interactions with the IDR involved in condensate

formation. Importantly, however, we determined that removing only three positively charged

residues in the IDR solubilized McdB condensates (Figure 3.10E) without affecting protein

structure (Figure 3.13) or hexamerization (Figure 3.10D).
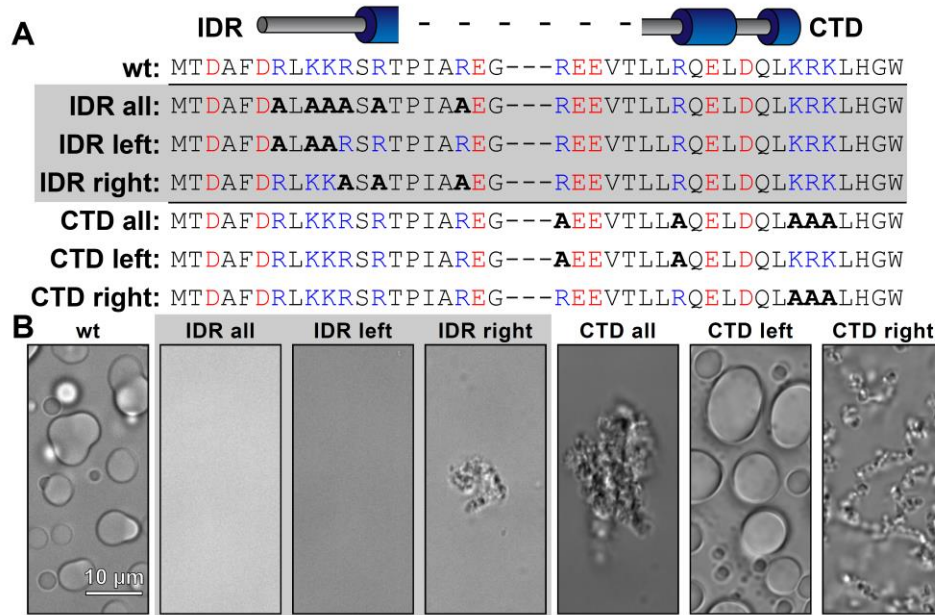
Figure 3.11: **Alanine-scanning of basic residues in the N- and C-termini of McdB.** (**A**) Table showing the sequence of WT McdB compared to the terminal A-substitution mutants. Acidic- and basic-residues are colored red and blue, respectively. A-substitutions are bolded. (**B**) Representative DIC microscopy images of all constructs listed in (A). Scale bar applies to all images.

49

A

Name	Domain Net Charge	Domain Sequence	Domain Model
CTD wt	0	EEVTLLRQELDQLKRKLHW	
CTD 2Q left	+2	QQVTLLRQELDQLKRKLHW	
CTD 2Q right	+2	EEVTLLRQQLQQLKRKLHW	
CTD all	+4	QQVTLLRQQLQQLKRKLHW	

B

	monomer (kDa)	oligomer (kDa)	multiple	oligomer
WT	17.4	96.4	x5.5	hexamer
CTD 2Q left	17.4	77.8	x4.5	mixed
CTD all	17.4	31.8	x1.8	dimer
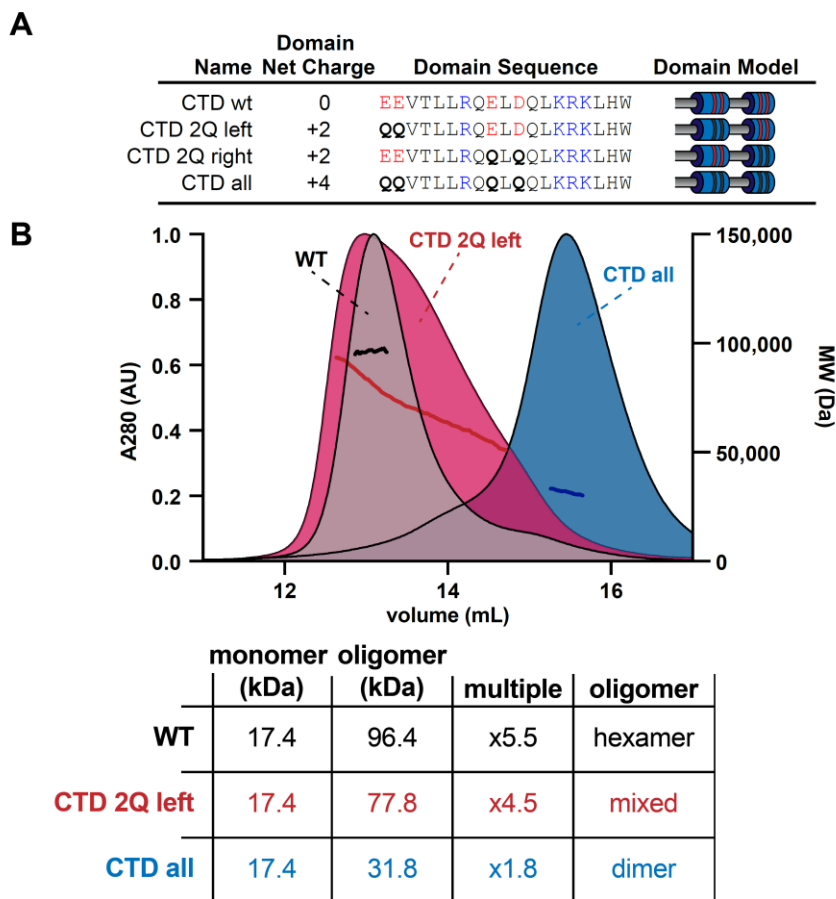
Figure 3.12: Mutations to the CTD destabilize the trimer-of-dimers hexamer. (**A**) Table showing the net charge and amino acid sequence of wild-type McdB compared to the Q substitution mutants in the CTD. Acidic and basic residues are colored red and blue, respectively. Q-substitutions are bolded. Graphical models of the McdB variants are also provided. (**B**) SEC-MALS graphs of the indicated variants (*top*) with a table summarizing results (*below*). Note the CTD 2Q left MALS data spans the MW from hexamer range to dimer range. The CTD 2Q right variant formed insoluble aggregates and is not shown.

A

Name	IDR Net Charge	IDR Sequence	Model
wt	+3	MTDAFDRLKKRSRTPIAREG	
2Q left	+1	MTDAFD**Q**L**Q**KRSRTPIAREG	
2Q middle	+1	MTDAFDRLK**QQ**SRTPIAREG	
2Q right	+1	MTDAFDRLKKRS**Q**TPIA**Q**EG	
3Q left	0	MTDAFD**Q**L**QQ**RSRTPIAREG	
3Q right	0	MTDAFDRLKK**Q**S**Q**TPIA**Q**EG	
all	-3	MTDAFD**Q**L**QQQ**S**Q**TPIA**Q**EG	
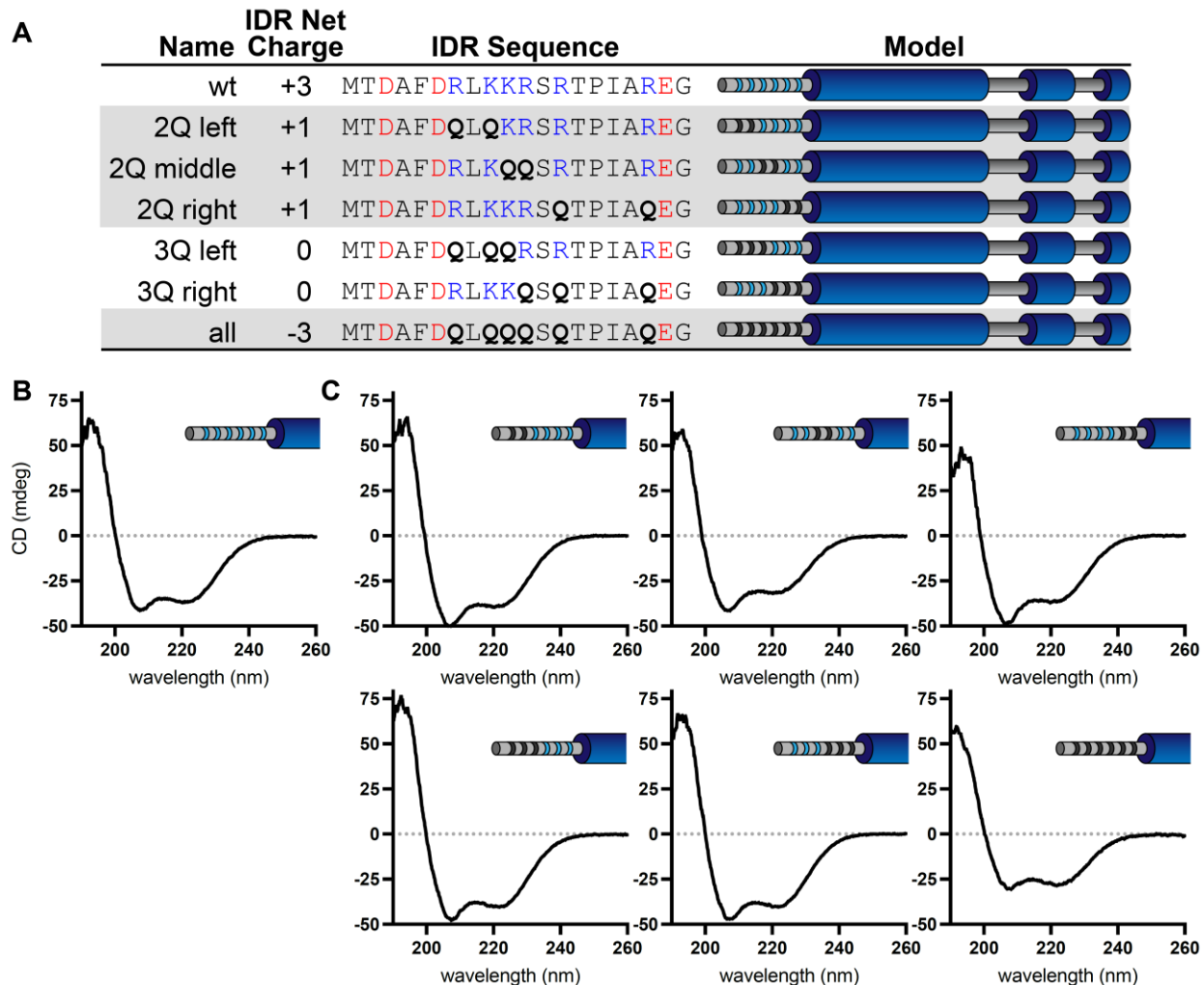
Figure 3.13: CD spectra of wild-type McdB and N-terminal glutamine substitution mutants. (**A**) Table showing the net charge and N-terminal IDR sequence of wild-type McdB compared to the glutamine-substitution mutants. Acidic- and basic-residues in the IDR are colored red and blue, respectively. Glutamine-substitutions are bolded. Graphical models of the McdB variants are also provided where blue stripes represent the six basic residues in the IDR. Black stripes represent the location of the glutamine substitutions. CD spectra of both (**B**) wild-type McdB and (**C**) mutants with the indicated glutamine substitutions in the N-terminal IDR of McdB.

51

3.3.8 McdB condensate formation is tunable through changes in IDR net charge

Substituting only three basic residues in the IDR (net charge -3) completely solubilized McdB condensates under our standard conditions (Figure 3.10E). We set out to determine the effect of fewer mutations in the IDR on condensate solubility. Pairs of basic residues in the IDR were substituted with glutamines, leaving an IDR net charge of +1 (Figure 3.14A). All +1 mutants still formed condensates, albeit smaller and fewer than that of wildtype McdB (Figure 3.14B).

The data suggest McdB condensate formation is tunable though changes to the net charge of the IDR. If correct, the triplet substitution mutants (IDR net charge 0) may still be capable of forming condensates at higher protein concentrations. Indeed, when we added a crowding agent, the net-charge 0 mutants formed condensates (Figure 3.14C). Moreover, a gradual increase in the proportion of McdB in the soluble phase was revealed as we incrementally removed positive charge from the IDR. Removing all six positive residues (net charge +3) still completely solubilized McdB even in the presence of a crowder (Figure 3.14C). Importantly, McdB mutants with the same IDR net charge, but with different residues substituted, showed similar changes to condensate solubility (Figure 3.14C). Substitution position showed slight differences in condensate size, but the overall effect on solubility was the same within each charge grouping. Together, the data show that it is the net charge of the IDR, and not a specific basic residue, that is critical for mediating condensate solubility.
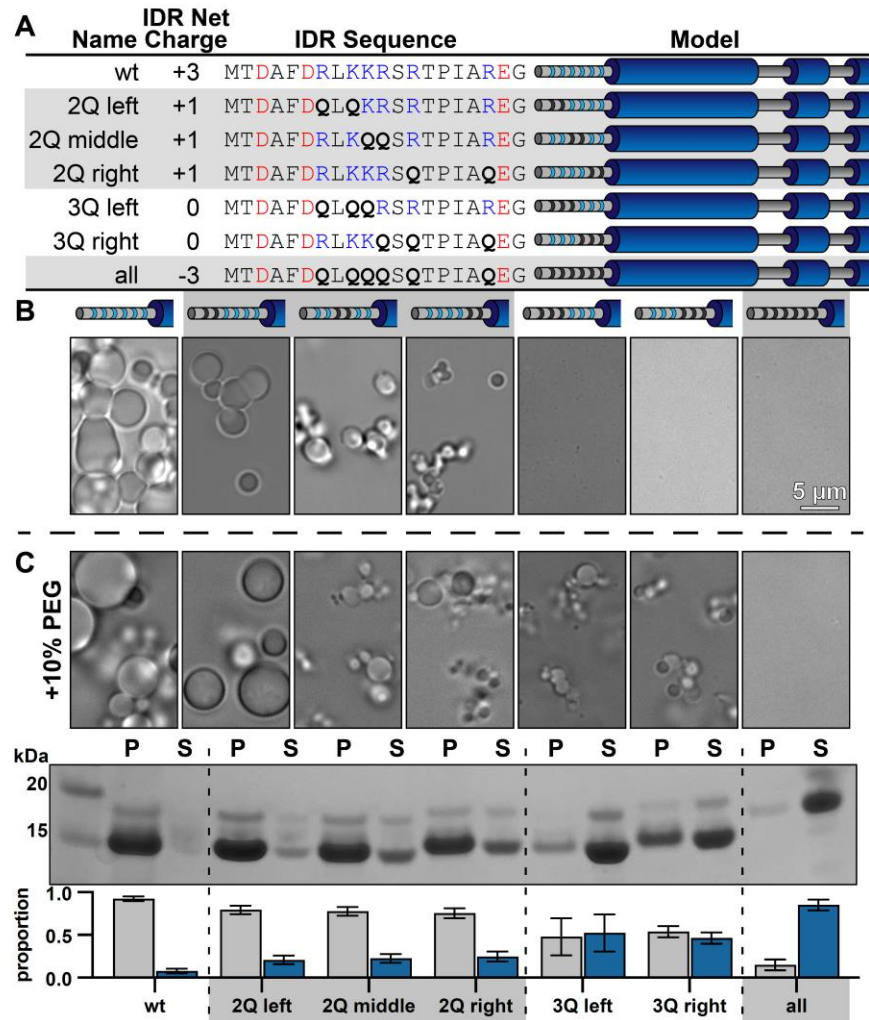
Figure 3.14: Net charge of the IDR can be used to tune the solubility of McdB condensates. (**A**) Table showing the net charge and N-terminal IDR sequence of wild-type McdB compared to the glutamine (Q)-substitution mutants. Acidic and basic residues in the IDR are colored red and blue, respectively. Q-substitutions are bolded. Graphical models of the McdB variants are also provided where blue stripes represent the six basic residues in the IDR. Black stripes represent the location of the Q-substitutions. (**B**) Representative DIC microscopy images for wild-type and the Q-substitution mutants of McdB at 100 µM in 150 mM KCl and 20 mM HEPES pH 7.2. Scale bar applies to all images. (**C**) As in (**B**), but with the addition of 10% PEG-8000 (*top*). McdB condensates were pelleted (P) and run on an SDS-PAGE gel along with the associated supernatant (S) (*middle*). P and S band intensities were then quantified (*bottom*). Mean and SD from 3 replicates are shown.

3.3.9 Net charge of the IDR affects McdB condensation in E. coli

To determine if the IDR can be used to tune McdB solubility in cells, we induced expression of fluorescent fusions of mCherry with both wildtype McdB (McdB[wt]) and the full glutamine-substitution mutant, with an IDR net charge of -3 ("McdB[-3]") in *E. coli* MG1655. As protein concentration increased, McdB[wt] formed polar foci that coexisted with a dilute cytoplasmic phase (Figure 3.15A), similar to the dense and dilute phases of McdB *in vitro*. After 3 hours of expression, nearly 70% of cells with McdB[wt] adopted this two-state regime (Figure 3.15B). The foci were indeed driven by McdB, as mCherry alone remained diffuse (Figure 3.15A-B). McdB[-3], on the other hand, was considerably more soluble than wild-type, where even after 3 hours of expression <10% of cells contained foci (Figure 3.15A-B). The change in solubility was not due to differences in protein levels or due to cleavage of the fluorescent tag (Figure 3.15C), but instead represents an increased solubility due to the IDR substitutions. Together the data show that adjustments to the net charge of the IDR can also affect McdB condensate solubility *in vivo*.
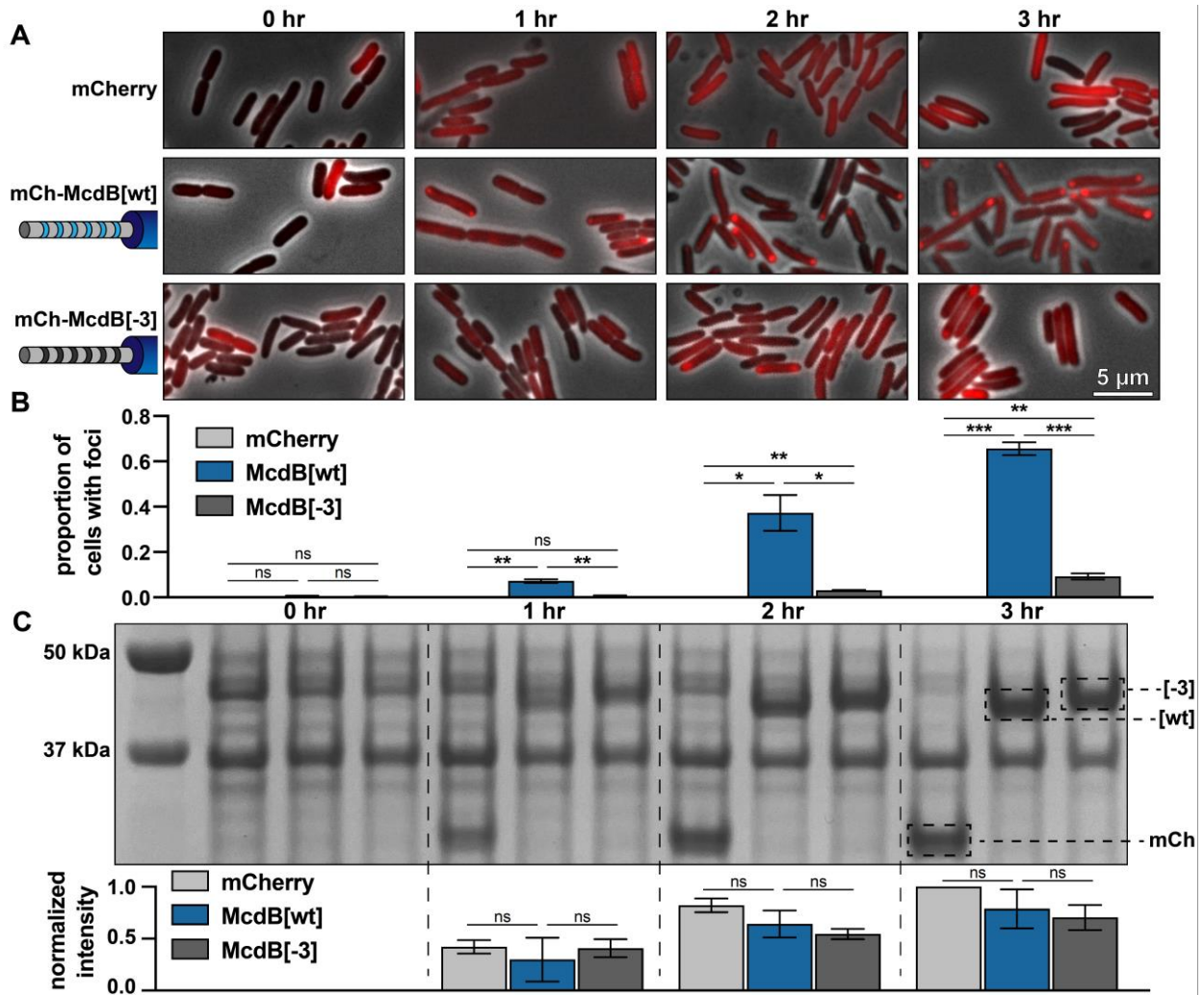
Figure 3.15: Net charge of the IDR affects McdB solubility in *E. coli*. (**A**) Representative fluorescence microscopy images monitoring the expression of the indicated constructs over time. Scale bar applies to all images. (**B**) Quantification of the proportion of cells containing foci from the images represented in (A). All quantifications were done on > 300 cells and from n = 3 technical replicates. Reported values represent means with SD. *p < 0.05 **p < 0.01 ***p < 0.001 by Welch's *t* test. (**C**) SDS-PAGE of cell lysates from the time course represented in (A). All samples were standardized to the same OD600 prior to loading. The expected MWs of the three constructs indicated are: mCherry 26.8 kDa; mCh-McdB[wt] 44.6 kDa; mCh-McdB[-3] 44.5 kDa. Normalized intensities from the indicated bands were quantified from 3 biological replicates (*below*). Reported values represent means with SD. Data were analyzed via Welch's *t* test.

3.3.10 McdB[-3] causes mispositioned carboxysomes, likely due to an inability to interact with

McdA

Having identified a mutant that solubilizes McdB condensates, without affecting structure

or hexamerization, we set out to determine its influence on carboxysome positioning in *Se*.

mNeonGreen (mNG) was N-terminally fused to either McdB[wt] or McdB[-3] and expressed at

its native locus. The small subunit of Rubisco (RbcS) was C-terminally fused to mTurquoise

(mTQ) to image carboxysomes. As shown previously [7], mNG-McdB[wt] supported well-

distributed carboxysomes along the cell length (Figure 3.16A, Figure 3.17). The mNG-McdB[-3]

strain, on the other hand, displayed carboxysome aggregates. However, it is important to note

that McdA is a ParA/MinD family ATPase, which typically interact with their adaptor proteins

via basic resides in the N-terminus of the adaptor protein, analogous to McdB [40, 41, 42, 43, 44,

45]. Therefore, it is highly likely that one or more of the basic residues removed from McdB[-3]

not only modulate condensate formation, but also mediate McdA interaction. A loss in McdA

interaction would explain the carboxysome aggregation phenotype, as we have shown previously

[7,12]. To investigate this possibility, we knocked out McdA in the McdB[-3] mutant and found

no significant differences in carboxysome mispositioning compared to the *ΔmcdA* strain alone

(Figure 3.17). Together, the data suggest that carboxysome aggregation in the McdB[-3] strain is

due to this mutant's inability to interact with McdA, and is not necessarily due to the effects

these mutations have on McdB condensate formation.
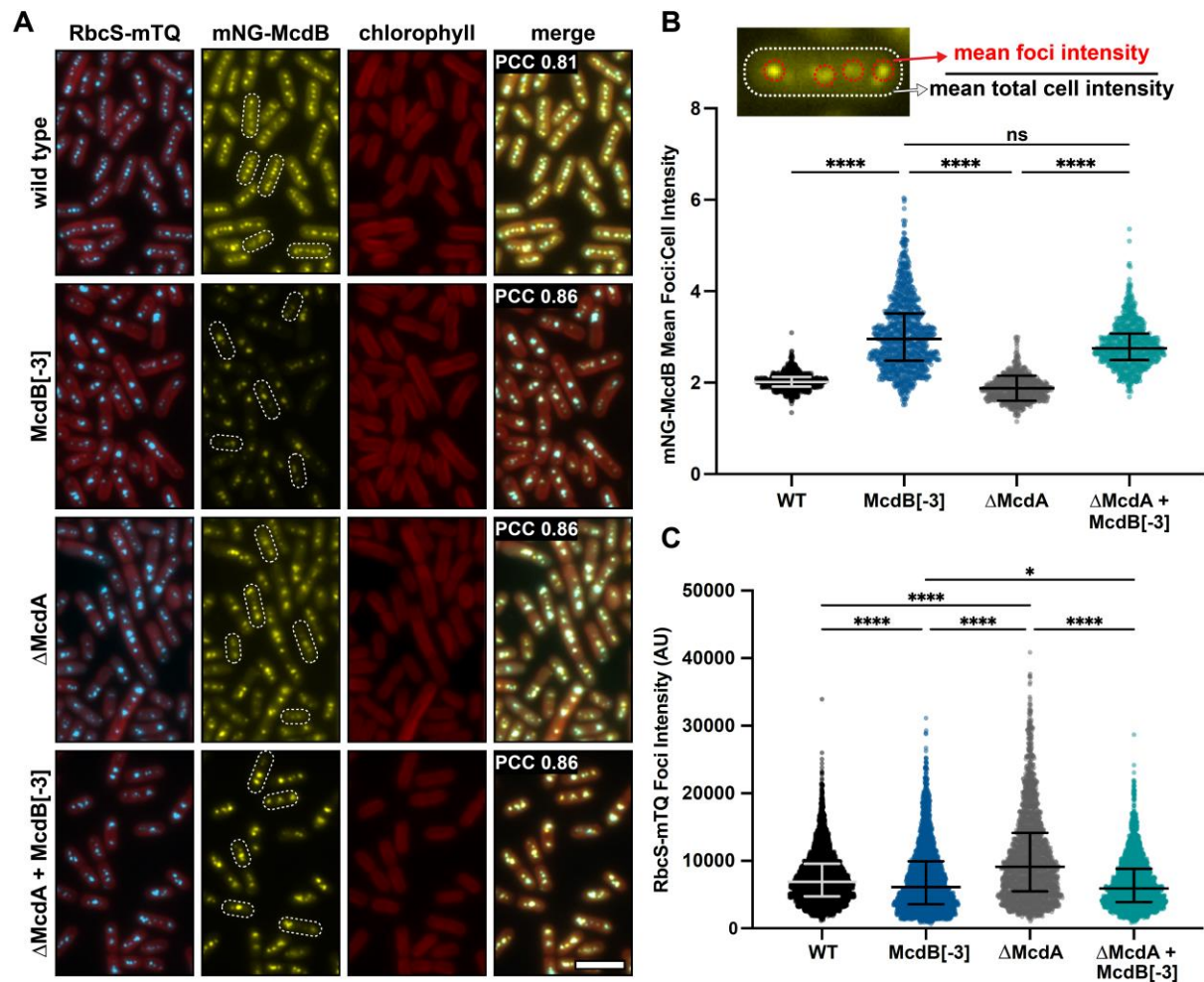
Figure 3.16: McdB[-3], which results in a high degree of condensate solubilization *in vitro* and in *E. coli*, alters the soluble fraction of McdB and carboxysome Rubisco levels *in vivo*. (**A**) Representative fluorescence microscopy images of the indicated strains. Scale bar = 5 µm and applies to all images. Pearson's Correlation Coefficients (PCC) are shown for mNG-McdB and RbcS-mTQ for each strain. PCC values are means from > 10,000 cells over 10 fields of view. (**B**) Quantification of (mean foci intensity / mean total cell intensities) for mNG-McdB of n > 500 cells. Medians and interquartile ranges are displayed. **** $p < 0.001$ based on Kruskal-Wallis ANOVA. (**C**) Quantification of mean RbcS-mTQ foci intensity for n > 500 cells. Medians and interquartile ranges are displayed. *$p < 0.05$; ****$p < 0.001$ based on Kruskal-Wallis ANOVA.
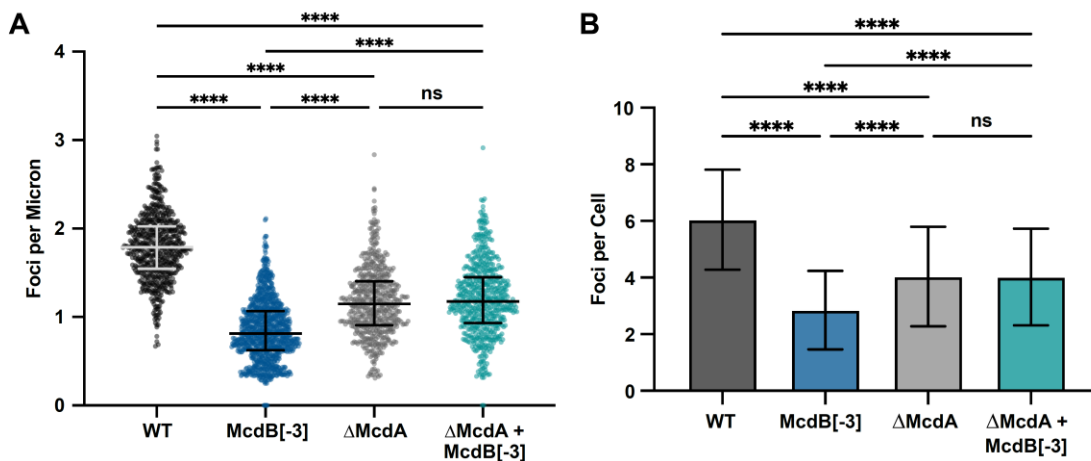
Figure 3.17: Deletion of McdA causes no additional loss of carboxysome positioning in McdB[-3] strain. (**A**) Quantification of RbcS-mTQ foci per micron from n > 500 cells. Medians and interquartile ranges are displayed. **** p < 0.001 based on Kruskal-Wallis ANOVA. (**B**) Quantification of RbcS-mTQ foci per cell for n > 500 cells. Medians and interquartile ranges are displayed. ****p < 0.001 based on Kruskal-Wallis ANOVA.

3.3.11 Condensate-defective McdB[-3] has a reduced cytoplasmic phase and associates with carboxysomes with lowered Rubisco content

Although carboxysome mispositioning by McdB[-3] cannot be directly ascribed to defects in McdB condensate formation specifically, two other observed phenotypes are not explained by a loss in McdA interaction. First, McdB[-3] still strongly colocalized with carboxysomes (PCC = 0.86 ± 0.01, n > 10,000 cells), similar to that of both McdB[wt] (PCC = 0.81 ± 0.01, n > 10,000 cells), and association did not change in the absence of McdA (PCC = 0.86 ± 0.02, n > 10,000 cells) (Figure 3.16A). The data strongly suggests that condensate formation is not required for McdB to associate with carboxysomes.

Strikingly, however, the cytoplasmic phase observed for McdB[wt] was significantly lower for McdB[-3], both in the presence or absence of McdA (Figure 3.16A). Indeed, when

quantifying the intensity ratio of carboxysome-associated McdB to that of the whole cell,

McdB[-3] showed a significant deviation in the ratio that was independent of McdA (Figure

3.16B). The data show that while McdB condensate formation is not required for carboxysome

association, without this activity, the cytoplasmic fraction of McdB notably declines.

Finally, we set out to directly determine the effect of McdB[-3] on the carboxysome itself

by quantifying encapsulated Rubisco. Intriguingly, the McdB[-3] strains, with or without McdA,

had carboxysomes with significantly lower Rubisco content as quantified by RbcS-mTQ

intensity (Figure 3.16C). This finding was particularly striking in the Δ*mcdA* background

because, as we have shown previously, deletion of McdA results in increased RbcS-mTQ foci

intensity due to carboxysome aggregation [46]. But with McdB[-3], RbcS-mTQ intensity

decreased, even with McdA deleted (Figure 3.16C). Together, these data show that McdB[-3]

increases the carboxysome-bound to soluble-McdB ratio and decreases Rubisco content in

carboxysomes. These phenotypes are not explained by the loss of interaction with McdA, and are

therefore potentially linked to defects in McdB condensate formation.

3.4 Discussion

In this report, we generate an initial structural model of *Se* McdB based on several

empirical and predictive approaches. We define a tripartite domain architecture with an N-

terminal IDR, a stable CC domain, and a CTD consisting of several α-helices (Figure 3.3). We

show that the CC dimerizes McdB and the CTD trimerizes the dimer, resulting in a trimer-of-

dimers hexamer. The IDR had no impact on oligomerization (Figure 3.4). Next, we found that

McdB forms condensates via pH-dependent PSCP, where condensates show time-dependent

viscoelastic properties and McdB forms pH-dependent clusters at concentrations far below the

observed c_{sat} (Figure 3.6). Using truncations, we found that the CC domain drives condensate

formation, the IDR modulates solubility, and the CTD provides further valency. Therefore, all

three domains are required for achieving wild-type levels of condensate formation (Figure 3.7).

We then identified positive residues in McdB important for stabilizing condensates. By

performing scanning mutagenesis in both the IDR and CTD, we show that while mutations to the

CTD destabilize the McdB hexamer, substituting out basic residues in the IDR solubilized

condensates without affecting McdB structure or oligomerization (Figure 3.10). These findings

allowed us to design a series of mutants where the net charge of the IDR tuned McdB condensate

solubility both *in vitro* (Figure 3.14) and in *E. coli* (Figure 3.15). Lastly, we found that a

solubilized McdB mutant, McdB[-3], impacts the carboxysome-bound to soluble McdB ratio in

the cell, as well as Rubisco content in carboxysomes (Figure 3.16). Overall, we determined

McdB domain architecture, its oligomerization domains, regions required for condensate

formation, and how to fine-tune condensate solubility, allowing us to link McdB condensate

formation to potential functions *in vivo* (Figure 3.18).
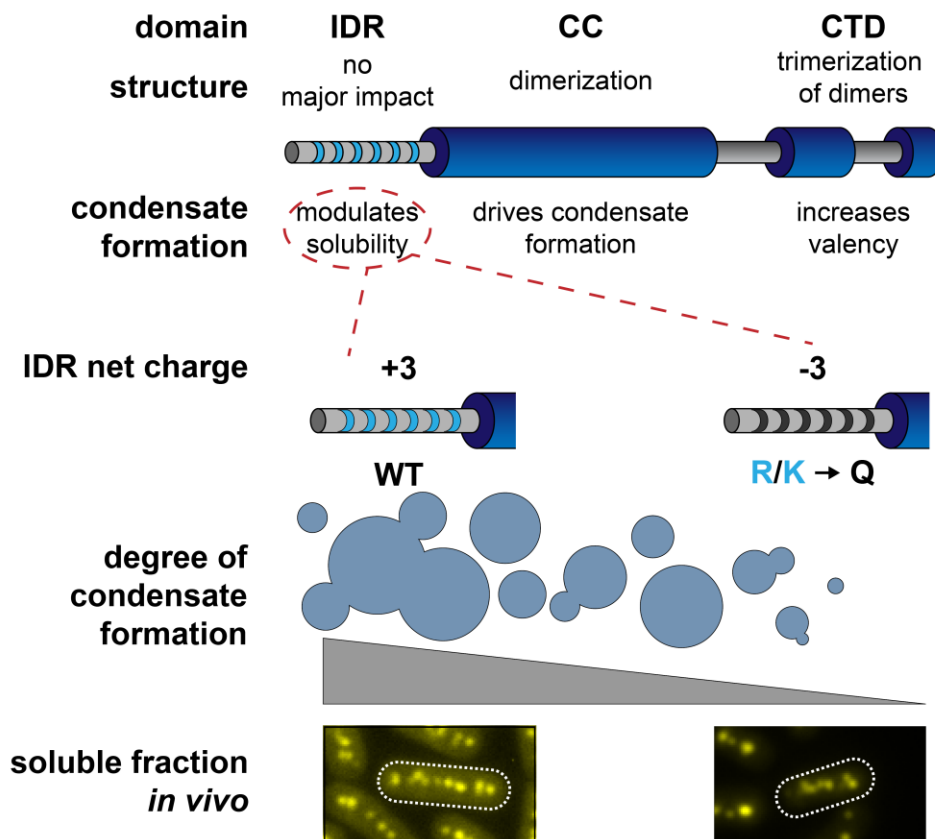
Figure 3.18: Proposed model of *Se* McdB domain structure and self-association. The central coiled-coil domain (CC) is necessary and sufficient for dimerization and driving condensate formation. The α-helical C-terminal domain (CTD) trimerize McdB dimers and increases the degree of condensate formation compared to the CC alone. The N-terminal IDR does not affect oligomerization and increases the degree of condensate formation compared to the CC alone. Substituting basic residues (K/R) in the IDR to glutamines (Q) can tune condensate solubility *in vitro* without affecting McdB oligomerization. These mutations allowed us to identify *in vivo* phenotypes correlated specifically to McdB phase separation, including the relative amount of soluble McdB.

3.4.1 McdB condensate formation follows a nuanced, multi-domain mechanism

As the field of biomolecular condensates advances, more nuanced mechanisms are arising that describe combinations of condensate-driver domains, solubility modulators, and influences of oligomerization. We see here that different domains of McdB influence condensate

formation via effects on solubility, self-association, and network formation. For many proteins, IDRs have been shown to be necessary and sufficient for driving condensate formation to a degree that is comparable to the full-length protein [47, 48, 49, 50, 51]. The IDR of McdB, on the other hand, did not form condensates even at high protein concentrations and in the presence of a crowder. Furthermore, while deleting the IDR did not prevent McdB condensate formation, substituting only six basic residues in the IDR with glutamines solubilized condensates both *in vitro* and in *E. coli*. In line with our findings, recent models have described how charged residues within IDRs can serve as key modulators of protein phase separation via mediating interactions with the solvent [52, 53, 54].

Glutamine-rich regions are known to be involved in stable protein-protein interactions such as in coiled-coils and amyloids [55, 56], and expansion of glutamine-rich regions in some condensate-forming proteins leads to amylogenesis and disease [57, 58]. Thus, one potential caveat here is that introduction of glutamines may lead to amylogenesis for McdB. However, when we introduced glutamines into the IDR of McdB, solubility was increased both *in vitro* and *in vivo* without any impact on hexamerization. These findings are especially striking for a condensate-forming protein as glutamines are generally thought to stabilize and coarsen protein condensates [34]. On the contrary, we see that increased glutamine content can solubilize condensates and is therefore largely context dependent. These findings expand our understanding of the molecular grammar of biomolecular condensates.

Oligomerization has also been found to influence protein condensate formation. For example, some proteins require oligomerization to provide the multivalency needed to form condensates [37, 59, 60], where some IDRs only induce condensate formation when fused to an oligomerizing domain [36]. In some cases, oligomerization domains, like coiled-coils, drive

condensate formation and are modulated by other domains [61], similar to our findings here with McdB. Such is the case for the bacterial protein PopZ, where an oligomerization domain forms condensates with solubility modulated by an IDR [19].

Truncated proteins have been useful in the study of biomolecular condensates. But it is important to note that using truncation data alone to dissect modes of condensate formation can lead to erroneous models since entire regions of the protein are missing. However, data from our truncation and substitution mutants were entirely congruent. For example, deletion of the CTD or substitutions to this region caused destabilization of the hexamer to a dimer, and deletion of the IDR or substitutions to this region caused solubilization of condensates without affecting hexamer formation. Furthermore, it should be noted that the McdB constructs used in our *in vitro* assays were free from fluorescent proteins, organic dyes, or other modifications that may influence phase separation. Therefore, the observed material properties of these condensates have full dependence on the McdB sequence.

3.4.2 McdB homologs have polyampholytic properties between their N- and C-termini

In our previous bioinformatic study, we found that McdB homologs possess features that enable condensate formation, including intrinsic disorder, low hydrophobicity, biased amino acid compositions, and multivalency [8]. Intriguingly, we also found that McdB proteins were potentially polyampholytes, with biphasic charge distributions between the NTD and CTD flanking the CC domain. For *Se* McdB, the N-terminal IDR has a pI of 10.8 and the CTD has a pI of 6.8, suggesting a potential electrostatic interaction between the two termini. The reason for such a shared feature was not obvious, but we proposed this polyampholytic nature was important for McdB self-association. A structure of the CC domain of a plasmid-encoded McdB-

like protein from the cyanobacterium *Cyanothece sp.* PCC 7424 displayed an antiparallel association to form a dimer [62]. Antiparallel dimerization of the CC domains of *Se* McdB would align these oppositely charged extensions. Consistently, our truncation data provides evidence suggesting a condensate-stabilizing interaction between the N-terminal IDR and the CTD. Condensates from the IDR+CC construct, which lack the CTD, were highly dynamic. But with the CTD present, condensates fused slowly. We were unable to dissect how the CTD contributes to condensate formation as substitutions to the CTD caused hexamer destabilization. However, these results have set the stage for several future studies that will (*i*) probe the orientation of McdB subunits within the hexamer, (*ii*) determine how basic residues in the IDR influence McdB condensate formation at a molecular level, and (*iii*) identify residues in the CTD that may influence condensate formation.

3.4.3 Considerations for McdB condensate formation in vivo

Investigating the roles of biomolecular condensates is especially challenging in bacteria, due in large part to size limitations that require a combination of techniques often including super-resolution microscopy [18]. While these approaches would certainly advance our understanding of the function of McdB in the cell, we begin here by taking a biochemical approach to identify mutation sets that affect condensate solubility *in vitro* and investigating the effects of these mutations *in vivo*. By mutating positively charged residues in the IDR of McdB to glutamines, we were able to solubilize condensates and shift McdB molecules into the dilute phase both *in vitro* and in *E. coli*. Surprisingly, this same mutation set seemed to have an opposite effect *in vivo*, where the dilute phase was diminished and carboxysome-associated McdB foci became brighter (see Figure 3.16). This suggests that McdB condensate formation is

64

not required for foci formation *in vivo*, but rather that phase separation may negatively regulate

carboxysome binding. *In vitro*, condensate formation represents a major form of McdB self-

association, sequestering away a large portion of McdB molecules into self-enriched condensates

where they have a lessened interaction with the surrounding solvent. *In vivo*, these self-

associations may similarly serve to regulate the amount of McdB free to interact with

carboxysomes. For instance, if positive charge in the N-terminus of McdB interacts with

negative charge in the C-terminus of McdB, and the C-terminal region is also required for

carboxysome association, self-associations may compete with carboxysome binding. Thus, by

greatly reducing the tendency of McdB to self-associate, we may be increasing the amount of

McdB free to interact with carboxysomes. We find models such as this especially attractive as

they provide a potential means for McdB to be regulated via pH in a manner consistent with the

literature described below. To test this model, we would first need to have a better understanding

of the regions of McdB required for carboxysome-association and how these relate to the model

of McdB condensate formation and its polyampholytic properties.

3.4.4 pH as a potential underlying regulator for McdB condensate solubility and its association with carboxysomes

In carbon-fixing organisms, the collection of processes that contribute to efficient carbon

fixation are referred to as the carbon concentrating mechanism (CCM). The development of a

model for the cyanobacterial CCM has provided insight into how different features, such the

presence of carboxysomes, affect overall carbon capture [63]. It was shown that incorporation of

a pH flux into CCM models provides values more consistent with experimentation [64]. This

updated 'pH-aware' model suggests that the carboxysome lumen maintains a lower pH than the

cytoplasm via Rubisco proton production; with the cytoplasm being ~ pH 8.5 and the carboxysome lumen being ~ pH 7.5 [64, 65].

Here, we show pH as a major regulator of McdB condensate solubility *in vitro*. Furthermore, we report a potential link between McdB condensate solubility and regulating both the carboxysome-bound fractions of McdB as well as Rubisco content within carboxysomes. We speculate that intracellular differences in pH may influence McdB self-associations and thus regulate carboxysome binding like our model described in the previous section.

Future studies will determine the nature of McdB association with carboxysomes, and how condensate formation influences this association. For example, we have shown that McdB strongly associates with carboxysome shell proteins via bacterial two-hybrid assays [7]. It is attractive to speculate that McdB undergoes pre-wetting interactions with the 2D surface of the carboxysome shell, which then nucleates surface-assisted condensation. Such 2D interactions would significantly impact McdB phase boundaries [66]. It is also possible that McdB phase separation directly modulates carboxysome fluidity and Rubisco content. Carboxysomes were traditionally thought of as paracrystalline, but recent data in *S. elongatus* shows that carboxysome biogenesis begins with Rubisco forming a condensate with its linker protein CcmM [67]. Intriguingly, our bacterial-two hybrid assays have shown that, in addition to shell proteins, McdB also interacts strongly with CcmM. It is therefore possible that McdB and its phase separation activity influences the carboxysome composition, fluidity, and function through interactions with the Rubisco-CcmM condensate. Such a model helps explain defects in Rubisco loading in our condensation-deficient McdB stain. Teasing apart the stable protein-protein interactions between McdB and carboxysomes from the dynamic processes governing condensate solubility will therefore be of significant importance.

66

3.4.5 Tunable protein condensates as useful tools for synthetic biology

A useful property of biomolecular condensates is the ability to regulate enzyme activity [68]. Specific chemistries within condensates can affect the degree to which certain metabolites and enzymes are soluble within the dense phase. Thus, condensates can serve as reaction centers that regulate the overall metabolism of a cell by transiently altering the activities of key reactions. For example, it has been shown that certain scaffolding proteins can form phase separated condensates with Rubisco [67, 69]. It is speculated that these Rubisco condensates were the original CCM, which then led to the evolution of carboxysomes and the modern CCM [65]. An exciting future direction for the field of biomolecular condensates is the prospect of designing condensate forming scaffolding proteins that can recruit specific enzymes, such as Rubisco, and implementing these designer enzyme-condensates in synthetic cells to engineer metabolism [19, 20, 68].

Here, we find that the IDR of McdB, which does not itself drive condensate formation, is amenable to mutations that fine-tune McdB condensate properties. The bacterial protein PopZ has already been engineered to fine-tune condensate formation and has also been developed as a tool called the "PopTag", which endows condensate forming activity to fusion proteins expressed in a variety of cell types, including human cells [20]. As McdB and PopZ differ at the primary, secondary, tertiary, and quaternary levels, it is expected that they could be used as separate tags to design coexisting but immiscible condensates, thus expanding design potential and the repertoire of condensate-related tools for synthetic biology.

3.5 Acknowledgements

3.6 Materials and Methods

3.6.1 Protein expression and purification

Proteolysis was performed on *Se* McdB at 30 µM in buffer containing 150 mM KCl, 50 mM HEPES pH 7.7, and 2 mM BME. Trypsin protease (Thermo-Fischer) was added at a 1:100 ratio of protease:protein. The reaction was incubated at 30ºC and samples were quenched at the indicated time points by diluting into 4X Laemmli SDS-PAGE sample buffer containing 8% SDS. Degradation over time was visualized by running time points on a 4–12% Bis-Tris NuPAGE gel (Invitrogen) and staining with InstantBlue Coomassie Stain (Abcam).

Bands that were N-terminally sequenced were separated via SDS-PAGE as above, but transferred to a PVDF membrane (Bio-Rad) prior to staining. Transfer of bands was performed using a Trans-Blot Turbo Transfer System (Bio-Rad). N-terminal sequences of these bands were then determined using Edman degradation.

3.6.2 Circular dichroism

For all protein samples analyzed, far-UV CD spectra were obtained using a J-1500 CD spectrometer (Jasco). All measurements were taken with 250 µL of protein at 0.25 mg/mL in 20 mM KPi, pH 8.0. Measurements were taken using a quartz cell with a path length of 0.1 cm. The spectra were acquired from 260 to 190 nm with a 0.1 nm interval, 50 nm/min scan speed, and at 25°C unless otherwise stated.

3.6.3 Microscopy of protein condensates

Samples for imaging were set up in 16 well CultureWells (Grace BioLabs). Wells were passivated by overnight incubation in 5% (w/v) Pluronic acid (Thermo-Fischer), and washed thoroughly with the corresponding buffer prior to use. All condensate samples were incubated for 30 minutes prior to imaging unless otherwise stated. For experiments where samples were imaged across pH titrations, the following buffers were used: phosphate buffer for pH 6.3-6.7, HEPES for pH 7.2-7.7, and Tris-HCl for 8.2-8.6. Imaging of condensates was performed using a Nikon Ti2-E motorized inverted microscope (60 × DIC objective and DIC analyzer cube) controlled by NIS Elements software with a Transmitted LED Lamp house and a Photometrics Prime 95B Back-illuminated sCMOS Camera. Image analysis was performed using Fiji v 1.0.

3.6.4 Fluorescence recovery after photobleaching (FRAP)

All FRAP measurements were performed using the indicated protein concentration with the addition of 1:1000 mNG-McdB based on molarity. All fluorescence imaging was performed

using a Nikon Ti2-E motorized inverted microscope controlled by NIS Elements software with a

SOLA 365 LED light source, a 100× objective lens (Oil CFI60 Plan Apochromat Lambda Series

for DIC), and a Photometrics Prime 95B Back-illuminated sCMOS camera. mNG signal was

acquired using a "GFP" filter set [excitation, 470/40 nm (450 to 490 nm); emission, 525/50 nm

(500 to 550 nm); dichroic mirror, 495 nm]. Bleaching was conducted with a 405-nm laser at 40%

power (20 mW) with a 200-µs dwell time. Recovery was monitored with a time-lapse video with

5 sec intervals for 20 mins. Image analysis was done in Fiji v 1.0. Intensities from bleached

regions of interest (ROIs) were background subtracted and normalized using an unbleached

condensate to account for any full field of view photobleaching. The values for each condensate

were then normalized such that a value of 1 was set to the pre-bleach intensity and a value of 0

was set to the intensity immediately post-bleaching. Data were exported, further tabulated,

graphed, and analyzed using GraphPad Prism 9.0.1 for macOS (GraphPad Software, San Diego,

CA, www.graphpad.com).

3.6.5 Dynamic light scattering (DLS)

All sizing and polydispersity measurements were carried out on an Uncle by Unchained

Labs (USA) at 25°C in triplicate. Samples were adjusted to the indicated concentrations in 150

mM KCl and 20 mM of the following buffers based on pH: HEPES, pH 7.2; Tris-HCl, pH 8.2;

CAPS, pH 10.2. Samples were analyzed both before and after a centrifugation step at 20,000 rcf

for 10 min as indicated. Data were exported, further tabulated, graphed, and analyzed using

GraphPad Prism 9.0.1 for macOS (GraphPad Software, San Diego, CA, www.graphpad.com).

3.6.6 Size-exclusion chromatography (SEC)

SEC was performed on full-length and truncated McdB proteins using a Superdex 200 Increase 10/300 GL (Cytiva) column connected to an AKTA pure system (Cytiva). 500 µL of sample at 1.5 mg/mL protein was passed through the column at 0.4 mL/min in buffer [150 mM KCl; 20 mM Tris-HCl pH 8.2] while monitoring absorbance at 220 nm.

3.6.7 SEC coupled to multi-angled light scattering (SEC-MALS)

For each sample analyzed, 50 µL at 1.5 mg/ml was passed over an SEC column (PROTEIN KW-804; Shodex) at a flow rate of 0.4 mL/min in buffer containing 150 mM KCl and 20 mM Tris-HCl, pH 8.2. Following SEC, the samples were analyzed using an A280 UV detector (AKTA pure; Cytiva), the DAWN HELEOS-II MALS detector with an internal QELs (Wyatt Technology), and the Optilab T-rEX refractive index detector (Wyatt Technology). The data were analyzed to calculate mass using ASTRA 6 software (Wyatt Technology). Bovine serum albumin was used as the standard for calibration.

3.6.8 Phase diagrams

Data for phase diagrams was collected using an Infinite M200 PRO plate reader (Tecan). Samples were set up in 96 well glass-bottom plates (Thomas Scientific) and absorbance at 350 nm was measured as previously described [15]. Reported values are averages of triplicates with buffer blanks subtracted, and error bars representing standard deviations. Protein concentration, KCl concentration, and pH values varied as indicated, but for each pH value tested, 20 mM of

the following buffers were used: phosphate buffer for pH 6.3-6.7, HEPES for pH 7.2-7.7, and Tris-HCl for pH 8.2-8.6.

3.6.9 Quantification of phase separation via centrifugation

Centrifugation was used to quantify the degree to which McdB and its variants condensed under certain conditions, as previously described [15]. Briefly, 100 µL of sample was incubated at the conditions specified for 30 minutes, and then centrifuged at 16,000 rcf for 2 minutes. The supernatant was removed and the pellet resuspended in an equal volume of McdB solubilization buffer [300 mM KCl, 20 mM CAPS pH 10.2]; McdB does not condense at pH 10.2. Samples were then diluted into 4X Laemmli SDS-PAGE sample buffer. Pellet and supernatant fractions were visualized on a 4–12% Bis-Tris NuPAGE gel (Invitrogen) by staining with InstantBlue Coomassie Stain (Abcam) for 1 hour and then destaining in water for 14-16 hours. The intensities of the bands were quantified using Fiji v 1.0 and resultant data graphed using GraphPad Prism 9.0.1 for macOS (GraphPad Software, San Diego, CA, www.graphpad.com).

3.6.10 Expression and visualization of mCherry fusions in E. coli

All constructs were expressed off a plasmid from a pTrc promoter in *E. coli* MG1655. Overnight cultures grown in LB + carbenicillin (100 µg/mL) were diluted at 1:100 into AB medium + carbenicillin (100 µg/mL) supplemented with (0.2% glycerol; 10 µg/mL thiamine; 0.2% casein; 25 µg/mL uracil). Cultures were grown at 37°C to an OD600 = 0.3 and induced with 1 mM IPTG. Following induction, cultures were grown at 37°C and samples taken at the indicated time points.

Cells used for imaging were prepared by spotting 3 µL of cells onto a 2% UltraPure agarose + AB medium pad on a Mantek dish. Images were taken using Nikon Ti2-E motorized inverted microscope controlled by NIS Elements software with a SOLA 365 LED light source, a 100X Objective lens (Oil CFI Plan Apochromat DM Lambda Series for Phase Contrast), and a Hamamatsu Orca Flash 4.0 LT + sCMOS camera. mCherry signal was imaged using a "TexasRed" filter set (C-FL Texas Red, Hard Coat, High Signal-to-Noise, Zero Shift, Excitation: 560/40 nm [540-580 nm], Emission: 630/75 nm [593-668 nm], Dichroic Mirror: 585 nm). Image analysis was performed using Fiji v 1.0.

To monitor expression levels, cells were harvested via centrifugation at the indicated time points, and resuspended in 4X Laemmli SDS-PAGE sample buffer to give a final OD600 = 4. Samples were boiled at 95ºC and 10 µL were then run on a 4–12% Bis-Tris NuPAGE gel (Invitrogen). Bands were visualized by staining with InstantBlue Coomassie Stain (Abcam) for 1 hour and then destaining in water for 14-16 hours. Quantifying the normalized band intensities was performed using Fiji v 1.0.

3.6.11 Growth and transformation of S. elongatus PCC 7942

All *S. elongatus* (ATCC 33912) strains were grown in BG-11 medium (Sigma) buffered with 1 g/L HEPES, pH 8.3. Cells were incubated with the following growth conditions: 60 µmol m–2 s–1 continuous LED 5600 K light, 32°C, 2% CO2, and shaking at 130 RPM. Transformations of *S. elongatus* cells were performed as previously described [70]. Transformants were plated on BG-11 agar with 12.5 µg/ml kanamycin. Single colonies were picked and transferred liquid BG-11 medium with corresponding antibiotic concentrations.

Complete gene insertions and absence of the wild-type gene were verified via PCR, and cultures were removed from antibiotic selection prior to imaging.

3.6.12 Live cell fluorescence microscopy and analysis

100 µL of exponentially growing cells (OD750 ~ 0.7) were harvested and spun down at 4000 rcf for 1 min and resuspended in 10 µl fresh BG-11. 2 µl of the resuspension were then spotted on 1.5% UltraPure agarose (Invitrogen) + BG-11 pad on a 35-mm glass-bottom dish (MatTek Life Sciences). All fluorescence and phase-contrast imaging were performed using a Nikon Ti2-E motorized inverted microscope controlled by NIS Elements software with a SOLA 365 LED light source, a 100× objective lens (Oil CFI Plan Apochromat DM Lambda Series for Phase Contrast), and a Photometrics Prime 95B back-illuminated sCMOS camera or Hamamatsu Orca-Flash 4.0 LTS camera. mNG-McdB variants were imaged using a "YFP" filter set (C-FL YFP, Hard Coat, High Signal-to-Noise, Zero Shift, Excitation: 500/20 nm [490–510 nm], Emission: 535/30 nm [520–550 nm], Dichroic Mirror: 515 nm). RbcS-mTQ-labeled carboxysomes were imaged using a "CFP" filter set (C-FL CFP, Hard Coat, High Signal-toNoise, Zero Shift, Excitation: 436/20 nm [426–446 nm], Emission: 480/40 nm [460-500 nm], Dichroic Mirror: 455 nm). Chlorophyll was imaged using a "TexasRed" filter set (C-FL Texas Red, Hard Coat, 583 High Signal-to-Noise, Zero Shift, Excitation: 560/40 nm [540-580 nm], Emission: 630/75 nm 584 [593-668 nm], Dichroic Mirror: 585 nm).

Image analysis including cell segmentation, quantification of foci number, intensities, and spacing were performed using Fiji plugin MicrobeJ 5.13n [71]. Cell perimeter detection and segmentation were done using the rod-shaped descriptor with default threshold settings. Carboxysome foci detection was performed using the point function with tolerance of 700 and

the sharpen image filter selected. McdB foci detection was performed using the smoothed foci function with tolerance of 100, Z-score of 3, and the minimum image filter selected. Pearson's correlation coefficients were calculated using ImageJ plugin JaCoP [72], and reported values represent means and standard deviations from > 10,000 cells over 10 fields of view. Data were exported, further tabulated, graphed, and analyzed using GraphPad Prism 9.0.1 for macOS (GraphPad Software, San Diego, CA, www.graphpad.com).

3.7 References

1. C. A. Kerfeld, C. Aussignargues, J. Zarzycki, F. Cai, M. Sutter, Bacterial microcompartments. *Nat. Rev. Microbiol.* **16**, 277–290 (2018).
2. M. Sutter, M. R. Melnicki, F. Schulz, T. Woyke, C. A. Kerfeld, A catalog of the diversity and ubiquity of bacterial microcompartments. *Nat. Commun.* **12**, 1–12 (2021).
3. T. O. Yeates, C. A. Kerfeld, S. Heinhorst, G. C. Cannon, J. M. Shively, Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat. Rev. Microbiol.* **6**, 681–691 (2008).
4. B. M. Long, *et al.*, Carboxysome encapsulation of the CO2-fixing enzyme Rubisco in tobacco chloroplasts. *Nat. Commun.* **9**, 1–14 (2018).
5. I. Flamholz, *et al.*, Functional reconstitution of a bacterial CO2 concentrating mechanism in E. coli. *eLife* **9**, 1–57 (2020).
6. D. F. Savage, B. Afonso, A. H. Chen, P. A. Silver, Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science (80-.).* **327**, 1258–1261 (2010).
7. J. S. MacCready, *et al.*, Protein gradients on the nucleoid position the carbon-fixing organelles of cyanobacteria. *eLife* **7** (2018).
8. J. S. MacCready, J. L. Basalla, A. G. Vecchiarelli, J. Echave, Origin and Evolution of Carboxysome Positioning Systems in Cyanobacteria. *Mol. Biol. Evol.* **37**, 1434–1451 (2020).
9. J. S. MacCready, L. Tran, J. L. Basalla, P. Hakim, A. G. Vecchiarelli, The McdAB system positions α-carboxysomes in proteobacteria. *Mol. Microbiol.* **116**, 277–297 (2021).
10. J. Lutkenhaus, The ParA/MinD family puts things in their place. *Trends Microbiol.* **20**, 411–418 (2012).
11. A. G. Vecchiarelli, K. Mizuuchi, B. E. Funnell, Surfing biological surfaces: exploiting the nucleoid for partition and transport in bacteria. *Mol. Microbiol.* **86**, 513–523 (2012).
12. B. E. Funnell, ParB partition proteins: Complex formation and spreading at bacterial and plasmid centromeres. *Front. Mol. Biosci.* **3**, 44 (2016).

13. P. Hakim, Y. Hoang, A. G. Vecchiarelli, Dissection of the ATPase active site of McdA reveals the sequential steps essential for carboxysome distribution. *Mol. Biol. Cell* **32** (2021).

14. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).

15. S. Alberti, A. Gladfelter, T. Mittag, Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).

16. T. Mittag, R. V. Pappu, A conceptual framework for understanding phase separation and addressing open questions and challenges. *Mol. Cell* **82**, 2201–2214 (2022).

17. G. L. Dignon, R. B. Best, J. Mittal, Biomolecular Phase Separation: From Molecular Driving Forces to Macroscopic Properties. *Annu. Rev. Phys. Chem.* **71**, 53–75 (2020).

18. C. A. Azaldegui, A. G. Vecchiarelli, J. S. Biteen, The emergence of phase separation as an organizing principle in bacteria. *Biophys. J.* **120**, 1123–1138 (2021).

19. W. Peeples, M. K. Rosen, Mechanistic dissection of increased enzymatic rate in a phase-separated compartment. *Nat. Chem. Biol.* **17**, 693–702 (2021).

20. K. Lasker, *et al.*, A modular platform for engineering function of natural and synthetic biomolecular condensates. *bioRxiv*, 2021.02.03.429226 (2021).

21. D. Viny, R. L. Levine, Drug modulation by nuclear condensates. *Science (80-.).* **368**, 1314–1315 (2020).

22. Y. Shin, *et al.*, Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell* **168**, 159-171.e14 (2017).

23. J. S. MacCready, A. G. Vecchiarelli, Positioning the model bacterial organelle, the carboxysome. *MBio* **12** (2021).

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

25. J. Yang, *et al.*, The I-TASSER suite: Protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2014).

26. J. Echave, S. J. Spielman, C. O. Wilke, Causes of evolutionary rate variation among protein sites. *Nat. Rev. Genet. 2016 172* **17**, 109–121 (2016).

27. F. Fiumara, L. Fioriti, E. R. Kandel, W. A. Hendrickson, Essential role of coiled-coils for aggregation and activity of Q/N-rich prions and polyQ proteins. *Cell* **143**, 1121 (2010).

28. P. P. De Laureto, *et al.*, Limited proteolysis of bovine α-lactalbumin: Isolation and characterization of protein domains. *Protein Sci.* **8**, 2290–2303 (2008).

29. S. Schopper, *et al.*, Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry. *Nat. Protoc.* **12**, 2391–2410 (2017).

30. Fontana, *et al.*, Probing protein structure by limited proteolysis *. *Acta Biochim Pol.* **51**, 299–321 (2004).

31. S. Xue, *et al.*, Low-complexity domain of U1-70K modulates phase separation and aggregation through distinctive basic-acidic motifs. *Sci. Adv.* **5**, 1–10 (2019).

32. M. Kar, *et al.*, Phase-separating RNA-binding proteins form heterogeneous distributions of clusters in subsaturated solutions. *Proc. Natl. Acad. Sci. U. S. A.* **119**, 1–12 (2022).

33. T. S. Harmon, A. S. Holehouse, M. K. Rosen, R. V. Pappu, Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *Elife* **6**, 1–31 (2017).

34. J. Wang, *et al.*, A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699.e16 (2018).

35. J. Stetefeld, S. A. McKenna, T. R. Patel, Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophys. Rev.* **8**, 409–427 (2016).

36. D. Bracha, *et al.*, Mapping Local and Global Liquid Phase Behavior in Living Cells Using Photo-Oligomerizable Seeds. *Cell* **175**, 1467-1480.e13 (2018).

37. J. Guillén-Boixet, *et al.*, RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell* **181**, 346-361.e17 (2020).

38. D. M. Mitrea, *et al.*, Self-interaction of NPM1 modulates multiple mechanisms of liquid–liquid phase separation. *Nat. Commun.* **9**, 1–13 (2018).

39. S. Xue, *et al.*, Low-complexity domain of U1-70K modulates phase separation and aggregation through distinctive basic-acidic motifs. *Sci. Adv.* **5** (2019).

40. L. Radnedge, B. Youngren, M. Davis, S. Austin, Probing the structure of complex macromolecular interactions by homolog specificity scanning: the P1 and P7 plasmid partition systems. *EMBO J.* **17**, 6076–6085 (1998).

41. N. V. Ravin, J. Rech, D. Lane, Mapping of Functional Domains in F Plasmid Partition Proteins Reveals a Bipartite SopB-recognition Domain in SopA. *J. Mol. Biol.* **329**, 875–889 (2003).

42. D. Barillà, E. Carmelo, F. Hayes, The tail of the ParG DNA segregation protein remodels ParF polymers and enhances ATP hydrolysis via an arginine finger-like motif. *Proc. Natl. Acad. Sci.* **104**, 1811–1816 (2007).

43. Y. Ah-Seng, F. Lopez, F. Pasta, D. Lane, J. Y. Bouet, Dual role of DNA in regulating ATP hydrolysis by the SopA partition protein. *J. Biol. Chem.* **284**, 30067–30075 (2009).

44. H. Ghasriani, *et al.*, Appropriation of the MinD protein-interaction motif by the dimeric interface of the bacterial cell division regulator MinE. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 18416–18421 (2010).

45. D. Schumacher, A. Harms, S. Bergeler, E. Frey, L. Søgaard-Andersen, PomX, a ParA/MinD ATPase activating protein, is a triple regulator of cell division in Myxococcus xanthus. *eLife* **10** (2021).

46. R. Rillema, Y. Hoang, J. S. MacCready, A. G. Vecchiarelli, Carboxysome mispositioning alters growth, morphology, and Rubisco level of the cyanobacterium Synechococcus elongatus PCC 7942. *MBio* **12** (2021).

47. S. Elbaum-Garfinkle, *et al.*, The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7189–7194 (2015).

48. L. D. Muiznieks, S. Sharpe, R. Pomès, F. W. Keeley, Role of Liquid–Liquid Phase Separation in Assembly of Elastin and Other Extracellular Matrix Proteins. *J. Mol. Biol.* **430**, 4741–4753 (2018).

49. H. J. Kim, *et al.*, Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467–473 (2013).

50. L. Darling, Y. Liu, C. J. Oldfield, V. N. Uversky, Intrinsically Disordered Proteome of Human Membrane-Less Organelles. *Proteomics* **18**, 1700193 (2018).

51. D. M. Shapiro, M. Ney, S. A. Eghtesadi, A. Chilkoti, Protein Phase Separation Arising from Intrinsic Disorder: First-Principles to Bespoke Applications. *J. Phys. Chem. B* **125**, 6740–6759 (2021).

52. M. J. Fossat, X. Zeng, R. V. Pappu, Uncovering Differences in Hydration Free Energies and Structures for Model Compound Mimics of Charged Side Chains of Amino Acids. *J. Phys. Chem. B* **125**, 4148 (2021).

53. X. Zeng, K. M. Ruff, R. V. Pappu, Competing interactions give rise to two-state behavior and switch-like transitions in charge-rich intrinsically disordered proteins. *bioRxiv*, 2022.01.11.475920 (2022).

54. M. J. Fossat, A. E. Posey, R. V. Pappu, Quantifying charge state heterogeneity for proteins with multiple ionizable residues. *Biophys. J.* **120**, 5438–5453 (2021).

55. F. Fiumara, L. Fioriti, E. R. Kandel, W. A. Hendrickson, Essential role of coiled-coils for aggregation and activity of Q/N-rich prions and polyQ proteins. *Cell* **143**, 1121 (2010).

56. M. D. Michelitsch, J. S. Weissman, A census of glutamine/asparagine-rich regions: Implications for their conserved function and the prediction of novel prions. *PNAS* **97**, 11910–11915 (2000).

57. B. Kokona, Z. P. Rosenthal, R. Fairman, Role of the coiled-coil structural motif in polyglutamine aggregation. *Biochemistry* **53**, 6738–6746 (2014).

58. M. J. Kwon, *et al.*, Coiled-coil structure-dependent interactions between polyQ proteins and Foxo lead to dendrite pathology and behavioral defects. *PNAS* **115**, E10748–E10757 (2018).

59. Wang, *et al.*, A single N-terminal phosphomimic disrupts TDP-43 polymerization, phase separation, and RNA splicing. *EMBO J.* **37**, e97452 (2018).

60. M. R. Marzahn, *et al.*, Higher-order oligomerization promotes localization of SPOP to liquid nuclear speckles. *EMBO J.* **35**, 1254–1275 (2016).

61. D. A. Ramirez, L. E. Hough, M. R. Shirts, Coiled-coil domains can drive liquid-liquid phase separation. *Biophys. J.* **122**, 209a (2023).

62. M. A. Schumacher, M. Henderson, H. Zhang, Structures of maintenance of carboxysome distribution Walker-box McdA and McdB adaptor homologs. *Nuceic Acids Research* **47**, 5950–5962 (2019).

63. N. Mangan, M. Brenner, Systems analysis of the CO2 concentrating mechanism in cyanobacteria. *eLife* **2014** (2014).

64. N. M. Mangan, A. Flamholz, R. D. Hood, R. Milo, D. F. Savage, pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5354–E5362 (2016).

65. B. M. Long, B. Förster, S. B. Pulsford, G. D. Price, M. R. Badger, Rubisco proton production can drive the elevation of CO2 within condensates and carboxysomes. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021).

66. J. A. Ditlev, Membrane-associated phase separation: organization and function emerge from a two-dimensional milieu. *J. Mol. Cell Biol.* **13**, 319–324 (2021).

67. H. Wang, *et al.*, Rubisco condensate formation by CcmM in β-carboxysome biogenesis. *Nature* **566**, 131–135 (2019).

68. B. G. O'Flynn, T. Mittag, The role of liquid–liquid phase separation in regulating enzyme activity. *Curr. Opin. Cell Biol.* **69**, 70–79 (2021).

69. L. M. Oltrogge, *et al.*, Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation. *Nat. Struct. Mol. Biol.* **27**, 281–287 (2020).

70. E. M. Clerico, J. L. Ditty, & S. S. Golden, Specialized Techniques for Site-Directed Mutagenesis in Cyanobacteria. *In: Rosato, E.* (eds) *Circadian Rhythms. Methods in Molecular Biology™, 362*. Humana Press. (2007).

71. A. Ducret, E. M. Quardokus, Y. V. Brun, MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. *Nat. Microbiol.* **1**, 16077 (2016).

72. S. Bolte, F.P. Cordelieres, A guided tour into subcellular colocalization analysis in light microscopy. *J Microsc* **224,** 213–232 (2006).

Chapter 4

An Invariant C-Terminal Tryptophan in McdB Mediates Its Interaction and Positioning Function with Carboxysomes[2]

4.1 Abstract

Bacterial microcompartments (BMCs) are widespread, protein-based organelles that regulate metabolism. The model for studying BMCs is the carboxysome, which facilitates carbon-fixation in several autotrophic bacteria. Carboxysomes can be distinguished as type α or ß, which are structurally and phyletically distinct. We recently characterized the Maintenance of Carboxysome Distribution (Mcd) systems responsible for spatially regulating α- and ß-carboxysomes, consisting of the proteins McdA and McdB. McdA is an ATPase that drives carboxysome positioning, and McdB is the adaptor protein that directly interacts with carboxysomes to provide cargo specificity. The molecular features of McdB proteins that specify their interactions with carboxysomes, and whether these are similar between α- and ß-carboxysomes, remain unknown. Here, we identify C-terminal motifs containing an invariant tryptophan necessary for α- and ß-McdBs to associate with α- and ß-carboxysomes, respectively. Substituting this tryptophan with other aromatic residues reveals corresponding gradients of

_____

[2] This chapter is based in full on the following preprint: *J. L. Basalla, M. Ghalmi, Y. Hoang, R. Dow, & A. G. Vecchiarelli. An invariant C-terminal tryptophan in McdB mediates its interaction and positioning function with carboxysomes. bioRxiv, 2023.11.21.568049 (2023)*. I generated all data, quantifications, and figures with the following exceptions: M. Ghalmi quantified much of the live cell microscopy data and generated several strains used for imaging; Y Hoang performed experiments and imaging related to Figure 4.12; R. Dow quantified data for Figure 4.12.

carboxysome colocalization and positioning by McdB *in vivo*. Intriguingly, these gradients also correlate with the ability of McdB to form condensates *in vitro*. The results reveal a shared mechanism underlying McdB adaptor protein binding to carboxysomes, and potentially other BMCs. Our findings also implicate condensate formation as playing a key role in this association.

4.2 Introduction

An important cellular feature across all domains of life is the compartmentalization of biological processes. Many bacteria possess protein-based organelles called bacterial microcompartments (BMCs) that provide subcellular compartmentalization and reaction isolation [1, 2]. BMCs consist of selectively-permeable protein shells that encapsulate a set of enzymes, thus serving as nanoscale reaction centers for key metabolic steps [2]. A recent bioinformatic survey identified 68 unique BMC types in 45 bacterial phyla [2], revealing that BMCs are widespread. Despite BMC prevalence and importance in diverse bacterial metabolisms, little is known about how BMCs are spatially regulated in the cell.

The best studied BMC type is the carboxysome [3]. Carboxysomes encapsulate the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) and co-concentrate it with its substrate CO_2 to significantly increase the efficiency of carbon fixation in many autotrophic bacteria [3]. As a result, carboxysomes are estimated to facilitate about 35% of all global carbon-fixation [4, 5], making carboxysomes of interest for developing carbon-capturing technologies [6, 7]. Beyond their biotechnological potential, carboxysomes are also the paradigm for understanding fundamental aspects of general BMC biology, such as assembly, structure, and spatial regulation [1, 2, 3]. Therefore, investigating the fundamental aspects of carboxysomes is

important for developing technologies as well as deepening our understanding of BMCs.

Two subtypes of carboxysomes exist, α and ß, where ß-carboxysomes are found in ß-cyanobacteria and α-carboxysomes are found in numerous phylogenetically distinct groups, including α-cyanobacteria and several types of chemoautotrophic bacteria [2]. While functionally equivalent, α- and ß-carboxysomes are structurally and phyletically distinct, with key differences in composition, mode of assembly, and regulation [8, 9]. In fact, α-carboxysomes are more closely related to other BMC-types than they are to β-carboxysomes [9]. Therefore, α- and ß-carboxysomes represent distinct BMC types, and comparative studies between the two have been critical for our understanding of carboxysome biology and BMCs in general.

Carboxysomes are spatially organized in the cell. In the model cyanobacterium *Synechococcus elongatus* PCC 7942 (*Se*, hereafter), a two-protein system is responsible for distributing ß-carboxysomes down the cell length [10, 11, 12]. One component, which we named Maintenance of carboxysome distribution protein A (McdA), is a member of the ParA/MinD family of ATPases known to position various genetic and protein-based structures in bacteria [13, 14]. The second component is a novel protein we named McdB, which interacts with McdA and also localizes to carboxysomes [10], thus acting as an adaptor to link the carboxysome cargo to its positioning ATPase [15]. Deletion of either McdA or McdB results in carboxysome aggregation and asymmetric inheritance of carboxysome clusters, slower cell growth, and a rapid loss of carboxysomes in the cell population [16, 17]. Therefore, uniform positioning maintains the carbon fixation efficiency of carboxysomes and ensures faithful inheritance of this vital organelle after cell division.

McdAB systems are widespread among β-cyanobacteria which contain β-carboxysomes, and proteobacteria which contain α-carboxysomes [11, 12]. Using the α-carboxysome model

organism *Halothiobacillus neapolitanus* (*Hn*, hereafter), we have shown that an McdAB system, distinct from that of β-carboxysomes, spatially distributes α-carboxysomes [12]. Therefore, McdAB is a cross-phylum two-protein system necessary for positioning both α- and β-carboxysomes. More broadly, putative McdAB systems were also identified for other BMCs involved in diverse metabolic processes. Understanding how the McdAB system spatially regulates carboxysomes therefore has broad implications for understanding BMC trafficking across bacteria.

One outstanding question is how the adaptor protein, McdB, connects to and provides specificity for the carboxysome cargo. Our previous studies of β-McdB from *Se* and α-McdB from *Hn* revealed extreme differences at the sequence and structural levels [11, 12]. For example, *Se* McdB is largely α-helical with a coiled-coil domain and forms a trimer-of-dimers hexamer [18], whereas *Hn* McdB is monomeric and completely intrinsically disordered [12]. Therefore, given the extreme diversity between α- and β-McdB proteins, it also remains to be determined whether they follow a similar mechanism to associate with α- and ß-carboxysomes. Despite their diversity, α- and β-McdB proteins have been shown to form condensates *in vitro* [11, 12, 18]. The processes underlying condensate formation *in vitro* can influence subcellular organization *in vivo* in both eukaryotes and prokaryotes [19, 20]. Along these lines, we recently found evidence suggesting that condensate formation by *Se* McdB may play a role in its association with carboxysomes *in vivo* [21]. To what extent condensate formation by McdB influences its association with carboxysomes, and whether this activity plays a role in the spatial organization of α- and ß-carboxysomes remains to be determined.

Here, we identified a C-terminal motif that contains an invariant tryptophan in both α - and β-McdB proteins. We determined this invariant tryptophan is essential for the association of McdB

with β-carboxysomes in *Se* and α-carboxysomes in *Hn*. Furthermore, expressing only this C-terminal motif containing the tryptophan and surrounding residues was necessary and sufficient for carboxysome association in *Hn*, but not in *Se*. We provide evidence to suggest *Se* McdB oligomerization is also required. We also show that putative McdB-like proteins associated with other BMC types encode invariant tyrosines, suggesting other aromatic residues can serve the same role as tryptophan [12]. By substituting the C-terminal tryptophan in *Hn* and *Se* McdB with other aromatic residues, we observed corresponding complementation gradients of carboxysome association and carboxysome positioning. Interestingly, these gradients of activity *in vivo* correlated with condensate formation *in vitro* for the purified aromatic mutants of McdB. Together, the results show that despite the extreme diversity between α- and β-McdB proteins and α- and β-carboxysomes, a similar mode of association is used. These results lay the groundwork for understanding the molecular mechanisms of protein association with the surface of the carboxysome and potentially other BMCs across bacteria.

4.3 Results

4.3.1 All McdB proteins encode an invariant tryptophan at their extreme C-terminus

We first performed multiple sequences alignments both within and across α- and ß-McdB types to identify regions of conservation that may be involved in associating with carboxysomes. On average, McdBs show low sequence identity; 14.8% among α-McdBs (Figure 4.1A), 14.9% among ß-McdBs (Figure 4.1B), and 6.7% across all McdBs (Figure 4.1C). One reason for this low
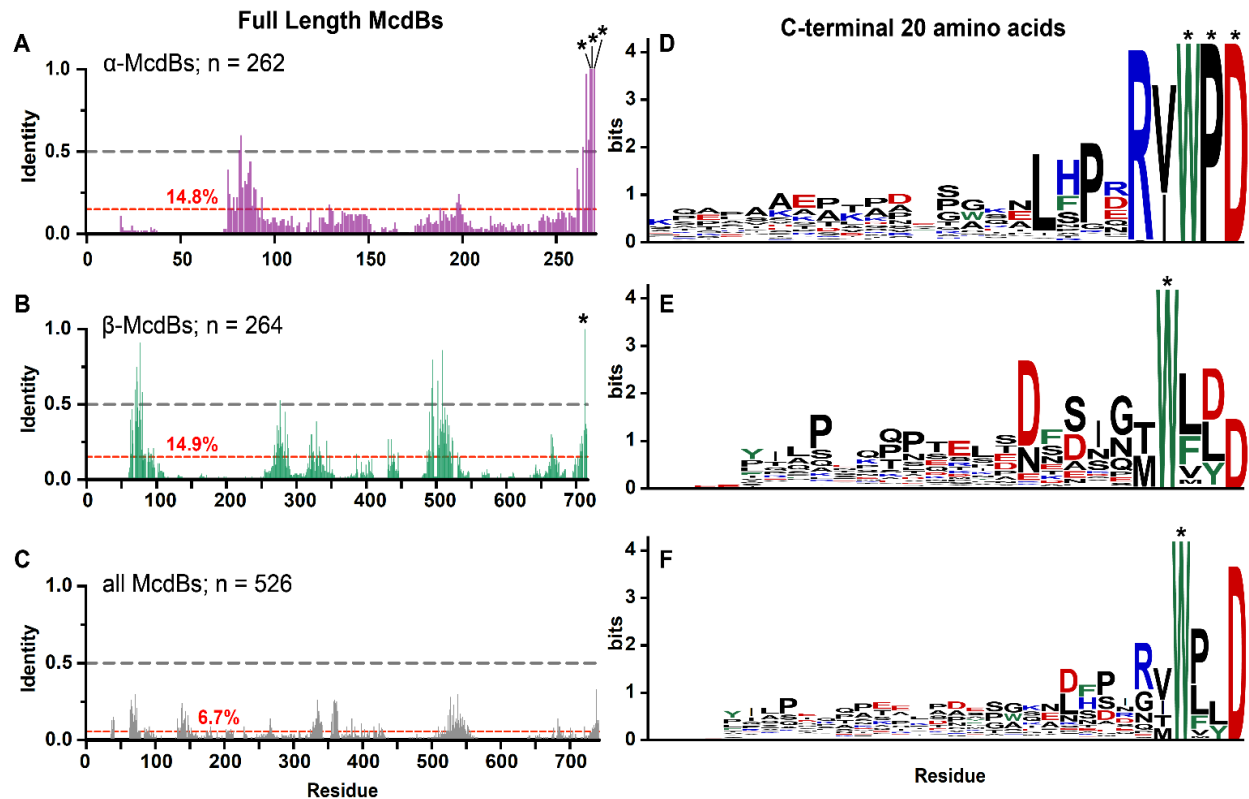
Figure 4.1: **All McdB proteins share an invariant C-terminal tryptophan.** (**A-C**) Percent identities from multiple sequence alignments of full-length (**A**) α-McdBs, (**B**) ß-McdBs, and (**C**) both α- and ß-McdBs. The average percent identity for each alignment is shown in red. (**D-F**) Sequence logos generated from multiple sequence alignments of only the last 20 C-terminal amino acids of (**D**) α-McdBs, (**E**) ß-McdBs, and (**F**) both α- and ß-McdBs. Positions that are invariant (100 percent identity) are indicated with an asterisk. Cationic residues are colored blue, anionic residues red, aromatic residues green, and all others black.

average identity is the large alignment gaps that stem from the high variance in sequence lengths, which range from 51 – 169 residues for α-McdBs and 132 – 394 residues for ß-McdBs (Figure 4.2A-C). Although the average identity was low, we identified three invariant residues (WPD) at the C-terminus of α-McdBs (Figure 4.1A), and one invariant residue (W) at the C-terminus of ß-McdBs (Figure 4.1B), consistent with our previous reports [11, 12]. Aligning all full-length McdB protein sequences did not identify any invariant residues (Figure 4.1C), again due to the

Figure 4.2: **McdB amino acid sequences are highly variable.** (**A**) A representative section of 50 sequences from the multiple sequence alignment of α-McdBs, centered on the sequence for *Hn* McdB (purple, starred). Dashes represent gaps in the alignment. (**B**) A representative section of 50 sequences from the multiple sequence alignment of ß-McdBs, centered on the sequence for *Se* McdB (green, starred). Dashes represent gaps in the alignment. (**C**) A representative section of 50 sequences from the multiple sequence alignment of all McdBs, centered on a region in which α-McdBs (purple) and ß-McdBs (green) align. Dashes represent gaps in the alignment. (**D**) Amino acid sequences of the full-length *Hn* and *Se* McdBs. The invariant tryptophan is colored green.

large variations in protein lengths, which prevented the C-termini of α- and ß-McdBs from

aligning (Figure 4.2C). We therefore repeated these alignments on only the last 20 C-terminal

amino acids of all McdB sequences to control for length variation (Figure 4.1D-F). These

alignments unveiled motif conservation specific to α - and ß-McdB types as well as conservation shared across all McdBs. α-McdBs contain a short consensus sequence of R(V/I)WPD at the extreme C-terminus (Figure 4.1D). The ß-McdB C-terminal motif is more degenerate, showing an enrichment of aspartic acids (D) within the last 10 C-terminal residues, along with the invariant tryptophan (Figure 4.1E). And across all McdBs, we identified the tryptophan as the only invariant residue (Figure 4.1F), suggesting a critical role in its functionality for both α- and ß-McdBs.

4.3.2 The C-terminal invariant tryptophan is required for McdB association with carboxysomes

Adaptor proteins typically associate with their cognate ParA/MinD positioning ATPase via basic residues in the N-terminus of the adaptor [22], and we have computationally shown that this is likely the case for McdB interactions with McdA [23]. Therefore, our identification of an invariant tryptophan within the C-terminus of all McdBs motivated our study for its potential role in associating with the carboxysome cargo, as opposed to McdA, which we expect McdB interacts with via its N-terminus.

We performed *in vivo* fluorescence microscopy in both *Se* and *Hn* cells to determine how McdB localization and carboxysome organization were altered for McdB mutants lacking the invariant C-terminal tryptophan. To visualize carboxysomes, the fluorescent protein monomeric Turquoise2 (mTQ) [24] was fused to the C-terminus of the small subunit of the Rubisco enzyme (RbcS in *Se*; CbbS in *Hn*). RbcS-mTQ and CbbS-mTQ were expressed using a second copy of their native promoters inserted at a neutral site, in addition to the wild type copy at the native locus. To simultaneously image McdB mutants in these carboxysome reporter strains, mutations

A

	mNG-McdB	RbcS-mTQ	merge

S. elongatus (Se) — wild type / ΔW152

B

PCC — wild type vs ΔW152 — ****

C

mTQ foci / micron — wild type vs ΔW152 — ****

D

	mNG-McdB	CbbS-mTQ	merge

H. neapolitanus (Hn) — wild type* / W94G

E

PCC — wild type* vs W94G — ****

F

unlabeled — wild type / W94G

G

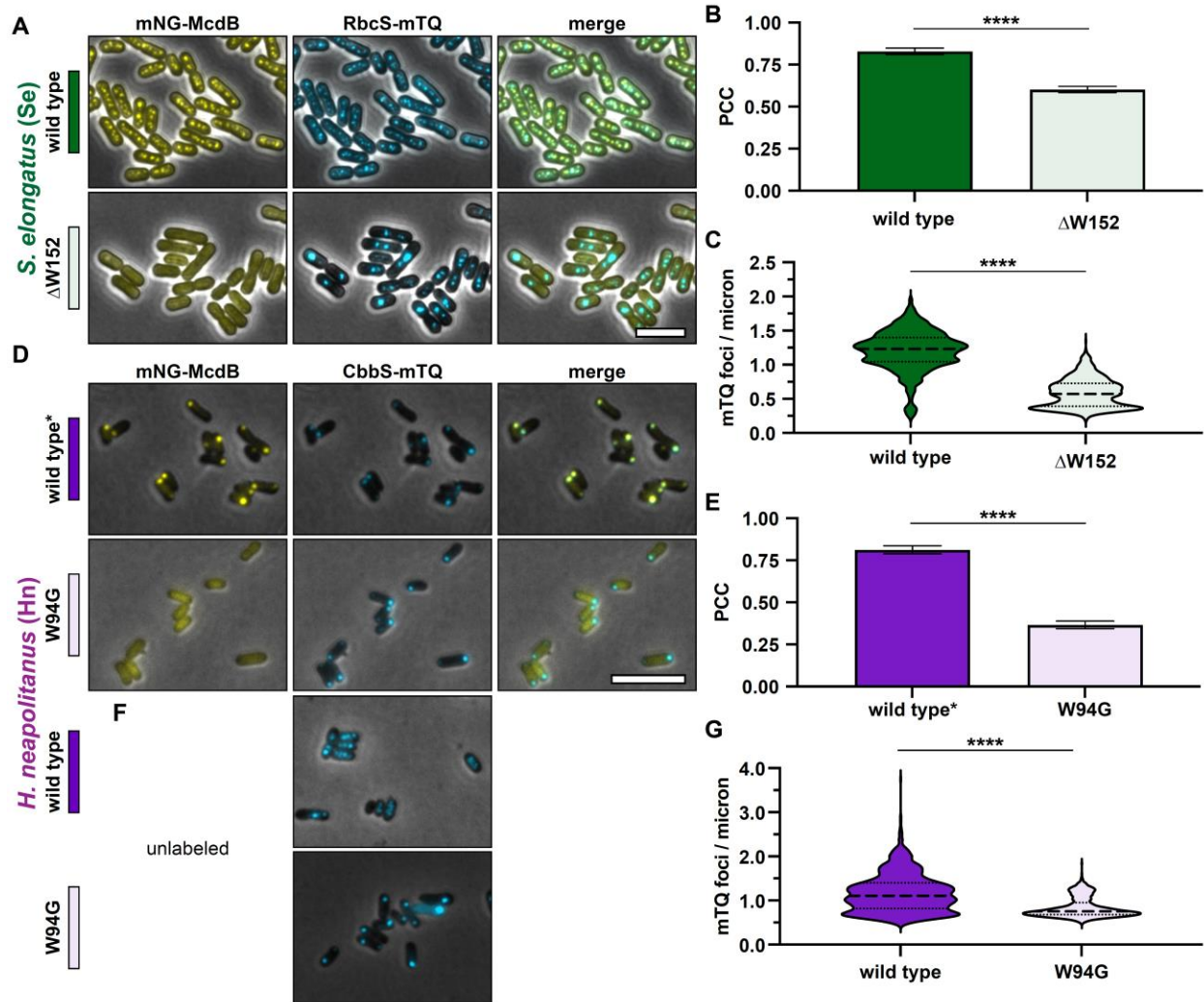mTQ foci / micron — wild type vs W94G — ****

88

Figure 4.3 (*previous page*): **The invariant tryptophan for both α- and β-McdBs mediates carboxysome localization.** (**A**) Representative microscopy images of the indicated *Synechococcus elongatus* (*Se*) strains. Phase contrast images are shown in black and white and overlaid with the fluorescence channels: mNG-McdB proteins are yellow and RbcS-mTQ labelled carboxysomes are cyan. Colored bars next to the strain names correspond to colors on the associated graphs. (**B**) Pearson's Correlation Coefficients (PCCs) quantified for the indicated *Se* strains. Graphs represent means and standard deviations from 7 technical replicates each with n > 500 cells. **** $p < 0.001$ from Welch's t-test. (**C**) Quantification of carboxysome spacing as number of mTQ foci divided by cell length. Graphs represent medians and interquartile ranges from 3 biological replicates each with n > 500 cells. **** $p < 0.001$ from Mann-Whitney U-test. (**D**) Representative microscopy images of the indicated *Halothiobacillus neapolitanus* (*Hn*) strains. Wild type* indicates the wild type McdB with an N-terminal mNG tag, which causes carboxysome aggregation in *Hn*. Phase contrast images are overlaid with the fluorescence channels: mNG-McdB proteins are yellow and CbbS-mTQ labelled carboxysomes are cyan. Colored bars next to the strain names correspond to colors on the associated graphs. (**E**). Pearson's Correlation Coefficients (PCCs) quantified for the indicated *Hn* strains. Graphs represent means and standard deviations from 7 technical replicates each with n > 500 cells. **** $p < 0.001$ from Welch's t-test. (**F**) As in (D), but with McdB not labeled with mNG. (**G**) Quantification of carboxysome spacing as number of mTQ foci divided by cell length. Graphs represent medians and interquartile ranges from 3 biological replicates each with n > 500 cells. **** $p < 0.001$ from Mann-Whitney U-test. Scale bars are 5 µm and apply to all images.

were made in an McdB variant that was N-terminally fused to the fluorescent protein monomeric NeonGreen (mNG) [25]. We have previously shown that, in *Se*, mNG-McdB is fully functional for carboxysome positioning when expressed as the only copy of McdB at its native locus [10]. In *Hn*, the mNG fusion unfortunately perturbs McdB interactions with McdA, resulting in carboxysome aggregation. However, this fusion still associates with carboxysomes and therefore remains a useful positive control for studying McdB-carboxysome association in *Hn* [12]. Finally, we also performed phase-contrast imaging to monitor cell morphology.

In *Se*, the invariant tryptophan is the final C-terminal amino acid (Figure 4.2D), therefore we simply deleted it to make McdB[ΔW152]. As shown previously in wild type *Se* cells [10], mNG-McdB colocalized with carboxysome foci that are uniformly distributed down the cell length (Figure 4.3A). Without the invariant tryptophan, mNG-McdB[ΔW152] was diffuse
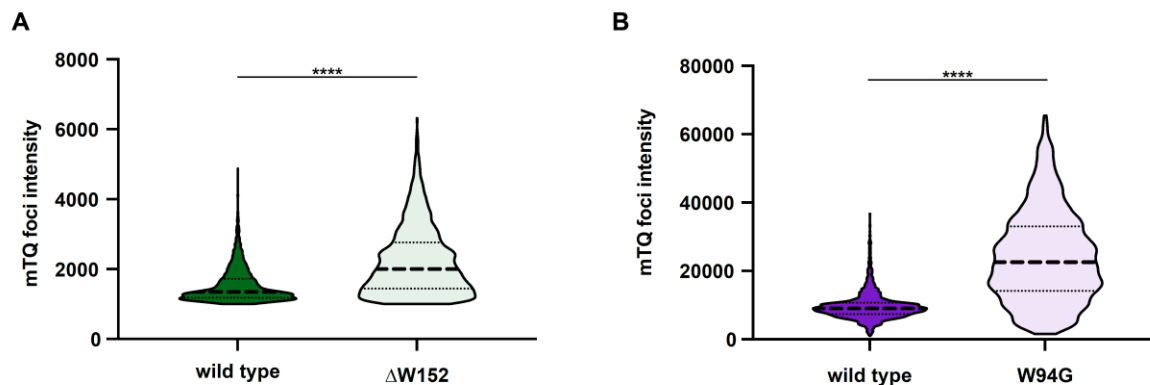
Figure 4.4: **Removal of the invariant tryptophan of McdB results in carboxysome aggregation.** (**A**) Quantification of RbcS-mTQ intensities from carboxysome foci in the indicated strains of *Se*. Graphs show medians and interquartile ranges from 3 biological replicates each with n > 1000 foci. **** p < 0.001 from Mann-Whitney U-test. (**B**) As in (A), but for *Hn* strains.

in the cell (Figure 4.3A), with no notable carboxysome colocalization compared to that of wild type (Figure 4.3B). Carboxysomes were also mispositioned (Figure 4.3C) and clustered into high-intensity aggregates in the McdB[ΔW152] strain (Figure 4.4B), which phenocopies a complete deletion of McdB, as shown previously [17].

For *Hn* McdB, the invariant tryptophan is the third amino acid from the C-terminus (Figure 4.2D). Therefore, a glycine substitution was used to make McdB[W94G]. Although the mNG-McdB fusion destroys its carboxysome positioning function, the fusion still strongly colocalized with the mispositioned carboxysome aggregates at the cell pole, hence the label "wild type*" (Figure 4.3D). As with *Se* McdB[ΔW152], mNG-McdB[W94G] was diffuse in *Hn* cells and did not colocalize with carboxysomes (Figure 4.3E).

While mNG-McdB[W94G] allowed us to observe changes in protein localization, we were unable to determine the effects of this mutant on carboxysome positioning when fused to mNG. Therefore, we also imaged carboxysome distribution in *Hn* cells with unlabeled

McdB[W94G] (Figure 4.3F). Consistent with *Se* McdB[ΔW152], carboxysomes became

mispositioned (Figure 4.3G) and clustered into high-intensity aggregates in the McdB[W94G]

mutant (Figure 4.4B), which phenocopies a complete deletion of McdB in *Hn*, as shown

previously [12]. Together, our results show that the C-terminal tryptophan found in all McdB

proteins is necessary for carboxysome association and positioning.

4.3.3 Loss of carboxysome association and positioning is not due to destabilization of McdB mutants

We found it striking that a single residue change completely destroyed McdB association and

positioning in both *Se* and *Hn* cells. We therefore set out to confirm that the observed phenotypes

were not a consequence of these mutations destabilizing McdB. We used circular dichroism

(CD) and size-exclusion chromatography with multi-angle light scattering (SEC-MALS) to

determine the secondary and quaternary structures of the purified proteins, respectively. As

shown previously [12, 18], wild type *Se* McdB forms a hexamer in solution with an α-helical

signature, whereas *Hn* McdB is monomeric and completely intrinsically disordered (Figure

4.5A-B). Since wild type *Hn* McdB is monomeric and disordered, there is no structure to disrupt

in McdB[W94G]. We therefore focused our analyses on McdB[ΔW152]. *Se* McdB[ΔW152]

remained hexameric and displayed the same α-helical signature as that of wild type *Se* McdB,

showing this mutation did not destabilize *Se* McdB structure or oligomerization *in vitro*.

Furthermore, we then confirmed that the tryptophan deletion did not destabilize *Se* McdB

structure *in vivo*, potentially leading to degradation. The average mNG fluorescence per cell was

quantified and compared between *Se* strains with wild type McdB and McdB[ΔW152]. No

decrease in the average mNG signal was observed in the McdB[ΔW152] strain, compared to that

of wild type (Figure 4.5C). Together, the results show that the tryptophan mutations have no significant effects on McdB stability either *in vivo* or *in vitro,* and indicate that the invariant tryptophan directly mediates α- and ß-McdB interactions with their respective carboxysomes.

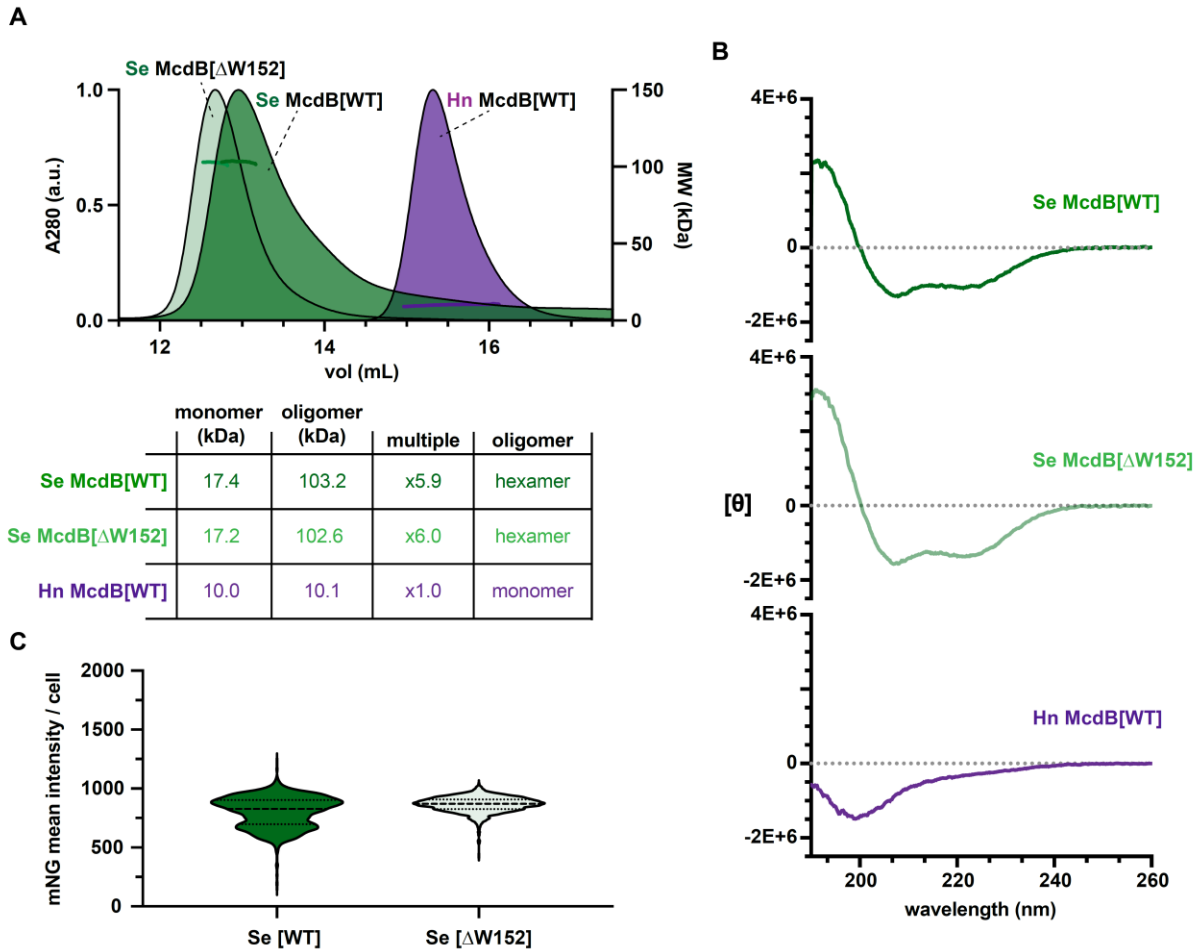**Figure 4.5**: **Mutations to the invariant Trp does not affect McdB protein stability *in vivo* or *in vitro*.** (**A**) SEC-MALS for the indicated McdB proteins, with a summary table below. WT = wild type. (**B**) CD spectra for the indicated McdB proteins. (**C**) Quantification of mNG-McdB intensities per cell for the indicated strains. Graphs represent the medians and interquartile ranges for n > 500 cells for each strain.

4.3.4 The C-terminus of monomeric Hn McdB is necessary and sufficient for carboxysome association, but not for hexameric Se McdB

We next set out to determine if the conserved C-terminal motifs we identified, which contain the invariant tryptophan, were sufficient to drive McdB association with carboxysomes. To investigate this, we N-terminally fused mNG to the last 31 amino acids of *Se* McdB and the last 10 amino acids of *Hn* McdB (Figure 4.6A). The choice of C-terminal domain (CTD) size was informed by our previous biochemical analysis of *Se* McdB, which revealed folded regions extending ~ 30 amino acids from the C-terminus [18]. We aimed to preserve this folding in the event that it was required for *Se* McdB association with carboxysomes. *Hn* McdB, on the other hand, is monomeric and completely disordered (Figure 4.5A-B). Therefore for the *Hn* McdB[CTD] construct, only the last 10 amino acids were included, comprising the highest identities from the α-McdB C-terminal alignments (see Figure 4.1D).

Intriguingly, the mNG-CTDs from *Se* and *Hn* McdB displayed different localizations. *Se* McdB[CTD] was completely diffuse in the cell (Figure 4.6B), and showed no association with carboxysomes, similar to that of McdB[ΔW152] (Figure 4.6B-C). *Hn* McdB[CTD], on the other hand, strongly colocalized with carboxysomes (Figure 4.6D), similar to that of wildtype* (Figure 4.6E). The data show that the last 10 amino acids of *Hn* McdB are necessary and sufficient for associating with α-carboxysomes. However, the last 31 amino acids of *Se* McdB, despite encoding the conserved motif and invariant tryptophan, is insufficient.

Recall that full-length *Se* McdB is a hexamer in solution whereas *Hn* McdB is monomeric (Figure 4.5A). Furthermore, we have previously shown that the CTD of *Se* McdB has an α-helical secondary structure [18], while *Hn* McdB is completely disordered [12]. It is
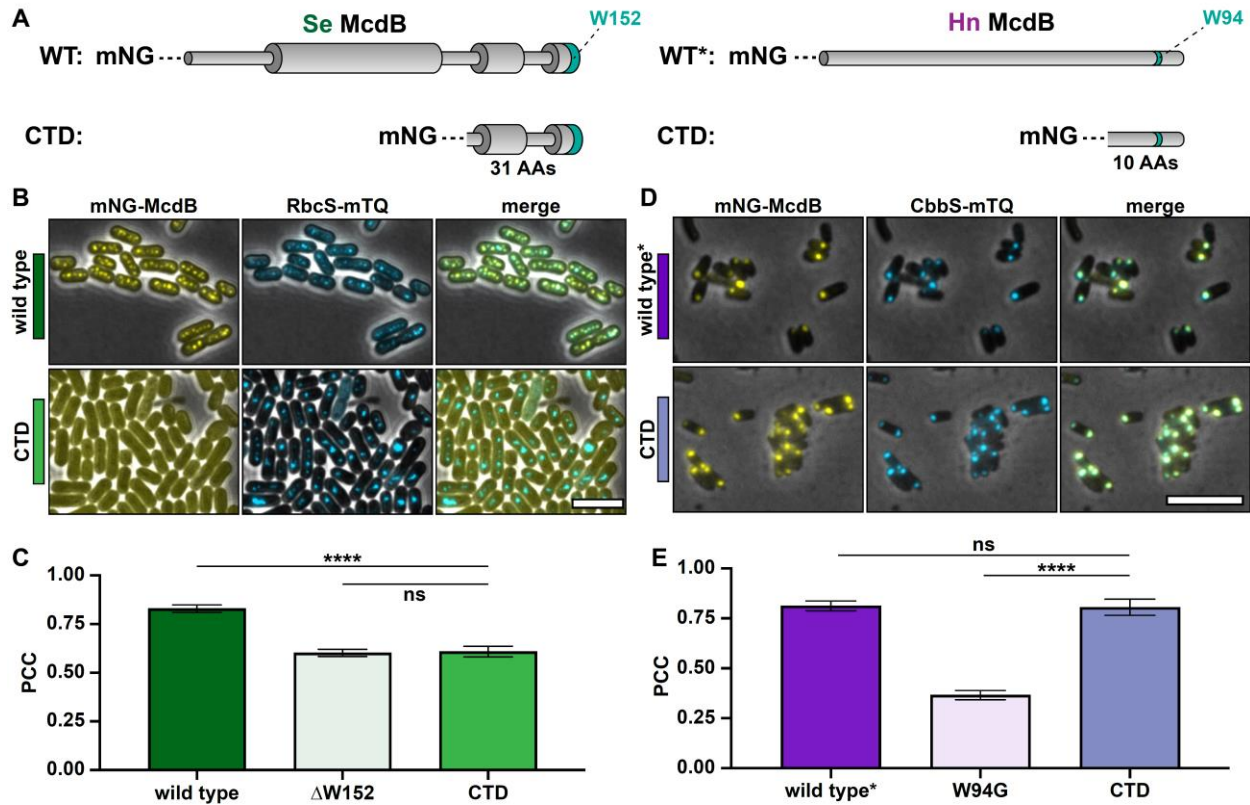
Figure 4.6: **The C-termini of α- and β-McdBs show differences in their ability to localize to carboxysomes.** (**A**) McdB protein models for *Se* and *Hn* wild type (WT) and C-terminal domains (CTD). Wide cylinders represent α-helical regions and narrow cylinders represent region of intrinsic disorder. The invariant tryptophan (W) is represented as a green stripe in the protein models. Sizes of the CTD truncations used are indicated below the respective model. (**B**) Representative microscopy images of the indicated *Se* strains. Phase contrast images are shown in black and white and overlaid with the fluorescence channels: mNG-McdB proteins are yellow and RbcS-mTQ labelled carboxysomes are cyan. Colored bars next to the strain names correspond to colors on the associated graphs. (**C**). Pearson's Correlation Coefficients (PCCs) quantified for the indicated *Se* strains. Graphs represent means and standard deviations from 7 technical replicates each with n > 500 cells. **** p < 0.001 and ns = non-significant from Welch's t-test. (**D**) Representative microscopy images of the indicated *Hn* strains. Wild type* indicates the wild type McdB with an N-terminal mNG tag, which causes carboxysome aggregation in *Hn*. Phase contrast images are overlaid with the fluorescence channels: mNG-McdB proteins are yellow and CbbS-mTQ labelled carboxysomes are cyan. Colored bars next to the strain names correspond to colors on the associated graphs. (**E**) Pearson's Correlation Coefficients (PCCs) quantified for the indicated *Se* strains. Graphs represent means and standard deviations from 7 technical replicates each with n > 500 cells. **** p < 0.001 and ns = non-significant from Welch's t-test. Scale bars are 5 µm and apply to all images.

therefore possible that the *Se* McdB[CTD] has altered protein structure and/or oligomerization that influences its association with carboxysomes. To investigate oligomerization, we performed SEC on *Se* McdB[CTD] (3.7 KDa), and used full-length *Hn* McdB (10 KDa) and an N-terminal peptide of *Se* McdB (*Se* McdB[NTD]; 2.3 kDa) as sizing standards. Both *Se* McdB[CTD] and *Se* McdB[NTD] eluted at the lower end of the separation range of the column (3 kDa) (Figure 4.7A), indicating that *Se* McdB[CTD] remains monomeric. CD analysis confirmed that *Se* McdB[CTD] retained an α-helical structure (Figure 4.7B). Full-length *Hn* McdB and *Se* McdB[NTD] are provided as disordered protein controls. The retained α-helical structure of *Se* McdB[CTD] explains why it eluted later than the disordered *Se* McdB[NTD] construct, despite having a higher molecular weight (Figure 4.7A).

Together, the results show that *Se* McdB[CTD] retains its α-helical structure, but does not form a hexamer like the full-length protein. We propose that the C-terminal association of *Se* McdB with carboxysomes requires higher avidity provided by hexamerization, whereas for the *Hn* McdB monomer, a single motif is necessary and sufficient for carboxysome association.
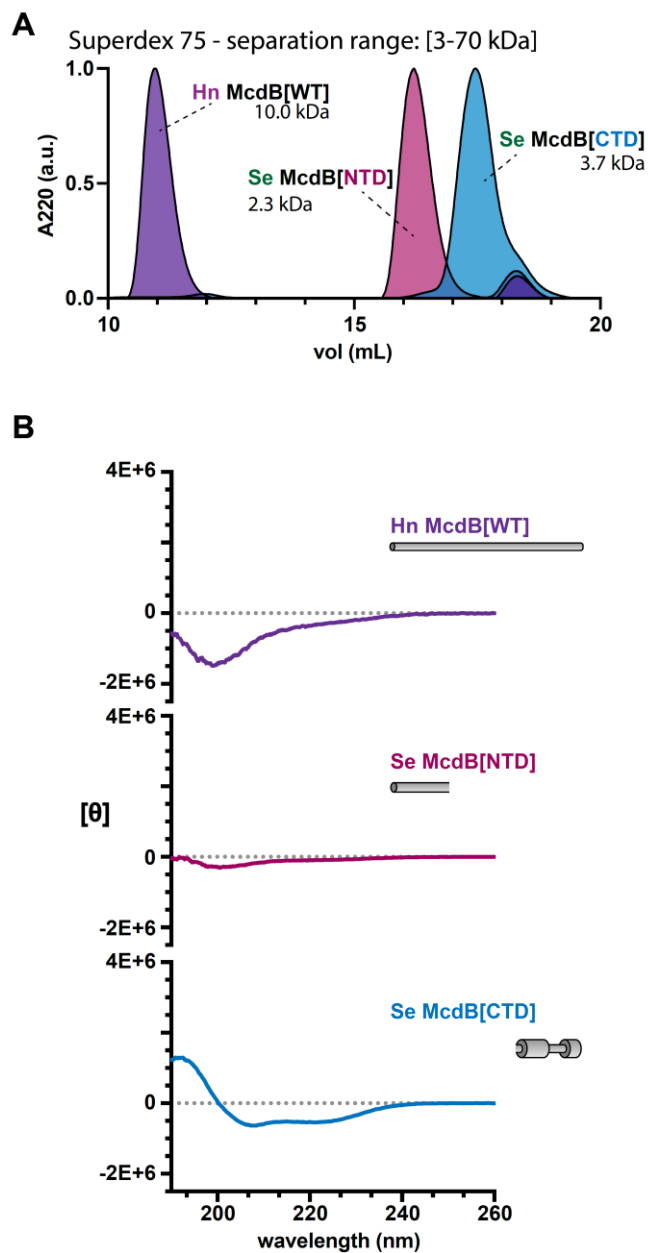
Figure 4.7: **The C-terminus of *Se* McdB alone does not oligomerize, but remains α-helical.** (**A**) SEC performed on the indicated column for the indicated protein variants. WT = wild type, NTD = N-terminal domain, CTD = C-terminal domain. Predicted monomeric weights are indicated for each construct. (**C**) CD spectra for the indicated proteins. Models of the constructs are shown, with wide cylinders representing α-helical regions and narrow cylinders representing region of intrinsic disorder.

4.3.5 Other aromatic residues functionally replace the invariant tryptophan with varying activity

Our previous bioinformatic analyses identified putative McdB-like proteins associated with other BMCs [11], including the 1,2-propanediol utilization microcompartment (PDU) and the

glycyl radical enzyme-containing microcompartment (GRM) [1]. Intriguingly, the C-termini of these McdB-like proteins lack the invariant tryptophan (Figure 4.8A). Instead, we found tyrosine (Y) or phenylalanine (F) residues conserved within the last five amino acids (Figure 4.8B), suggesting other C-terminal aromatics could potentially fulfill the role of the invariant tryptophan.
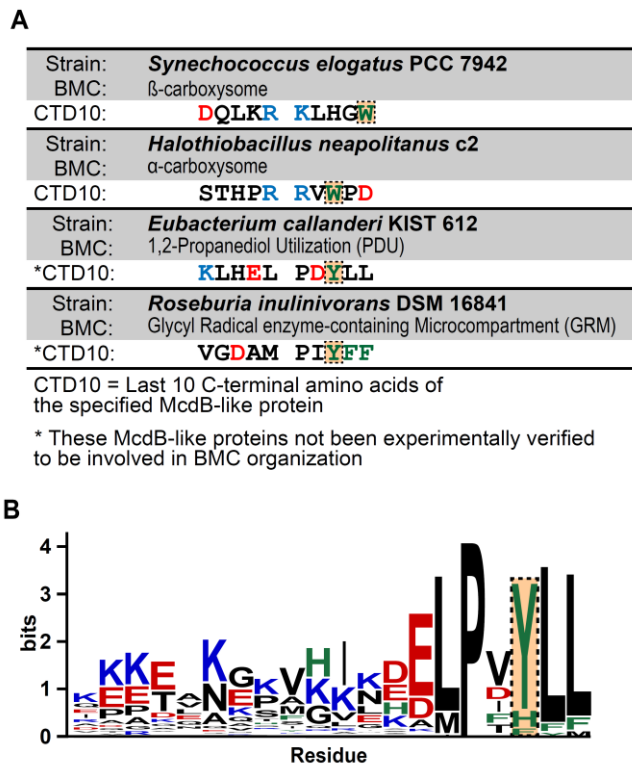
Figure 4.8: **McdB-like proteins found near the operons of non-carboxysome BMCs have aromatics other than tryptophan at their C-termini.** (**A**) Table displaying the last 10 C-terminal amino acids of McdBs from *Se* and *Hn*, as well as from putative McdBs from bacteria containing the indicated BMCs. Acidic residues are colored red, basic residues blue, and aromatic residues green. The residues at the position of the invariant tryptophan are boxed and highlighted. (**B**) Sequence logos generated from a multiple sequence alignments of the last 20 C-terminal amino acids of McdBs from non-carboxysome BMCs. The position corresponding to the invariant tryptophan as shown in (A) is boxed and highlighted.

To test this, we substituted the invariant tryptophan with Y or F in both the *Se* and *Hn* McdB proteins. These McdB variants were N-terminally fused to mNG, expressed at the native locus, and imaged in the carboxysome-labeled strains. We found that *Se* McdB[W152Y] colocalized with carboxysomes to the same degree as wild type McdB (Figure 4.9A-B). *Se* McdB[W152F] showed weaker association, but still greater than *Se* McdB[ΔW152]. Intriguingly, this gradient of carboxysome colocalization strongly correlated with the carboxysome positioning function of each *Se* McdB variant (Figure 4.9C, Figure 4.10A). *Se* McdB[W152Y] still distributed

carboxysomes, albeit with slightly perturbed spacing and higher foci intensities compared to that of wild type. *Se* McdB[W152F] showed an even lesser degree of carboxysome positioning compared to McdB[W152Y], but still greater than *Se* McdB[ΔW152].

In *Hn*, substituting the tryptophan with other aromatic residues was significantly less permissive. McdB[W94Y] only moderately colocalized with carboxysomes, whereas McdB[W94F] was completely diffuse in the cell, similar to that of McdB[W94G] (Figure 4.9E). And none of the McdB mutants were capable of positioning carboxysomes in *Hn* (Figure 4.9F, Figure 4.10B).

To summarize, we found a striking gradient of McdB variant colocalization with carboxysomes that followed similar trends: W152 ≈ W152Y > W152F > ΔW152 in *Se* (Figure 4.9B), and W94 > W94Y > W94F ≈ W94G in *Hn* (Figure 4.9E). Furthermore, in *Se*, the gradient of carboxysome association directly correlated with the carboxysome positioning function of the McdB mutants (Figure 4.9C, Figure 4.10A). In *Hn*, however, none of the McdB mutants restored carboxysome positioning (Figure 4.9F, Figure 4.10B).

Overall, in both *Se* and *Hn*, substituting the conserved tryptophan with tyrosine provided strong McdB localization to carboxysomes, compared to that of the phenylalanine substitution. In fact, *Se* McdB[W152Y] localized to carboxysomes to a comparable degree as wild type McdB and provided near-wild type carboxysome positioning function. This is striking given that all carboxysome-associated McdBs bioinformatically identified encode an invariant C-terminal W instead of Y, and W is generally considered the least substitutable amino acid [26]. We conclude that C-terminal aromatic residues can drive the localization of McdBs to both α- and ß-carboxysomes, and suggests a similar role for the conserved C-terminal aromatic residues found in putative McdB-like proteins associated with other BMC types.
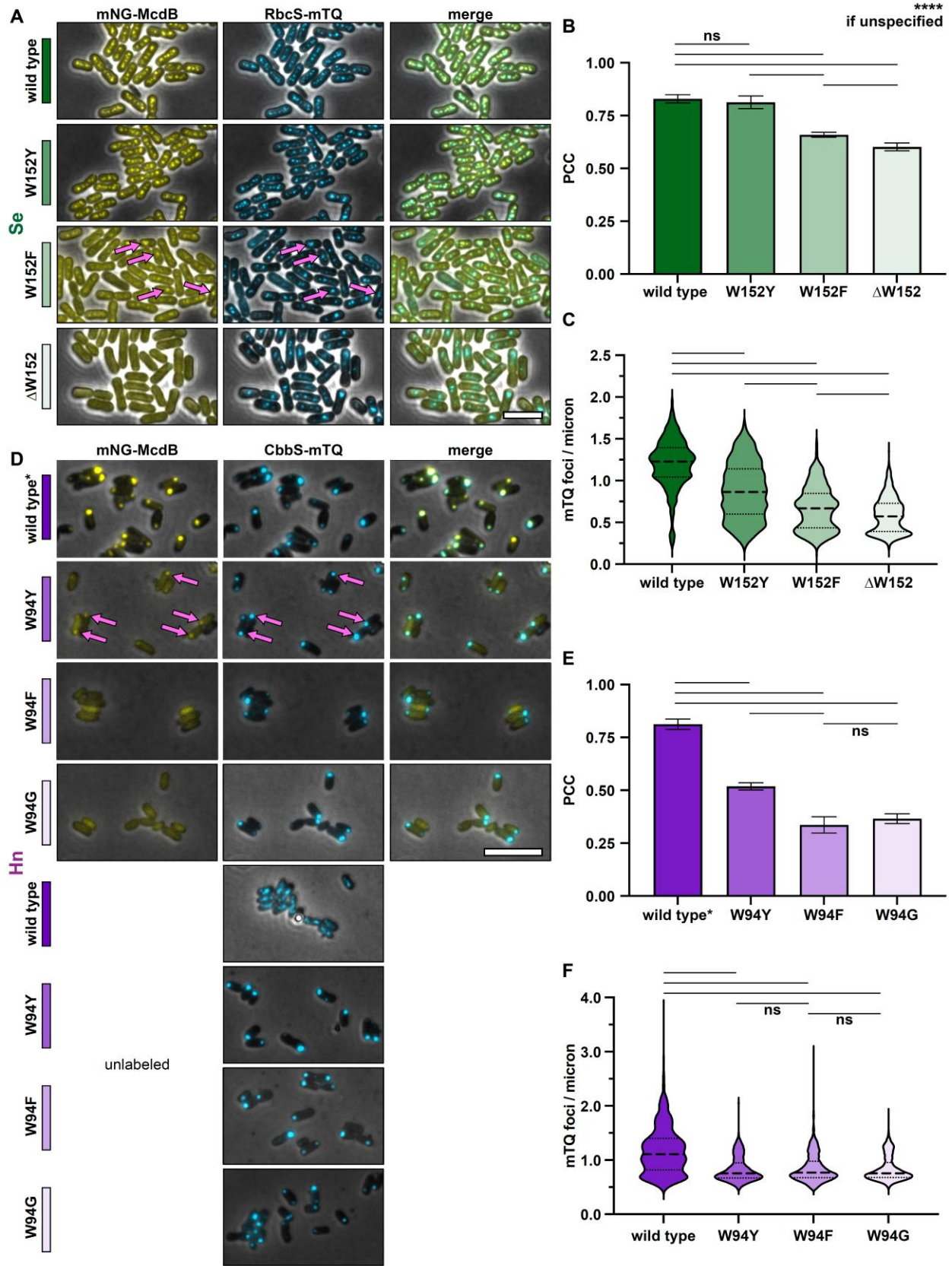
Figure 4.9 (*previous page*): Changing the invariant tryptophan to other aromatic residues reveals a gradient of McdB colocalization with carboxysomes. (A) Representative microscopy images of the indicated *Se* strains. Phase contrast images are shown in black and white and overlaid with the fluorescence channels: mNG-McdB proteins are yellow and RbcS-mTQ labelled carboxysomes are cyan. Magenta arrows highlight moderate McdB colocalization with carboxysomes. Colored bars next to the strain names correspond to colors on the associated graphs. (**B**). Pearson's Correlation Coefficients (PCCs) quantified for the indicated *Se* strains. Graphs represent means and standard deviations from 7 technical replicates each with n > 500 cells. **** $p < 0.001$ and ns = non-significant from Welch's t-test. (**C**) Quantification of carboxysome spacing as number of mTQ foci divided by cell length. Graphs represent medians and interquartile ranges from 3 biological replicates each with n > 500 cells. **** $p < 0.001$ from Mann-Whitney U-test. (**D-F**) As in (A-C), but in *Hn* strains. "Unlabeled" refers to the strain set with the indicated mutations in McdB, but McdB is not labeled with mNG. Wild type* indicates the wild type McdB with an N-terminal mNG tag, which causes carboxysome aggregation in *Hn*. Scale bars are 5 µm and apply to all

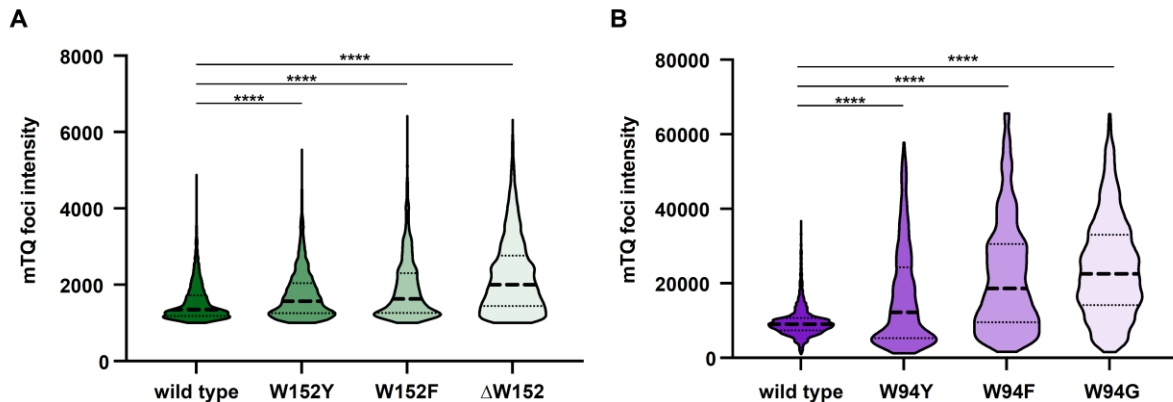**Figure 4.10: Changing the invariant tryptophan to other aromatic residues reveals a gradient of McdB function in positioning carboxysomes. (A)** Quantification of RbcS-mTQ intensities from carboxysome foci in the indicated strains of *Se*. Graphs represent medians and interquartile ranges from 3 biological replicates each with n > 1000 foci. **** $p < 0.001$ from Mann-Whitney U-test. (**B**) As in (A), but for *Hn* strains.

4.3.6 The invariant tryptophan influences McdB condensate formation

We previously found that several McdB proteins, including those from *Se* and *Hn*, can form biomolecular condensates *in vitro* [11, 12, 18, 20]. For *Se* McdB, positively charged residues within the disordered N-terminus mediate the degree to which McdB forms condensates - as positive charges were removed, the ability to form condensates decreased [21]. Our biochemical characterization also suggested these N-terminal positive residues may associate with the C-terminus of other McdB molecules to drive condensation. However, we did not identify C-terminal McdB mutants that altered condensation without destroying hexamerization.

Some protein condensates form via cation-π networks, where positively charged residues interact with electron-dense aromatic residues through electrons in their π orbitals [27, 28]. Interestingly, these studies have shown gradients of condensate formation by changing the type of aromatic residues involved [27]. We set out to determine if the invariant tryptophan influenced McdB condensate formation *in vitro*, and whether other aromatic residues at this position resulted in a gradient of condensate forming activity that could provide mechanistic insight into our observations *in vivo*.

We purified the *Se* McdB variants and compared the degree to which each formed condensates under conditions we previously found to facilitate condensate formation for wild type McdB [18]. The level of condensate formation for McdB[W152Y] was slightly lower than wild type, and lower still for McdB[W152F]. *Se* McdB[ΔW152] could form condensates, but with the lowest activity (Figure 4.11A). Intriguingly, this gradient of condensation activity (W152 > W152Y > W152F > ΔW152) directly correlates with the ability of these *Se* McdB variants to associate with carboxysomes and drive their positioning reactions *in vivo* (See Figure 4.9A-C).
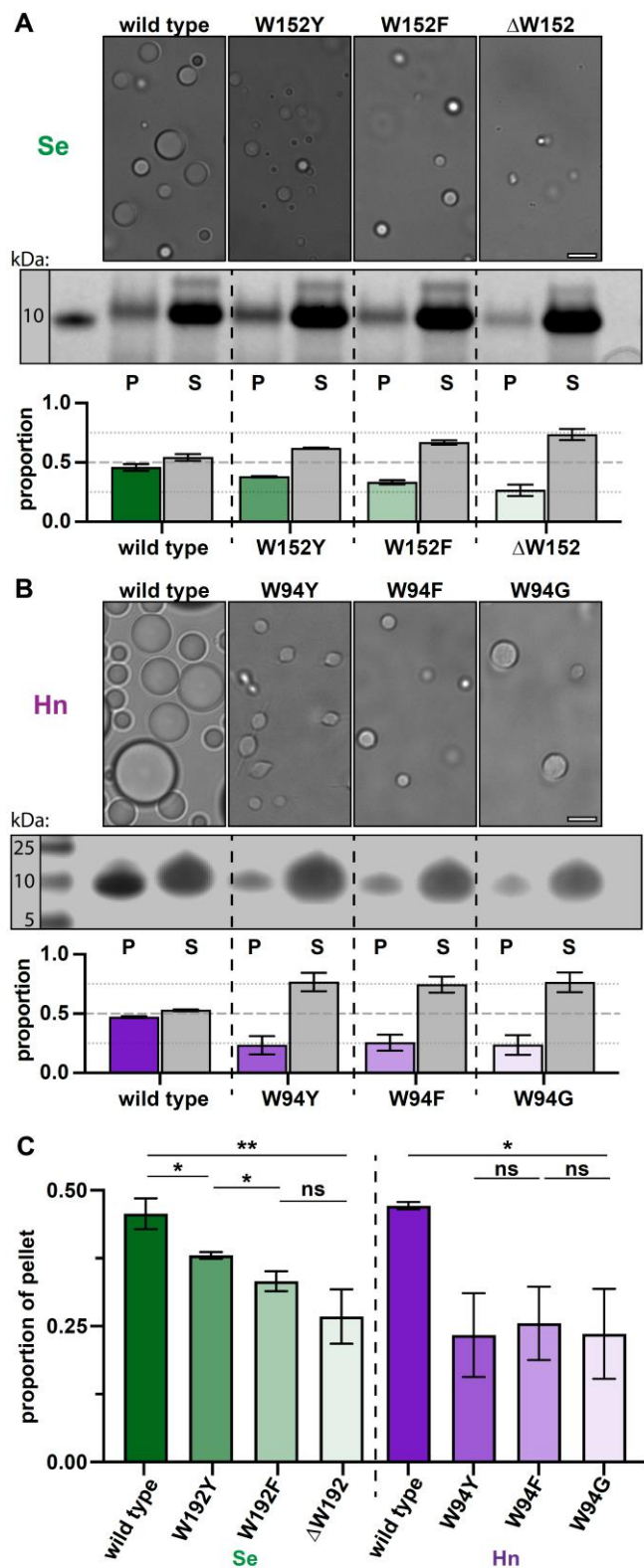
Figure 4.11: **Changing the invariant tryptophan to other aromatic residues reveals a gradient of condensate formation for *Se* McdB but not *Hn* McdB.** (**A**) (*top*) Representative DIC microscopy images of the indicated *Se* McdB variants at 50 µM (100 mM KCl; 20 mM HEPES, pH 7.2) after 1 hour. (*middle*) Samples were pelleted (P) and run on SDS-PAGE gel along with associated supernatant (S). (*bottom*) Gel bands were quantified. Graphs represent the proportion of total intensity from a P / S pair, and are reported as the mean and standard deviation from 3 replicates. (**B**) As in (A), but for *Hn* McdB at 700 µM (100 mM KCl; 20 mM HEPES, pH 7.2; 15% PEG-8000) after 18 hours. (**C**) Quantification summary of pellet fractions from *Se*- and *Hn*-McdB variants. ** $p < 0.01$, * $p < 0.05$, ns = non-significant from Welch's t-test. Scale bars are 5 µm and apply to all microscopy images.

When performing the same comparison with the purified McdB variants from *Hn*, we did not observe a gradient of condensate formation as we did with the *Se* proteins. Instead, all *Hn* McdB mutants showed the same significant loss of condensate formation (Figure 4.11B-C), which correlates with the inability of these *Hn* McdB variants to associate with and position carboxysomes *in vivo* (See Figure 4.9D-F).

Finally, to investigate the condensate forming activity and relative expression levels of McdB variants *in vivo*, we expressed mCherry-tagged McdB constructs in *E. coli* and monitored the formation of foci as well as expression levels over time. Using this approach, we have recently shown that the fluorescent foci formed by *Se* mCherry-McdB in *E.coli* cells (Figure 4.12A) are liquid-like condensates [18, 29]. Intriguingly, the tryptophan mutants once again displayed the same functional gradient in forming condensates *in vivo* (Figure 4.12B). Importantly, all *Se* McdB variants showed the same levels of expression compared to wildtype, and with no notable degradation (Figure 4.12C). The data provide an additional line of evidence showing that mutation to the invariant tryptophan did not result in the destabilization and/or degradation of McdB variants. Together, we once again find a gradient of condensation activity *in vivo* that mirrors our *in vitro* results with *Se* McdB (W152 > W152Y > W152F > ΔW152).

Hn McdB remained soluble in *E. coli*, even at the highest expression levels achievable in this assay (Figure 4.12D). This is consistent with *Hn* McdB requiring significantly higher protein concentrations to form condensates *in vitro*, compared to *Se* McdB (see Figure 4.11A-B). Since we could not form foci with wildtype *Hn* McdB, mutant versions of *Hn* McdB were not pursued using this approach.

Overall, the data implicate the invariant C-terminal tryptophan as a major contributor to

condensate formation for both α- and β-McdBs. Furthermore, aromatic residue substitutions at

the tryptophan position can affect McdB condensate formation *in vitro* in a manner that directly

correlates with how aromatics affect McdB function *in vivo*.
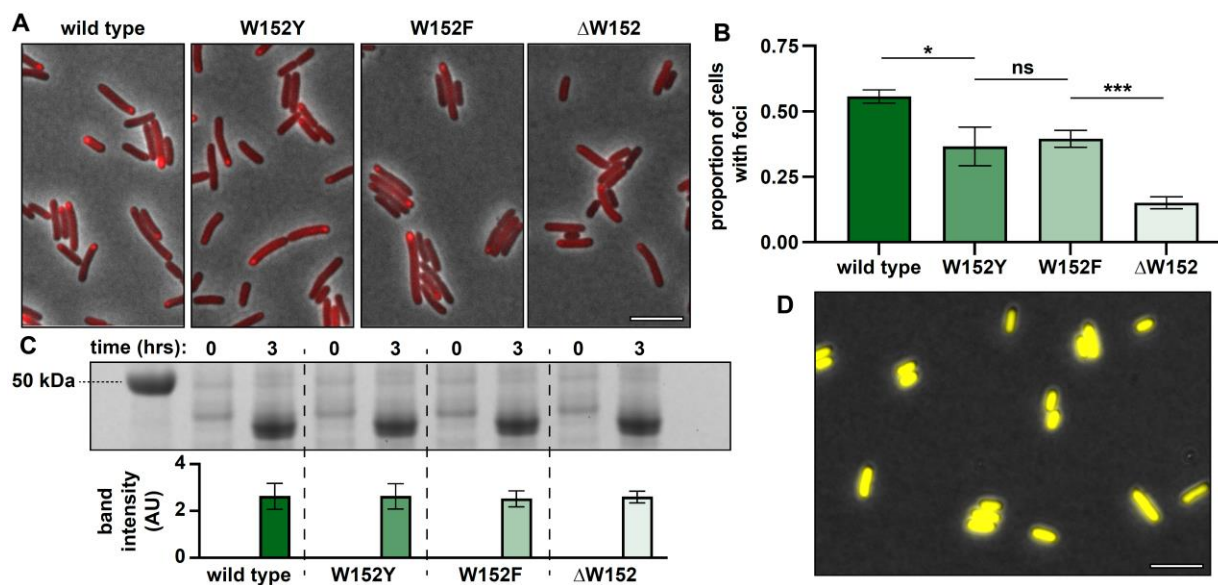
Figure 4.12: **Differences in the solubilities of *Se* McdB aromatic substitutions in *E. coli*.**
(**A**) Representative microscopy image of the indicated *Se* mCherry-McdB variant after 3
hours of expression in *E. coli* MG1655. Phase contrast images are shown in black and white
and overlaid with the fluorescence channel for mCherry-McdB as red. (**B**) Quantification of
the proportion of cells from (A) with foci. Graphs represent means and standard deviations
from 3 technical replicates each with n > 750 cells. *** $p < 0.001$, * $p < 0.05$, ns = non-
significant from Welch's t-test. (**C**) (*top*) SDS-PAGE gel from the experiment shown in (A).
Cells were harvested at the indicated times of expression and standardized by OD600 prior to
running on the gel. Expected size of mCherry-McdB constructs from *Se* is roughly 45 kDa.
(*bottom*) Quantification of the normalized band intensities from the above gel. Graph
represents means and standard deviations from 3 technical replicates. Comparisons of all
variants at the 3-hour time point were non-significant from Welch's t test. (**D**) Representative
microscopy image of wild type *Hn* mNG-McdB after 3 hours of expression in *E. coli* BL21.
Phase contrast images are shown in black and white and overlaid with the fluorescence
channel for mNG-McdB as yellow. Note that the protein remained completely soluble, and so
mutant variants of *Hn* McdB were excluded from further analysis. Scale bars are 5 µm and
apply to all images.

4.4 Discussion

Here, we identified and probed the function of conserved C-terminal motifs containing a tryptophan residue that is invariant across all McdB proteins bioinformatically identified to date. The invariance is striking because even when comparing small regions of McdB sequences of fixed length, this tryptophan is the only residue with 100% identity (Figure 4.1). We found that the invariant tryptophan is necessary for α- and ß-McdBs to colocalize with α- and ß-carboxysomes, respectively. With the invariant tryptophan removed, McdB became diffuse in the cytoplasm (Figure 4.3). Interestingly, the C-terminal motif containing the invariant tryptophan was necessary and sufficient for carboxysome localization by α-McdB of *Hn*, but not for ß-McdB of *Se* (Figure 4.6). We suggest this discrepancy may be due to differences in the minimal oligomeric unit of the proteins, whereby full-length *Hn* McdB and the C-terminal fragment are both monomers, while full-length *Se* McdB is a hexamer and its C-terminal fragment is monomeric (Figure 4.7).

We also found that putative McdB-like proteins that are associated with other BMCs have conserved C-terminal aromatic residues other than tryptophan (Figure 4.8). We therefore attempted to complement the removal of the invariant tryptophan in both α- and ß-McdBs by substituting it with other aromatic amino acids. Intriguingly, we observed a gradient of carboxysome association that correlated with the carboxysome positioning function of each McdB variant (Figure 4.9). Lastly, this gradient of function *in vivo* directly correlated with the ability of these McdB variants to form condensates both *in vitro* (Figure 4.11) and *in vivo* (Figure 4.12D). A summary of our findings is provided in Figure 4.13. Together, the data provide a foundation for future studies on the molecular nature of McdB-carboxysome interactions and the role protein condensation may play in this association.

S. elongatus

Localization and Oligomerization of CTD

Se McdB[WT] hexamer

Se McdB[CTD] monomer

Carboxysome Localization

Condensate Formation

Trp

Tyr

Phe

Gly / Δ

H. neapolitanus

Condensate Formation

Carboxysome Localization

Localization and Oligomerization of CTD

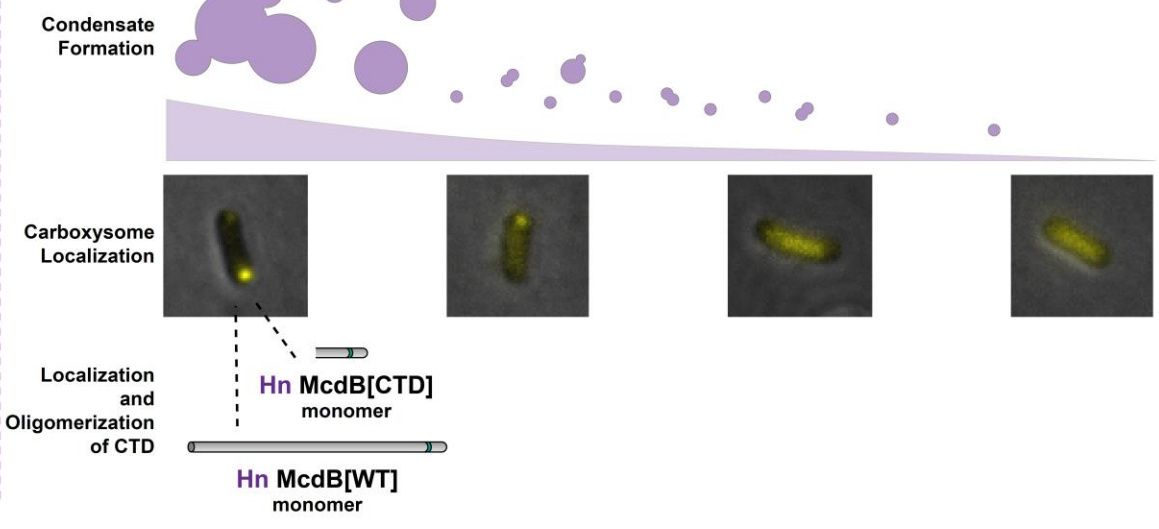Hn McdB[CTD] monomer

Hn McdB[WT] monomer
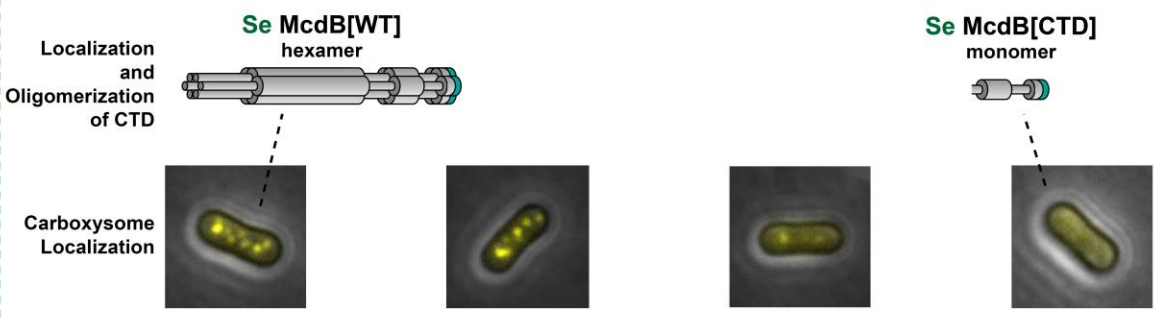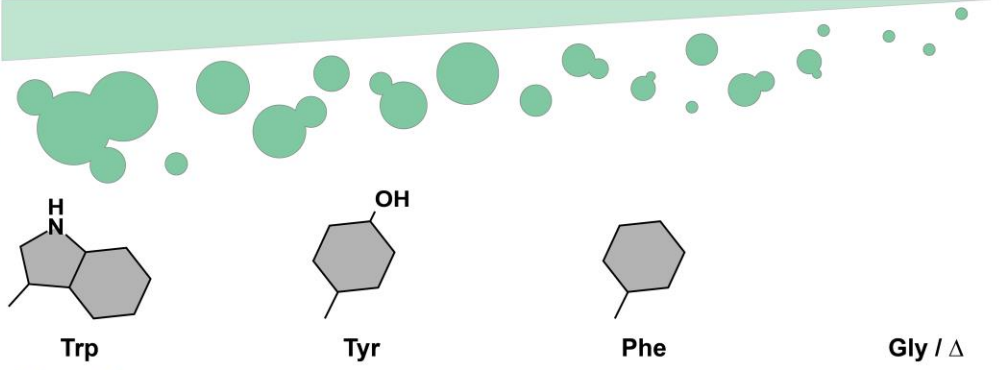
Figure 4.13 (*previous page*): **Summary of condensate formation and carboxysome localization for the aromatic substitution mutants of *Se* (*top*) and *Hn* (*bottom*) McdB.** Wild type (WT) McdB proteins from *Se* and *Hn* have an invariant C-terminal tryptophan (trp), depicted as a green stripe in the cartoon protein models. *Se* McdB functions as a hexamer whereas *Hn* McdB is monomeric. We substituted this trp with tyrosine (tyr), phenylalanine (phe), glycine (gly) as well as deleted the trp (Δ) (*center*), which revealed a gradient of condensate formation activity and carboxysome localization (ramps). The C-terminal domain (CTD) containing the invariant tryptophan was sufficient to localize *Hn* McdB to carboxysomes (*bottom*). However this was not the case for *Se* McdB[CTD] suggesting oligomerization of *Se* McdB is also a requirement (*top*).

4.4.1 Comparative analyses on BMC shell proteins could further our understanding of their molecular interactions with McdB proteins

It remains to be determined how, at a molecular level, the invariant tryptophan drives McdB association with carboxysomes. Carboxysomes, like all BMCs, are comprised of a selectively permeable protein shell and an enzymatic core [1]. While the set of encapsulated core enzymes are highly diverse, the outer shell proteins of BMCs are well-conserved in sequence and structure [1, 2]. For all BMCs, several different types of protein oligomers build the outer shell to give rise to the characteristic polyhedral shape of a BMC [1]; the most abundant of which is a hexamer (BMC-H). For β-carboxysomes, the major hexameric shell protein is called CcmK2. In *Se*, we have shown that McdB strongly associates with CcmK2 [10]. However, the regions and residues of CcmK2 required for the McdB-CcmK2 association remain to be determined.

Although BMC-H proteins have regions of high conservation, there are also variable regions that have led to a diversity of BMC-H subtypes [2, 30]. Case in point, the BMC-H shell proteins from α- and ß-carboxysomes (CsoS1A and CcmK2, respectively) are structurally and phyletically distinct from one another, forming distant clades on a phylogenetic tree of all BMC-H protein sequences [2, 30]. Therefore α- and ß-carboxysomes have distinct evolutionary

histories, but converged on a functionally homologous BMC type [8, 9, 30]. It is intriguing that α- and ß-McdB have seemingly also converged onto a similar mechanism of association with their respective carboxysomes; both mediated by a C-terminal invariant tryptophan.

Going forward, we will leverage this knowledge to examine conserved regions of CsoS1A and CcmK2 to identify co-occurring surface-exposed regions as candidate sites for interaction with the C-terminal motifs of α- and ß-McdB. Future studies such as this will help deepen our understanding of McdB localization to carboxysomes, and protein localization to BMCs in general.

4.4.2 Tryptophan mediates the assembly of several viral and phage capsids

BMC shells share several analogous features to viral capsids [31]. For instance, both are primarily comprised of hexameric proteins and some pentamers, the combination of which results in the characteristic polyhedral shape of BMCs and capsids [1, 32]. Although structurally similar, BMC shells and viral capsids likely evolved independently, representing multiple convergent events [31]. It is therefore insightful to compare these analogous structures and identify general features involved in self-assembly.

Intriguingly, many capsid proteins encode C-terminal tryptophan residues that are essential for the assembly of viral particles [33, 34, 35]. Similarly, several carboxysome BMC-H proteins, including *Se* CcmK2, contain tryptophan residues at the interface of shell proteins that are often oriented toward the outer facet of the shell [36, 37]. Albeit, none have been experimentally verified to be involved in assembly. However, the fact that tryptophan mediates interactions among viral capsid proteins and is found at the interface of BMC-H proteins suggests that tryptophan may be critical for mediating protein-protein interactions in these contexts. Whether

McdB proteins interact with carboxysome shells via π–π stacking of tryptophan residues is an attractive mechanism of association for future study. Importantly, several groups are working to purify minimalized α- and ß-carboxysome shells to be used as tools in synthetic biology and biotechnological applications [38, 39]. These minimalized shells could also serve as useful *in vitro* tools to reconstitute and study the molecular nature of McdB interactions with carboxysome components, and BMC-H proteins in particular.

4.4.3 Kinesin-1 recognizes a tryptophan-acidic motif to interact with protein-based cargos

In eukaryotic cells, the motor protein kinesin-1 is critical for transporting diverse protein-based cargos on microtubules [40]. How kinesin-1 recognizes and binds to a diverse set of cargo proteins to facilitate their transport has been under investigation for decades [41]. It is now understood that these different cargo proteins all contain tryptophan-acidic motifs (such as EWD) that facilitate their binding to positively charged pockets on kinesin-1 [42].

Analogous to kinesin-1, McdB-like proteins must bind to diverse protein-based BMCs to facilitate their spatial regulation. Intriguingly, we show here that both α- and ß-McdBs tend to have acidic residues (often D) within 5 amino acids of the conserved tryptophan (see Figure 4.1). This is also true for the putative McdB-like proteins we identified for other BMCs (see Figure 4.8), although these contain aromatics residues other than tryptophan. It is therefore attractive to speculate that McdBs follow an analogous mechanism to bind BMCs as does kinesin-1 to its cargos, using tryptophan-acidic motifs to bind positively charged pockets on BMC shells. Consistently, the carboxysome is known to have positively charged pockets within the pores of different shell proteins [43]. Future investigations will therefore focus on the involvement of the

conserved acidic residues in the C-termini of McdBs as well as the surface-exposed positive residues on BMC shells that could also mediate this association.

4.4.4 The role of protein condensates in ParA/MinD-based positioning systems

McdB functions as an adaptor protein for the carboxysome positioning ATPase, McdA, which is a member of the ParA/MinD-family of positioning ATPases [13, 14, 19]. This family is widespread in bacteria, and is responsible for spatially regulating a variety of genetic- and protein-based cargos [13, 14]. The ATPases themselves do not interact directly with the cargo, but instead rely on adaptor proteins that either interact with, or are essential components of, the positioned cargos [15]. How these adaptor proteins confer specificity to a wide variety of disparate cargos, ranging from the chromosome, divisome, flagella, chemotaxis arrays, and a diversity of BMCs, has been an active area of research for decades [13, 14, 15, 44].

Recently, the ability of some of these adaptor proteins to form condensates has been proposed to influence their localizations and functionality inside cells. For instance, the adaptor protein ParB which localizes to DNA molecules to aid in their segregation, has been shown to form condensates *in vitro* [45]. ParB foci in the cell have also been seen to exhibit liquid-like behaviors [46], suggesting a dynamic, condensate-like mechanism underlying their formation and maintenance. Similar evidence exists for the co-complex of adaptor proteins, PomX and PomY, which localize cell division to mid-cell in some bacteria [47]. The protein PomY forms liquid-like structures *in vivo* that, when perturbed, produce defects in cell division [48]. Here we show a correlation between McdB localization to its carboxysome cargo, the ability of McdB to distribute carboxysomes, and its ability to form condensates both *in vitro* and *in vivo*. Whether

the decrease in functionality we observed for our McdB variants is a direct consequence of the corresponding decrease in condensation activity will be an exciting area of future investigation.

4.5 Acknowledgements

4.6 Materials and Methods

4.6.1 Multiple sequence alignments

McdB amino acid sequences were obtained from our gene neighborhood analyses previously described for both α- [12] and ß- McdBs [11]. Multiple sequence alignments (MSAs) were performed using Clustal Omega [49] and were exported and viewed using Geneious Prime (v 2020.02.02). Identity graphs were generated in Geneious Prime, and represent the percentage of pairwise residues that are identical in the alignment, including gap versus non-gap residues but excluding gap versus gap residues. Sequence logos were created using the above mentioned MSAs via WebLogo (v 2.8.2) [50].

4.6.2 Construct design

All constructs used in this study were generated using Gibson Assembly [51]. Cloning of plasmids was performed in chemically competent *E. coli* Top10 cells (Takara Bio). To replace native McdB with mutant variants in both *Se* and *Hn*, homology regions of 750 bp from both upstream and downstream of the native *mcdB* loci were added to the flanking regions of the generated constructs [52]. Fluorescent fusions to proteins of interest were added to the indicated termini with a GSGSGS linker between the two proteins.

4.6.3 Growth and transformation of Se strains

All *Se* strains were grown in BG-11 media (Sigma) buffered with 1 g/L HEPES, pH 8.3. Cultures were grown in a Minitron incubation system (Infors-HT) with 60 µmol m^{-2} s^{-1} continuous LED 5600 K light, 32°C, 2% CO_2, and shaking at 130 RPM. Cells were transformed using 250-1000 ng of total plasmid DNA added to 300 µL of culture at $OD_{750} = 0.7$, and incubated in the dark for 16-24 hrs [52]. Transformations were then plated on BG-11 media plus agar with the addition of 12.5 µg/mL kanamycin or 12.5 µg/mL chloramphenicol. Single colonies were picked and grown in BG-11 liquid media containing the same antibiotic concentrations, verified for full insertion via colony PCR, and then removed from antibiotics.

4.6.4 Growth and transformation of Hn strains

All *Hn* strains were grown in ATCC® Medium 290: S6 medium for Thiobacilli [53]. Cultures were grown in a Minitron incubation system (Infors-HT) at 30°C, 5% CO_2, and shaking at 130 RPM. Competent *Hn* cells were generated by growing 1 L of log culture in 2.8 L flasks,

which were harvested by centrifugation at 3,000g for 45 min. Cell pellets were washed twice with 0.5 volumes of ice-cold water, and finally resuspended in 1 mL of ice-cold water. Competent cells were mixed with 250-1000 ng of total plasmid DNA and incubated on ice for 5 min. This mixture was then transferred to 5 mL of ice-cold S6 medium and incubated on ice for 5 min. Cells were then incubated for 16–24 hrs at 30°C, 5% CO_2, and 130 RPM. Transformations were then plated on S6 media plus agar with 50 µg/mL kanamycin or 25 µg/mL chloramphenicol. Single colonies were picked and grown in S6 liquid media containing the same antibiotic concentrations, verified for full insertion via colony PCR, and then removed from antibiotics.

4.6.5 Live cell fluorescence microscopy

For both *Se* and *Hn* cells, early log phase cultures grown in the absence of antibiotics were used for imaging. Two microliters of culture were spotted onto a 2 cm x 2 cm pad containing 1.5% UltraPure agarose (Invitrogen) + either BG-11 (for *Se*) or S6 media (for *Hn*). Cells were then imaged on a 35-mm glass-bottom dish (MatTek Life Sciences). All fluorescence and phase-contrast imaging was performed using a Nikon Ti2-E motorized inverted microscope controlled by NIS Elements software with a SOLA 365 LED light source, a ×100 objective lens (Oil CFI Plan Apochromat DM Lambda Series for Phase Contrast), and a Hamamatsu Orca-Flash 4.0 LTS camera. mNG constructs were imaged using a 'YFP' filter set (C-FL YFP, Hard Coat, High Signal-to-Noise, Zero Shift, excitation: 500/20 nm [490–510 nm], emission: 535/30 nm [520–550 nm], dichroic mirror: 515 nm). mTQ constructs were imaged using a 'CFP' filter set (C-FL CFP, Hard Coat, High Signal-to-Noise, Zero Shift, excitation: 436/20 nm [426–446 nm], emission: 480/40 nm [460–500 nm], dichroic mirror: 455 nm).

4.6.6 Image quantification using MicrobeJ

Image analysis including cell segmentation, quantification of foci spacing, and foci and cell intensities were performed using Fiji plugin MicrobeJ 5.13n [54, 55]. Cell perimeter detection and segmentation were done using the rod-shaped descriptor with default threshold settings at a tolerance of 55 for both *Hn* and *Se* cells. Carboxysome foci were detected from both *Se* and *Hn* using maxima detection set to point detection with a tolerance of 1000 and the sharpen image filter selected. PCCs were calculated using ImageJ plugin JaCoP [56]. Data were exported, further tabulated, graphed, and analyzed using GraphPad Prism 9.0.1 for macOS (GraphPad Software, San Diego, CA, https://www.graphpad.com).

4.6.7 Protein expression and purification

Wild type and mutant variants for both *Se* and *Hn* McdB were expressed with an N-terminal His-SUMO tag off a pET11b vector in *E. coli* BL21-AI (Invitrogen). All cells were grown in LB + carbenicillin (100 µg/mL) at 37°C. One liter cultures used for expression were inoculated using overnight cultures at a 1:100 dilution. Cultures were grown to an OD_{600} of 0.5 and expression was induced using final concentrations of IPTG at 1 mM and L-arabinose at 0.2%. Cultures were grown for an additional 4 hours, pelleted, and stored at -80°C.

Pellets were resuspended in 30 mL lysis buffer [300 mM KCl; 50 mM Tris-HCl pH 8.4; 5 mM BME; 10% glycerol; 50 mg lysozyme (Thermo-Fischer); protease inhibitor tablet (Thermo-Fischer)] and sonicated with cycles of 10 seconds on, 20 seconds off at 50% power for 7 minutes. Lysates were clarified via centrifugation at 15,000 rcf for 30 minutes. Clarified lysates

were passed through a 0.45 µm filter and loaded onto a 1 mL HisTrap HP (Cytiva) equilibrated in buffer A [300 mM KCl; 50 mM Tris-HCl pH 8.4; 5 mM BME; 10% glycerol]. Columns were washed with 5 column volumes of 5% buffer B [300 mM KCl; 20 mM Tris-HCl pH 8.4; 5 mM BME; 500 mM imidazole; 10% gylcerol]. Elution was performed using a 5-100% gradient of buffer B via an AKTA Pure system (Cytiva). Peak fractions were pooled and diluted with buffer A to a final imidazole concentration of < 100 mM. Ulp1 protease was added at 1:100 protease:sample, and incubated overnight at 25°C with gentle rocking. The pH was then adjusted to ~10 and samples were concentrated to a volume of < 5 mL, passed through a 0.45 µm filter and passed over a sizing column (HiLoad 16/600 Superdex 200 pg; Cytiva) equilibrated in buffer C [300 mM KCl; 20 mM CAPS pH 10.2; 5 mM BME; 10% glycerol]. Peak fractions were pooled, concentrated, and stored at -80°C.

4.6.8 SEC coupled to multi-angled light scattering (SEC-MALS)

For each sample analyzed, 50 µL at 1.5 mg/ml was passed over an SEC column (PROTEIN KW-804; Shodex) at a flow rate of 0.4 mL/min in buffer [150 mM KCl and 20 mM Tris-HCl, pH 8.2]. Following SEC, the samples were analyzed using an A$_{280}$ UV detector (AKTA pure; Cytiva), the DAWN HELEOS-II MALS detector with an internal QELs (Wyatt Technology), and the Optilab T-rEX refractive index detector (Wyatt Technology). The data were analyzed to calculate mass using ASTRA 6 software (Wyatt Technology). Bovine serum albumin was used as the standard for calibration.

4.6.9 Circular dichroism

For all protein samples analyzed, far-UV CD spectra were obtained using a J-1500 CD

spectrometer (Jasco). All measurements were taken with 250 µL of protein at 0.25 mg/mL in 20

mM KPi, pH 8.0. Measurements were taken using a quartz cell with a path length of 0.1 cm. The

spectra were acquired from 260 to 190 nm with a 0.1 nm interval, 50 nm/min scan speed, and at

25ºC.

4.6.10 Size-exclusion chromatography (SEC)

SEC was performed on full-length and truncated McdB proteins using a Superdex 75

Increase 10/300 GL (Cytiva) column connected to an AKTA pure system (Cytiva). 500 µL of

sample at 1.5 mg/mL was passed through the column at 0.4 mL/min in buffer [150 mM KCl; 20

mM Tris-HCl pH 8.2] while monitoring absorbance at 220 nm.

4.6.11 Microscopy of protein condensates

Samples for imaging were set up in 16 well CultureWells (Grace BioLabs). Wells were

passivated by overnight incubation in 5% (w/v) Pluronic acid (Thermo-Fischer), and washed

thoroughly with the corresponding buffer prior to use. All *Se* McdB samples were incubated for

at least 30 minutes prior to imaging condensates, and all *Hn* McdB samples for at least 18 hours

unless otherwise stated. Imaging of condensates was performed using a Nikon Ti2-E motorized

inverted microscope (60 × DIC objective and DIC analyzer cube) controlled by NIS Elements

software with a Transmitted LED Lamp house and a Photometrics Prime 95B Back-illuminated

sCMOS Camera. Image analysis was performed using Fiji v 1.0.

4.6.12 Quantification of phase separation via centrifugation

Centrifugation was used to quantify the degree to which McdB and its variants condensed under certain conditions, as previously described [57]. Briefly, 50 µL of sample was incubated at the conditions specified for 30 minutes, and then centrifuged at 16,000g for 10 minutes at 20°C. The supernatant was removed and the pellet resuspended in an equal volume of McdB solubilization buffer [300 mM KCl, 20 mM CAPS pH 10.2]. Samples were then diluted into 4X Laemmli SDS-PAGE sample buffer. Pellet and supernatant fractions were visualized on a 4–12% Bis-Tris NuPAGE gel (Invitrogen) by staining with InstantBlue Coomassie Stain (Abcam) for 1 hour and then destaining in water for 14-16 hours. The intensities of the bands were quantified using Fiji v 1.0 and resultant data graphed using GraphPad Prism 9.0.1 for macOS (GraphPad Software, San Diego, CA, www.graphpad.com).

4.6.13 Expression of proteins to quantify condensate formation in E. coli

All *Se* McdB constructs were expressed as N-terminal mCherry fusions [58] on plasmids regulated by the pTrc promoter in *E.coli* MG1665. Overnight cultures were grown in 5 mL LB media + carbenicillin (100 µg/mL). The overnight culture was then diluted 1:50 into 5 mL AB Media + carbenicillin (100 µg/mL) supplemented with (0.2% glycerol; 10 µg/mL thiamine; 0.2% casein; 25 µg/mL uracil). *Hn* McdB was expressed as N-terminal mNG fusions [58] on plasmids regulated by the T7 promoter in *E.coli* BL21. Overnight cultures were grown in 5 mL LB media + carbenicillin (100 µg/mL). The overnight culture was then diluted 1:50 into 5 mL LB Media + carbenicillin (100 µg/mL). All cultures were allowed to grow at 37°C for until OD = 0.2-0.6 and

then induced with 500 µM IPTG for *Se* strains, and 5 mM IPTG for *Hn*. The cultures continued

to grow post-incubation for 3 hours before imaging. Cells used for imaging were prepared by

spotting 2 µL of cells on to a 2% UltraPure agarose + AB medium pad on a Mantek dish. Images

were taken using Nikon Ti2-E motorized inverted microscope controlled by NIS Elements

software with a SOLA LED light source, a 100X Objective lens (Oil CFI Plan Apochromat DM

Lambda Series for Phase Contrast), and a Hamamatsu Orca Flash 4.0 LT + sCMOS camera.

mCherry signal was imaged using a "TexasRed" filter set (C-FL Texas Red, Hard Coat, High

Signal-to-Noise, Zero Shift, Excitation:560/40 nm [540-580 nm], Emission: 630/75 nm [593-668

nm], Dichroic Mirror: 585 nm). For monitoring expression levels, cells were harvested either at

the time of induction (t = 0 hours) or at the time of imaging (t = 3 hours). Cell lysates were

normalized based on OD600, and were visualized via SDS-PAGE. Intensity of the bands

corresponding to the McdB fusions were normalized to the background cell lysate and

quantified. Image analysis was performed using Fiji v 1.0.

4.7 References

1. C. A. Kerfeld, C. Aussignargues, J. Zarzycki, F. Cai, M. Sutter, Bacterial microcompartments. *Nat. Rev. Microbiol.* **16**, 277–290 (2018).
2. M. Sutter, M. R. Melnicki, F. Schulz, T. Woyke, C. A. Kerfeld, A catalog of the diversity and ubiquity of bacterial microcompartments. *Nat. Commun.* **12**, 1–12 (2021).
3. T. O. Yeates, C. A. Kerfeld, S. Heinhorst, G. C. Cannon, J. M. Shively, Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat. Rev. Microbiol.* **6**, 681–691 (2008).
4. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, E. Stackebrandt, The Prokaryotes (Springer New York, 2006).
5. N. C. Hill, J. W. Tay, S. Altus, D. M. Bortz, J. C. Cameron, Life cycle of a cyanobacterial carboxysome. *Sci. Adv.* **6** (2020).
6. A. I. Flamholz, *et al.*, Functional reconstitution of a bacterial co2 concentrating mechanism in E. coli. *Elife* **9**, 1–57 (2020).

7.   B. M. Long, *et al.*, Carboxysome encapsulation of the CO2-fixing enzyme Rubisco in tobacco chloroplasts. *Nat. Commun.* **9**, 1–14 (2018).

8.   B. D. Rae, *et al.*, Cyanobacterial Carboxysomes: Microcompartments that Facilitate CO2 Fixation. *J Mol Microbiol Biotechnol* **23**, 300–307 (2013).

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

10.   J. S. MacCready, *et al.*, Protein gradients on the nucleoid position the carbon-fixing organelles of cyanobacteria. *eLife* **7** (2018).

11.   J. S. MacCready, J. L. Basalla, A. G. Vecchiarelli, Origin and Evolution of Carboxysome Positioning Systems in Cyanobacteria. *Mol. Biol. Evol.* **37**, 1434–1451 (2020).

12.   J. S. MacCready, L. Tran, J. L. Basalla, P. Hakim, A. G. Vecchiarelli, The McdAB system positions α-carboxysomes in proteobacteria. *Mol. Microbiol.* **116**, 277–297 (2021).

13.   J. Lutkenhaus, The ParA/MinD family puts things in their place. *Trends Microbiol.* **20**, 411–418 (2012).

14.   A. G. Vecchiarelli, K. Mizuuchi, B. E. Funnell, Surfing biological surfaces: exploiting the nucleoid for partition and transport in bacteria. *Mol. Microbiol.* **86**, 513–523 (2012).

15.   B. E. Funnell, ParB partition proteins: Complex formation and spreading at bacterial and plasmid centromeres. *Front. Mol. Biosci.* **3**, 44 (2016).

16.   D. F. Savage, B. Afonso, A. H. Chen, P. A. Silver, Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science (80-.).* **327**, 1258–1261 (2010).

17.   R. Rillema, Y. Hoang, J. S. MacCready, A. G. Vecchiarelli, Carboxysome mispositioning alters growth, morphology, and Rubisco level of the cyanobacterium Synechococcus elongatus PCC 7942. *MBio* **12** (2021).

18.   J. L. Basalla, *et al.*, Dissecting the phase separation and oligomerization activities of the carboxysome positioning protein McdB. *Elife* **12** (2023).

19.   P. Hakim, Y. Hoang, A. G. Vecchiarelli, Dissection of the ATPase active site of McdA reveals the sequential steps essential for carboxysome distribution. *Mol. Biol. Cell* **32** (2021).

20.   S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).

21.   S. Alberti, A. Gladfelter, T. Mittag, Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).

22.   D. Schumacher, A. Harms, S. Bergeler, E. Frey, L. Søgaard-Andersen, PomX, a ParA/MinD ATPase activating protein, is a triple regulator of cell division in Myxococcus xanthus. *eLife* **10** (2021).

23.   L. T. Pulianmackal, *et al.*, Multiple ParA/MinD ATPases coordinate the positioning of disparate cargos in a bacterial cell. *Nat. Commun.* **14**, 1–15 (2023).

24.   J. Goedhart, *et al.*, Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nat. Commun. 2012 31* **3**, 1–9 (2012).

25.   N. C. Shaner, *et al.*, A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum. Nat. Methods 2013 105* **10**, 407–409 (2013).

26.   V. E. Gray, R. J. Hause, D. M. Fowler, Analysis of large-scale mutagenesis data to assess the impact of single amino acid substitutions. *Genetics* **207**, 53–61 (2017).

27. J. Wang, *et al.*, A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699.e16 (2018).

28. S. Das, Y. H. Lin, R. M. Vernon, J. D. Forman-Kay, H. S. Chan, Comparative roles of charge, π, and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 28795–28805 (2020).

29. Y. Hoang, C. A. Azaldegui, M. Ghalmi, J. S. Biteen, A. G. Vecchiarelli, An experimental framework to assess biomolecular condensates in bacteria. *bioRxiv*, 2023.03.22.533878 (2023).

30. M. R. Melnicki, M. Sutter, C. A. Kerfeld, Evolutionary relationships among shell proteins of carboxysomes and metabolosomes. *Curr. Opin. Microbiol.* **63**, 1–9 (2021).

31. N. G. A. Abrescia, D. H. Bamford, J. M. Grimes, D. I. Stuart, Structure Unifies the Viral Universe. *Annu. Rev. Biochem.* **81**, 795–822 (2012).

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

33. U. Skoging, P. Liljestro, Role of the C-terminal Tryptophan Residue for the Structure-Function of the Alphavirus Capsid Protein. *J. Mol. Biol.* **279**, 865–872 (1998).

34. I. Komla-Soukha, C. Sureau, A Tryptophan-Rich Motif in the Carboxyl Terminus of the Small Envelope Protein of Hepatitis B Virus Is Central to the Assembly of Hepatitis Delta Virus Particles. *J. Virol.* **80**, 4648–4655 (2006).

35. M. Tsuboi, S. A. Overman, K. Nakamura, A. Rodriguez-Casado, G. J. Thomas, Orientation and Interactions of an Essential Tryptophan (Trp-38) in the Capsid Subunit of Pf3 Filamentous Virus. *Biophys. J.* **34** (2003).

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

37. L. F. Garcia-Alles, *et al.*, Occurrence and stability of hetero-hexamer associations formed by β-carboxysome CcmK shell components. *PLoS One* **14**, e0223877 (2019).

38. M. Sutter, *et al.*, Structure of a Synthetic β-Carboxysome Shell. *Plant Physiol.* **181**, 1050–1058 (2019).

39. Y. Q. Tan, *et al.*, Structure of a Minimal α-Carboxysome-Derived Shell and Its Utility in Enzyme Stabilization. *Biomacromolecules* **22**, 4095–4109 (2021).

40. N. Hirokawa, Y. Noda, Y. Tanaka, S. Niwa, Kinesin superfamily motor proteins and intracellular transport. *Nat. Rev. Mol. Cell Biol.* **10**, 682–696 (2009).

41. J. G. Gindhart, Towards an understanding of kinesin-1 dependent transport pathways through the study of protein–protein interactions. *Brief. Funct. Genomics* **5**, 74–86 (2006).

42. S. Pernigo, A. Lamprecht, R. A. Steiner, M. P. Dodding, Structural basis for kinesin-1:cargo recognition. *Science* **340**, 356–359 (2013).

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

44. R. E. Debaugny, *et al.*, A conserved mechanism drives partition complex assembly on bacterial chromosomes and plasmids. *Mol. Syst. Biol.* **14**, 1–15 (2018).

45. L. Babl, *et al.*, CTP-controlled liquid–liquid phase separation of ParB. *J. Mol. Biol.* **434**, 167401 (2022).
46. B. Guilhas, *et al.*, ATP-Driven Separation of Liquid Phase Condensates in Bacteria. *Mol. Cell* **79**, 293-303.e4 (2020).
47. D. Schumacher, *et al.*, The PomXYZ Proteins Self-Organize on the Bacterial Nucleoid to Stimulate Cell Division. *Dev. Cell* **41**, 299-314.e13 (2017).
48. B. Ramm, *et al.*, Biomolecular condensate drives polymerization and bundling of the bacterial tubulin FtsZ to regulate cell division. *Nat. Commun. 2023 141* **14**, 1–24 (2023).
49. F. Madeira, *et al.*, Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276–W279 (2022).
50. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, WebLogo: A Sequence Logo Generator. *Genome Res.* **14**, 1188–1190 (2004).
51. D. G. Gibson, *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
52. E. M. Clerico, J. L. Ditty, & S. S. Golden, Specialized Techniques for Site-Directed Mutagenesis in Cyanobacteria. *In: Rosato, E.* (eds) *Circadian Rhythms. Methods in Molecular Biology™, 362*. Humana Press. (2007).
53. M. Hutchinson, K. I. Johnstone, D. White, The taxonomy of certain Thiobacilli. *J. Gen. Microbiol.* **41**, 357–366 (1965).
54. J. Schindelin, *et al.*, Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
55. A. Ducret, E. M. Quardokus, Y. V. Brun, MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. *Nat. Microbiol. 2016 17* **1**, 1–7 (2016).
56. S. Bolte, F. P. Cordelières, A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232 (2006).
57. S. Alberti, A. Gladfelter, T. Mittag, Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).
58. N. C. Shaner, *et al.*, Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma sp*. red fluorescent protein. *Nat. Biotechnol.* **22**, 1567–1572 (2004).

Chapter 5

Conclusions and Future Directions

5.1 Conclusions

In this dissertation, I describe my work contributing to our understanding of the biochemistry and mechanism of function of McdB proteins as spatial regulators of carboxysomes and BMCs. As I began my graduate studies, McdB from the cyanobacterium *S. elongatus* was newly identified as the adaptor protein linking the activity of the ATPase McdA to the spatial regulation of carboxysomes [1]. How McdB interacted with itself, with McdA, and with carboxysomes to drive its function were unknown. My work first established McdB as being widespread in BMC-containing bacteria and having the ability to form condensates [2]. I then dissected the phase separation and oligomerization activities of McdB from Se [3]. After these studies in McdB self-association, I identified the specificity determinants that allow McdB proteins to associate with their respective α- or ß-carboxysomes [4]. Together the findings provide some major and overarching conclusions: 1) McdB proteins robustly form condensates both *in vivo* and *in vitro*, and this activity can be tuned by changing the net charge of the N-terminus. 2) Hn McdB forms a disordered monomer and Se McdB forms a trimer-of-dimers hexamer. These differences in oligomeric potential have implications on how different McdB proteins associate with their respective BMCs. 3) Changes in the ability of McdB to form condensates *in vitro* correlate with changes in carboxysome association and positioning function *in vivo*. 4) McdB proteins use similar modes of carboxysome association regardless of subtype,

122

relying on a C-terminal motif containing an invariant tryptophan. My data further suggests that other aromatic residues, specifically tyrosine, could serve a similar role in McdB-like proteins and their association with other BMC types. Overall, my work has significantly advanced our understanding of McdB biochemistry and mechanism, and has provided the groundwork for future directions as detailed below.

5.2 Future Directions

5.2.1 Investigate how the N- and C-terminus of Se McdB interact to influence condensate formation

Several pieces of evidence suggest the N-terminal domain (NTD) and C-terminal domain (CTD) of Se McdB interact to regulate condensate formation and stability. However, the exact regions and residues responsible for McdB self-association remain to be determined. Deleting or mutating positively charged residues in the NTD of Se McdB significantly decreased its condensation activity. Indeed, I found that the net positive charge of the NTD, and not specific residues, was critical for condensate formation; whereby a less positive NTD decreased condensation activity. But the NTD alone was insufficient to form condensates. These findings support the proposal that there are key residues in the CTD of McdB that associate with the NTD, and this association is required for condensate formation. Consistent with this hypothesis, I found that the CTD of McdB was also necessary but insufficient for condensate formation. Our previous bioinformatic analyses of McdB described in Chapter 2 identified NTD-CTD charge asymmetry that is conserved across many McdB proteins. Therefore, it is possible that the negative charge of the CTD plays an important role in McdB condensate formation through

123

electrostatic interactions with the positively charged NTD. In addition, I found that the invariant C-terminal tryptophan was important for McdB condensate formation, suggesting cation-pi interactions may also play a role. Together the data suggest that weak electrostatic interactions between the N- and C-terminus of McdB may mediate condensate formation, consistent with our polyampholyte model described in the Chapter 3 discussion.

To test this proposal, I mixed NTD and CTD peptides of Se McdB, both 20 amino acids long, at a 1:1 molar ratio in buffer [100 mM KCl, 20 mM HEPES pH 7.2, 10% PEG-8000]. No condensates were observed under these conditions. However, it is possible that factors such as oligomerization influence the interactions of these two peptides. One potential future direction would therefore be to test constructs containing the N- and C-terminal peptides fused to different domains that provide varying degrees of oligomerization, such as: 1) disordered, monomeric linkers of varying lengths, 2) a parallel coiled-coil dimer [5], 3) an anti-parallel coiled-coil dimer [5], 4) a parallel coiled-coil hexamer [6], and 6) an anti-parallel coiled-coil hexamer [7]. If condensate formation is reconstituted in these contexts, the N- and C-terminal peptides could be developed as tools to facilitate condensate formation for interacting pairs of proteins, which is currently a major interest for engineering enzyme activity [8].

It may be the case that N- and C-terminal peptides do not interact to drive condensate formation, and none of the above constructs give rise to condensates. However, if this is the case, I still have determined the net charge of the NTD can mediate the solubility of condensates, potentially by affecting solvent interactions [9]. It may therefore be beneficial to try using the NTD as a "condensate solubility tuner" by fusing it onto a condensate forming protein. By making mutations to the net charge of the NTD tag, it may be possible to tune solubility of

condensates. This could allow for control of condensate formation for a range of different proteins, which would be useful for investigating their effects in the cell.

5.2.2 Determine a more detailed structural model of the Se McdB hexamer

My work has generated a rough domain architecture for Se McdB, defined by a disordered N-terminal domain (NTD), a central coiled-coil region, and an α-helical C-terminal domain (CTD). I also determined the contribution of each of these domains to the oligomerization of full-length McdB, where the coiled-coil formed a dimer and the CTD trimerized the dimer to form a trimer-of-dimers hexamer. However, the orientation of the monomers in the hexamer remains to be determined. This information is important for developing our models on how Se McdB interacts with carboxysomes and forms condensates, as I have implemented oligomerization in both of these processes. For instance, parallel monomers within the hexamer with all N- and C-termini oriented in the same direction would indicate that only McdB C-termini interact with carboxysomes and N-termini face the cytoplasm. This would be consistent with the C-terminal tryptophan mediating carboxysome binding and N-terminal basic residues mediating McdA interaction. However, a previous crystal structure of an unconfirmed McdB-like protein from the cyanobacterium *Cyanothece* PCC7424 has identified this putative McdB dimer to be antiparallel [10]. An antiparallel orientation for Se McdB would make sense for mediating interactions between N- and C-termini as described in the above section. Therefore, determining a more detailed structure of the McdB hexamer is important for advancing our models on McdB-carboxysome interactions as well as McdB condensate formation.

As described in Chapter 3, I had difficulty with crystal screens for Se McdB due to its insolubility, where condensate formation occurred across a wide range of buffer conditions screened. However, with the series of NTD mutants I have defined that solubilize McdB without affecting hexamerization, an attractive future direction is to optimize conditions for crystal formation and obtain a crystal structure. Additionally, while the full-length McdB hexamer (~104 kDa) is generally too large for NMR, it may be a useful approach for the stable coiled-coil dimer (~22 kDa). To this end, I have recently established a collaboration with Dr. Nathanial Nucci who specializes in NMR and is interested in coiled-coils. Pursuing this collaboration in the future would be beneficial for investigating the structure of the many coiled-coil containing McdB proteins, which I have shown can vary from one another [11] and potentially influence carboxysome association [4]. Additionally, Nathanial is interested in how coiled-coil domains affect protein condensate formation. My work described in Chapter 3 highlights Se McdB coiled-coil domain as an excellent model for investigating coiled-coil condensate formation, which is an avenue that we are exploring with Nathanial.

5.2.3 Defining the carboxysome association motifs for α- and ß-McdB proteins

My work has identified C-terminal motifs in McdB proteins that contain an invariant tryptophan that is necessary for α- or ß-carboxysome association and positioning. Intriguingly, while the last 10 amino acids of Hn McdB (α-McdB) were sufficient for α-carboxysome localization, the C-terminus of Se McdB (ß-McdB) was not sufficient for ß-carboxysome localization. Therefore, an important future direction is to determine what additional factors drive ß-McdB localization to ß-carboxysomes. For instance, in Se we proposed that McdB hexamerization may provide the required avidity for ß-carboxysome association, which our data

suggest is not required for α-McdB with α-carboxysomes in Hn. Therefore, fusing the C-terminus of Se McdB onto hexameric coiled-coils [6, 7] could help investigate this hypothesis. Additionally, in both α- and ß-McdB proteins, as well as in putative McdB proteins identified for other BMCs, there is an enrichment of acidic residues in the conserved C-terminal motifs surrounding the invariant tryptophan. Determining the effects of mutations to these acidic residues in McdB would therefore also help elucidate how McdB localizes to carboxysomes. Ultimately, determining the full motifs that drive McdB association with carboxysomes in several cases could potentially lead to the development of a general carboxysome localization motif, which would be of great interest to the field.

5.2.4 Determining the McdB interface on carboxysomes

While my work has provided a foundation for probing the McdB determinants required for carboxysome association, the carboxysome components, regions, and residues involved in associating with McdB remain unknown. One difficulty in investigating what residues on the carboxysome are important in this interaction is that McdB was determined to interact with several carboxysome shell components via B2H assays [1]. However, this single assay is the only data we have collected on what aspects of the carboxysome are involved in this interaction. Therefore, several future experiments could be done to further investigate what aspects of the carboxysome interact with the tryptophan motifs of McdB. Here, I introduce several potentially useful techniques, combining microscopy, biochemistry, and *in silico* modelling.

One useful future study would be to investigate the localization of McdB at the surface of carboxysomes using correlative light and electron microscopy (CLEM). CLEM combines both fluorescence and super-resolution microscopy techniques to investigate the interactions of

specific molecular components [12]. For instance, CLEM has been used to investigate details of the interaction between viral particles and host cell membranes [13], which is on the same length scale as carboxysomes. Therefore, this technique could be useful for determining what specific components of carboxysomes McdB is interacting with *in vivo*. Additionally, super-resolution microscopy could be combined with proximity based labelling approaches to further determine components McdB interacts with *in vivo*. A useful approach for proximity labelling proteins within the cytoplasm of cyanobacteria has recently been developed, which works on the length scales of 10-20 nm [14]. The technique works by fusing the enzyme APEX2 onto a protein of interest, which then locally biotinylates proteins when briefly exposed to H_2O_2. Therefore, fusing McdB to APEX2 could be done to identify which carboxysome components are labeled. The information from these approaches could be used to determine which carboxysome components would be useful to include in modeling the interaction with McdB. We have previously established a collaboration with another graduate student, Miguel Limcaoco, who has helped us develop a method for modelling the interaction between ParA/MinD ATPases and their adaptor proteins [15]. Briefly, this approach involves generating a relatively low-energy structural model of the interaction between two proteins. Then, *in silico* scanning mutagenesis can be done on one of the two interacting proteins to monitor how specific mutations are predicted to affect the overall energy of the interacting pair. Lastly, this data can be used to direct mutations *in vivo* to determine the residues within the carboxysome shell that are important for interacting with McdB. These approaches together would make a useful future study that could identify components of the carboxysome shell that are necessary for interacting with external proteins.

5.2.5 Cell-free reconstitution of McdB association with carboxysomes and McdA in vitro

I was able to provide evidence that suggests condensate formation by McdB contributes to its localization to carboxysomes *in vivo*. However, this data is largely correlative and not conclusive. One difficulty is that most of my investigations were done trying to alter the biochemistry of McdB *in vivo* via mutations that were characterized *in vitro*. However, other *in vivo* factors that are affected by these mutations can be confounding. For instance, in Chapter 3 I saw that the interaction between McdB and McdA made interpreting the effects of the N-terminal mutations difficult. Additionally, the C-terminal tryptophan seems to be involved in both McdB condensate formation and in driving McdB localization to carboxysomes. It would therefore be useful to reconstitute the assembly of McdB with carboxysomes *in vitro* to minimize confounding factors and have more control over conditions.

Although carboxysomes have yet to be assembled with recombinant proteins *in vitro*, native carboxysomes can be purified from both Se [16] and Hn [17]. Purifying versions of these carboxysomes with fluorescent Rubisco, would allow for their visualization *in vitro* as we have done *in vivo*. An initial test could be visualizing the colocalization of mTQ labeled carboxysomes with mNG labeled McdB. One question we have, which I describe in the discussion section of Chapter 3, is whether pH fluxes *in vivo* affect McdB localization to carboxysomes. Therefore, investigating the interaction between labeled McdB and carboxysomes *in vitro* is especially attractive because it allows for the control of buffer conditions, and could be used to investigate the effects of pH on McdB-carboxysome association.

Additionally, I have defined several interesting series of mutations to different regions of McdB that affect properties in the cell. For instance, in Se McdB, substituting the invariant tryptophan with tyrosine resulted in an equal amount of McdB localization to carboxysomes as

the wild type McdB. This was especially striking given that tryptophan can be a relatively difficult amino acid to substitute [18]. Further characterizing the tyrosine containing McdB *in vitro* would therefore be useful to understand if it truly is restoring the tryptophan activity. For instance, in addition to observing the binding to purified carboxysomes as described above, Kd values could also be obtained between McdB variants and purified carboxysome shell components using fluorescence anisotropy. Specifically characterizing the how the tyrosine substitution restores tryptophan activity in McdB could have implications for the many McdB-like proteins found near other BMCs that contain tyrosines instead of tryptophans (see Figure 4.8).

Lastly, *in vitro* reconstitution would be useful for studying the interaction between McdB and McdA. Specifically, some of the mutation sets I defined could investigate whether McdB condensate formation affects McdA ATPase activity. One general interest in the field of condensates is how sequestering an enzyme in a condensate may affect its activity [19, 20]. The McdB-McdA pair from Se could be a useful model to test some of these ideas. Recall in Chapter 3, I have defined a series of mutations in the N-terminus of McdB that seem to affect both its condensate forming activity and interaction with McdA. With these mutations, I showed that substituting positive charge in McdB N-terminus with polar glutamines incrementally decreased McdB condensate activity [3]. Importantly, I saw that no specific residue was itself necessary for mediating condensate formation, but rather the overall net charge of the N-terminus was important, where different mutation sets with the same affect on net charge had a similar affect on condensate formation. Recall that positive charge in the N-terminus of adaptor proteins typically mediates their interactions with ParA/MinD ATPases [21], which I provide evidence for for McdB and McdA [3]. For this interaction, there are often only one or two necessary

130

residues in the adaptor protein N-terminus that destroy the interaction with the ATPase when

deleted [21]. Therefore, it could be possible to mutate some positive charge in McdB N-terminus

to diminish its condensate formation without destroying its interaction with McdA. A general

future direction could therefore be to investigate whether McdB condensate formation affects

McdA ATPase activity.

5.2.6 Develop Hn McdB[CTD] as a tool for BMC engineering

An important finding from my work was identifying a 10 amino acid long peptide that

was necessary and sufficient to localize to carboxysomes in Hn. From a biotechnological

standpoint, this 1 kDa peptide can potentially function as a carboxysome-localization tag. For

instance, this tag localized the 45 kDa mNG protein to carboxysomes with a similar efficiency as

wild-type McdB. Importantly, shell components from α-carboxysomes derived from Hn are

being used to develop minimized, synthetic BMC shells for biotechnological purposes [22, 23].

The tag I identified here can expand these efforts to allow for adhering proteins to the outer

surface of these designer shells. These could include both structural components that could

localize BMCs within cells (e.g. via a membrane targeting sequence) as well as enzymes that

could localize reactions to the outer surface of the BMC shell (e.g. to allow local substrate

generation and increase efficiency). Initial experiments could be done with purified synthetic

shells to show that 1) the Hn McdB[CTD] could still facilitate localization of mNG to the surface

of shells from purified carboxysomes *in vitro*, and 2) that the Hn McdB[CTD] fused to a

membrane targeting sequence could help localize these synthetic shells to a membrane *in vivo*.

5.3 References

1. J. S. MacCready, *et al.*, Protein gradients on the nucleoid position the carbon-fixing organelles of cyanobacteria. *eLife* **7** (2018).
2. J. S. MacCready, J. L. Basalla, A. G. Vecchiarelli, J. Echave, Origin and Evolution of Carboxysome Positioning Systems in Cyanobacteria. *Mol. Biol. Evol.* **37**, 1434–1451 (2020).
3. J. L. Basalla, *et al.*, Dissecting the phase separation and oligomerization activities of the carboxysome positioning protein McdB. *Elife* **12** (2023).
4. J. L. Basalla, M. Ghalmi, Y. Hoang, R. Dow, A. G. Vecchiarelli, An invariant C-terminal tryptophan in McdB mediates its interaction and positioning function with carboxysomes. *bioRxiv*, 2023.11.21.568049 (2023).
5. M. G. Oakley, P. S. Kim, A buried polar interaction can direct the relative orientation of helices in a coiled coil. *Biochemistry* **37**, 12603–12610 (1998).
6. N. R. Zaccai, *et al.*, A de novo peptide hexamer with a mutable channel. *Nat. Chem. Biol.* **7**, 935–941 (2011).
7. R. K. Spencer, A. I. Hochbaum, X-ray Crystallographic Structure and Solution Behavior of an Antiparallel Coiled-Coil Hexamer Formed by de Novo Peptides. *Biochemistry* **55**, 3214–3223 (2016).
8. Y. Dai, L. You, A. Chilkoti, Engineering synthetic biomolecular condensates. *Nat. Rev. Bioeng.* **1**, 466–480 (2023).
9. T. Mittag, R. V. Pappu, A conceptual framework for understanding phase separation and addressing open questions and challenges. *Mol. Cell* **82**, 2201–2214 (2022).
10. M. A. Schumacher, M. Henderson, H. Zhang, Structures of maintenance of carboxysome distribution Walker-box McdA and McdB adaptor homologs. *Nucleic Acids Res.* **47**, 5950–5962 (2019).
11. J. S. MacCready, L. Tran, J. L. Basalla, P. Hakim, A. G. Vecchiarelli, The McdAB system positions α-carboxysomes in proteobacteria. *Mol. Microbiol.* **116**, 277–297 (2021).
12. Y. W. Chang, *et al.*, Correlated cryogenic photoactivated localization microscopy and cryo-electron tomography. *Nat. Methods* **11**, 737–739 (2014).
13. R. Santarella-Mellwig, *et al.*, Correlative Light Electron Microscopy (CLEM) for Tracking and Imaging Viral Protein Associated Structures in Cryo-immobilized Cells. *JoVE (Journal Vis. Exp.* **139**, e58154 (2018).
14. K. K. Dahlgren, C. Gates, T. Lee, J. C. Cameron, Proximity-based proteomics reveals the thylakoid lumen proteome in the cyanobacterium Synechococcus sp. PCC 7002. *Photosynth. Res.* **147**, 177–195 (2021).
15. L. T. Pulianmackal, *et al.*, Multiple ParA/MinD ATPases coordinate the positioning of disparate cargos in a bacterial cell. *Nat. Commun.* **14**, 1–15 (2023).
16. M. Faulkner, *et al.*, Direct characterization of the native structure and mechanics of cyanobacterial carboxysomes. *Nanoscale* **9**, 10662–10673 (2017).

17. L. A. Metskas, *et al.*, Rubisco forms a lattice inside alpha-carboxysomes. *Nat. Commun.* **13**, 1–9 (2022).

18. R. Trivedi, H. A. Nagarajaram, Substitution scoring matrices for proteins - An overview. *Protein Sci.* **29**, 2150–2163 (2020).

19. C. D. Reinkemeier, E. A. Lemke, Synthetic biomolecular condensates to engineer eukaryotic cells. *Curr. Opin. Chem. Biol.* **64**, 174–181 (2021).

20. Y. Dai, L. You, A. Chilkoti, Engineering synthetic biomolecular condensates. *Nat. Rev. Bioeng.* **1**, 466–480 (2023).

21. D. Schumacher, A. Harms, S. Bergeler, E. Frey, L. Søgaard-Andersen, PomX, a ParA/MinD ATPase activating protein, is a triple regulator of cell division in Myxococcus xanthus. *eLife* **10** (2021).

22. Y. Q. Tan, *et al.*, Structure of a Minimal α-Carboxysome-Derived Shell and Its Utility in Enzyme Stabilization. *Biomacromolecules* **22**, 4095–4109 (2021).

23. C. A. Kerfeld, M. Sutter, Engineered bacterial microcompartments: apps for programming metabolism. *Curr. Opin. Biotechnol.* **65**, 225–232 (2020).