An invariant C-terminal tryptophan in McdB mediates its interaction and positioning function with carboxysomes

Joseph L. Basalla, Maria Ghalmi, Y. Hoang, Rachel E. Dow, and Anthony G. Vecchiarelli*
Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI 48109

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 β, which are structurally and phyletically distinct. We recently characterized the maintenance of carboxysome distribution (Mcd) systems responsible for spatially regulating α- and β-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 β-carboxysomes, remain unknown. Here, we identify C-terminal motifs containing an invariant tryptophan necessary for α- and β-McdBs to associate with α- and β-carboxysomes, respectively. Substituting this tryptophan with other aromatic residues reveals corresponding gradients in the efficiency of 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.**

SIGNIFICANCE STATEMENT

- Maintenance of carboxysome distribution protein B (McdB) is necessary for positioning a widespread class of bacterial organelles that regulate metabolism. Without McdB, these organelles aggregate and lose functionality. How McdB interacts with and positions these organelles is unknown.
- The authors determined that an invariant tryptophan is necessary for McdB to interact with and position carboxysomes. A similar mechanism occurs in two diverse bacterial cell types, both relying on the invariant tryptophan.
- This class of bacterial organelles has been implemented in aspects of human pathology and the development of carbon-capturing technologies. Our results advance our understanding and applications of these organelles.

Monitoring Editor
Erin Goley
Johns Hopkins University

Received: Nov 20, 2023
Revised: Jun 10, 2024
Accepted: Jun 17, 2024

Cross-Species

Cross-Validation

New Hypothesis

New Methods

This article was published online ahead of print in MBoC in Press (http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E23-11-0443) on July 11, 2024.

Author contributions: J.L.B. conceived and designed the experiments; J.L.B., M.G., Y.H., and R.D. performed the experiments; J.L.B., M.G., and Y.H. analyzed the data; J.L.B. drafted the article; J.L.B. prepared the digital images.

Conflicts of interests: The authors declare no financial conflict of interest.

Materials availability statement: Any materials used in this study are described in detail in *Materials and Methods* or can be accessed by contacting the authors.

Data availability statement: The data that support the findings of this study are available in the Supplemental Material of this article or can be accessed by contacting the authors.

*Address correspondence to: Anthony G. Vecchiarelli (ave@umich.edu).

Abbreviations used: BMC, bacterial microcompartment; CD, circular dichroism; CTD, C-terminal domain; DIC, differential interference contrast; GRM, glycyl radical enzyme-containing microcompartment; Hn, *halothiobacillus neapolitanus*; IPTG, sopropyl β-D-1-thiogalactopyranoside; McdA, maintenance of carboxysome distribution protein A; McdB, maintenance of carboxysome distribution protein B; mNG, mNeonGreen; MSA, multiple sequence alignment; mTQ, mTurquoise2; NTD, N-terminal domain; PDU, 1,2-propanediol utilization microcompartment; Rubisco, ribulose-1/5,-bisphosphate carboxylase/oxygenase; Se, synechococcus elongatus PCC 7942; SEC, size exclusion chromatography; SEC-MALS, size exclusion chromatography with multi-angle static light scattering.

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 (Kerfeld *et al.*, 2018; Sutter *et al.*, 2021). BMCs consist of selectively-permeable protein shells that encapsulate a set of enzymes, thus serving as nanoscale reaction centers for key metabolic steps (Sutter *et al.*, 2021). A recent bioinformatic survey identified 68 unique BMC types in 45 bacterial phyla (Sutter *et al.*, 2021), 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 (Yeates *et al.*, 2008). Carboxysomes encapsulate the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) and coconcentrate it with its substrate CO_2 to significantly increase the efficiency of carbon fixation in many autotrophic bacteria (Yeates *et al.*, 2008). As a result, carboxysomes are estimated to facilitate about 35% of all global carbon fixation (Dworkin *et al.*, 2006; Hill *et al.*, 2020), making carboxysomes of interest for developing carbon-capturing technologies (Long *et al.*, 2018; Flamholz *et al.*, 2020). Beyond their biotechnological potential, carboxysomes are also the paradigm for understanding fundamental aspects of general BMC biology, such as assembly, structure, and spatial regulation (Yeates *et al.*, 2008; Kerfeld *et al.*, 2018; Sutter *et al.*, 2021). 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 β, where β-carboxysomes are found in β-cyanobacteria and α-carboxysomes are found in numerous phylogenetically distinct groups, including α-cyanobacteria and several types of chemoautotrophic bacteria (Sutter *et al.*, 2021). While functionally equivalent, α- and β-carboxysomes are structurally and phyletically distinct, with key differences in composition, mode of assembly, and regulation (Rae *et al.*, 2013; Kerfeld and Melnicki, 2016). In fact, α-carboxysomes are more closely related to other BMC types than they are to β-carboxysomes (Kerfeld and Melnicki, 2016). Therefore, α- and β-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 β-carboxysomes down the cell length (Maccready *et al.*, 2018; MacCready *et al.*, 2020; MacCready *et al.*, 2021). 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 (Lutkenhaus, 2012; Vecchiarelli *et al.*, 2012). The second component is a novel protein we named McdB, which interacts with McdA and also localizes to carboxysomes (Maccready *et al.*, 2018), thus acting as an adaptor to link the carboxysome cargo to its positioning ATPase (Funnell, 2016). 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 (Savage *et al.*, 2010; Rillema *et al.*, 2021). 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 (MacCready *et al.*, 2020; MacCready *et al.*, 2021). 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 (MacCready *et al.*, 2021). 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 (MacCready *et al.*, 2020; MacCready *et al.*, 2021). For example, *Se* McdB is largely α-helical with a coiled-coil domain and forms a trimer-of-dimers hexamer (Basalla *et al.*, 2023), whereas *Hn* McdB is monomeric and completely intrinsically disordered (MacCready *et al.*, 2021). 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 β-carboxysomes. Despite their diversity, α- and β-McdB proteins have been shown to form condensates in vitro (MacCready *et al.*, 2020; MacCready *et al.*, 2021; Basalla *et al.*, 2023). The processes underlying condensate formation in vitro can influence subcellular organization in vivo in both eukaryotes and prokaryotes (Banani *et al.*, 2017; Hakim *et al.*, 2021). 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 (Alberti *et al.*, 2019). 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 β-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 (MacCready *et al.*, 2021). 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.

RESULTS

All McdB proteins contain an invariant tryptophan in their C-terminal region

We first performed multiple sequences alignments both within and across α- and β-McdB types to identify regions of conservation that

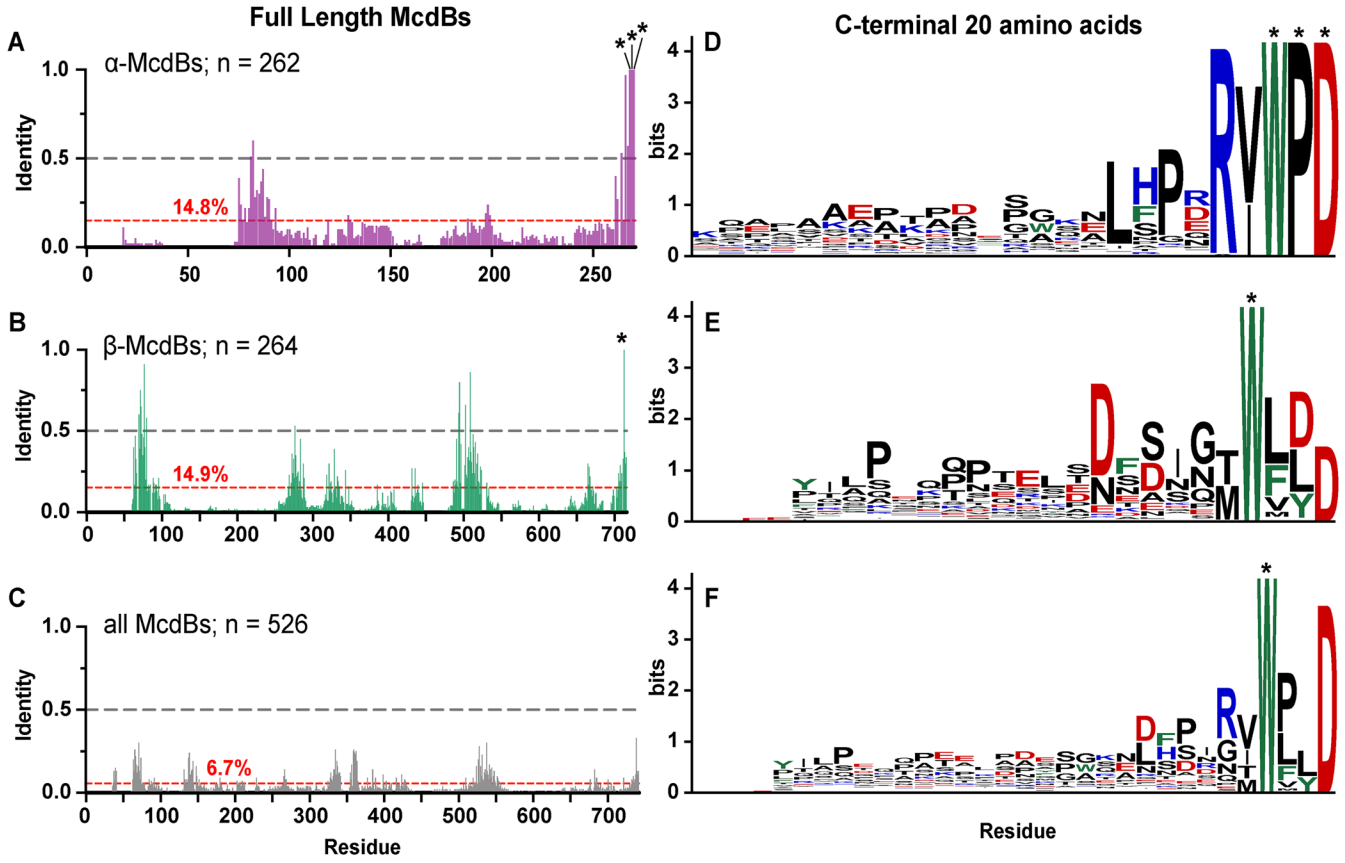

FIGURE 1: All McdB proteins share an invariant C-terminal tryptophan. (A–C) Percent identities from MSAs of full-length (A) α-McdBs, (B) β-McdBs, and (C) both α- and β-McdBs. The average percent identity for each alignment is shown in red. (D–F) Sequence logos generated from MSAs of only the last 20 C-terminal amino acids of (D) α-McdBs, (E) β-McdBs, and (F) both α- and β-McdBs. Positions that are invariant (100% identity) are indicated with an asterisk. Cationic residues are colored blue, anionic residues red, aromatic residues green, and all others black.

may be involved in associating with carboxysomes. On average, McdBs show low sequence identity; 14.8% among α-McdBs (Figure 1A), 14.9% among β-McdBs (Figure 1B), and 6.7% across all McdBs (Figure 1C). One reason for this low average identity is the large alignment gaps that stem from the high variance in sequence lengths, which range from 51 to 169 residues for α-McdBs and 132 to 394 residues for β-McdBs (Supplemental Figure S1, A–C). Although the average identity was low, we identified three invariant residues (WPD) in the C-terminal region of α-McdBs (Figure 1A), and one invariant residue (W) in the C-terminal region of β-McdBs (Figure 1B), consistent with our previous reports (MacCready *et al.*, 2020; MacCready *et al.*, 2021). Aligning all full-length McdB protein sequences did not identify any invariant residues (Figure 1C), again due to the large variations in protein lengths, which prevented the C-termini of α- and β-McdBs from aligning (Supplemental Figure S1C). We therefore repeated these alignments on only the last 20 C-terminal amino acids of all McdB sequences to control for length variation (Figure 1, D–F). These alignments unveiled motif conservation specific to α - and β-McdB types as well as conservation shared across all McdBs. α-McdBs contain a short consensus sequence of R(V/I)WPD at the C-terminus (Figure 1D). The β-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 1E). And across all McdBs, we identified the tryptophan as the only invariant residue (Figure 1F), suggesting a critical role in its functionality for both α- and β-McdBs.

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 (Schumacher *et al.*, 2021), and we have computationally shown that this is likely the case for McdB interactions with McdA (Pulianmackal *et al.*, 2023). Therefore, our identification of an invariant tryptophan within the C-terminal region 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) (Goedhart *et al.*, 2012) 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 were made in an McdB variant that was N-terminally fused to the fluorescent protein monomeric NeonGreen (mNG) (Shaner *et al.*, 2013). 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 (Maccready *et al*., 2018). 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* (MacCready *et al*., 2021). Finally, we also performed phase-contrast imaging to monitor cell morphology.

In *Se*, the invariant tryptophan is the final C-terminal amino acid (Supplemental Figure S1D), therefore we simply deleted it to make McdB[ΔW152]. As shown previously in wild type *Se* cells (Maccready *et al*., 2018), mNG-McdB colocalized with carboxysome foci that are uniformly distributed down the cell length (Figure 2A). Without the invariant tryptophan, mNG-McdB[ΔW152] was diffuse in the cell (Figure 2A), with no notable carboxysome colocalization compared with that of wild type (Figure 2B). Carboxysomes were also mispositioned (Figure 2C) and clustered into high-intensity aggregates in the McdB[ΔW152] strain (Supplemental Figure S2B), which phenocopies a complete deletion of McdB, as shown previously (Rillema *et al*., 2021).

For *Hn* McdB, the invariant tryptophan is the third amino acid from the C-terminus (Supplemental Figure S1D). Therefore, a glycine substitution was used to make McdB[W94G]. Although the mNG-McdB fusion destroys the carboxysome positioning function of McdB, the fusion still strongly colocalized with the mispositioned carboxysome aggregates at the cell pole, hence the label "wild type*" (Figure 2D). As with *Se* McdB[ΔW152], mNG-McdB[W94G] was diffuse in *Hn* cells and did not colocalize with carboxysomes (Figure 2E).

While mNG-McdB[W94G] allowed us to observe changes in protein localization, we were unable to determine the effects of this mutant protein on carboxysome positioning when fused to mNG. Therefore, we also imaged carboxysome distribution in *Hn* cells with unlabeled McdB[W94G] (Figure 2F). Consistent with *Se* McdB[ΔW152], carboxysomes became mispositioned (Figure 2G) and clustered into high-intensity aggregates in the McdB[W94G] mutant strain (Supplemental Figure S2B), which phenocopies a complete deletion of McdB in *Hn*, as shown previously (MacCready *et al*., 2021). Together, our results show that the C-terminal tryptophan found in all McdB proteins is necessary for carboxysome association and positioning.

Loss of carboxysome association and positioning is not due to destabilization of mutant McdB proteins

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 (SEC) with multi-angle light scattering (SEC-MALS) to determine the secondary and quaternary structures of the purified proteins, respectively. As shown previously (MacCready *et al*., 2021; Basalla *et al*., 2023), wild type *Se* McdB forms a hexamer in solution with an α-helical signature, whereas *Hn* McdB is monomeric and completely intrinsically disordered (Supplemental Figure S3, A and B). Because 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 with that of wild type (Supplemental Figure S3C). 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 β-McdB interactions with their respective carboxysomes.

The C-terminal region of monomeric *Hn* McdB is necessary and sufficient for carboxysome association, but not for hexameric *Se* McdB

We next set out to determine whether 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 3A). 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 (Basalla *et al*., 2023). 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 (Supplemental Figure S3, A and 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 1D).

Intriguingly, the mNG-CTDs from *Se* and *Hn* McdB displayed different localizations. *Se* McdB[CTD] was completely diffuse in the cell (Figure 3B), and showed no association with carboxysomes, similar to that of McdB[ΔW152] (Figure 3, B and C). *Hn* McdB[CTD], on the other hand, strongly colocalized with carboxysomes (Figure 3D), similar to that of wild type* (Figure 3E). The data show that the last 10 amino acids of *Hn* McdB are sufficient for associating with α-carboxysomes. However, the last 31 amino acids of *Se* McdB, despite encoding the conserved motif and invariant tryptophan, are insufficient.

Recall that full-length *Se* McdB is a hexamer in solution whereas *Hn* McdB is monomeric (Supplemental Figure S3A). Furthermore, we have previously shown that the CTD of *Se* McdB has an α-helical secondary structure (Basalla *et al*., 2023), while *Hn* McdB is completely disordered (MacCready *et al*., 2021). It is 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 domain (NTD) 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) (Supplemental Figure S4A), indicating that *Se* McdB[CTD] remains monomeric. CD analysis confirmed that *Se* McdB[CTD] retained an α-helical structure (Supplemental Figure S4B). 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 (Supplemental Figure S4A).

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.

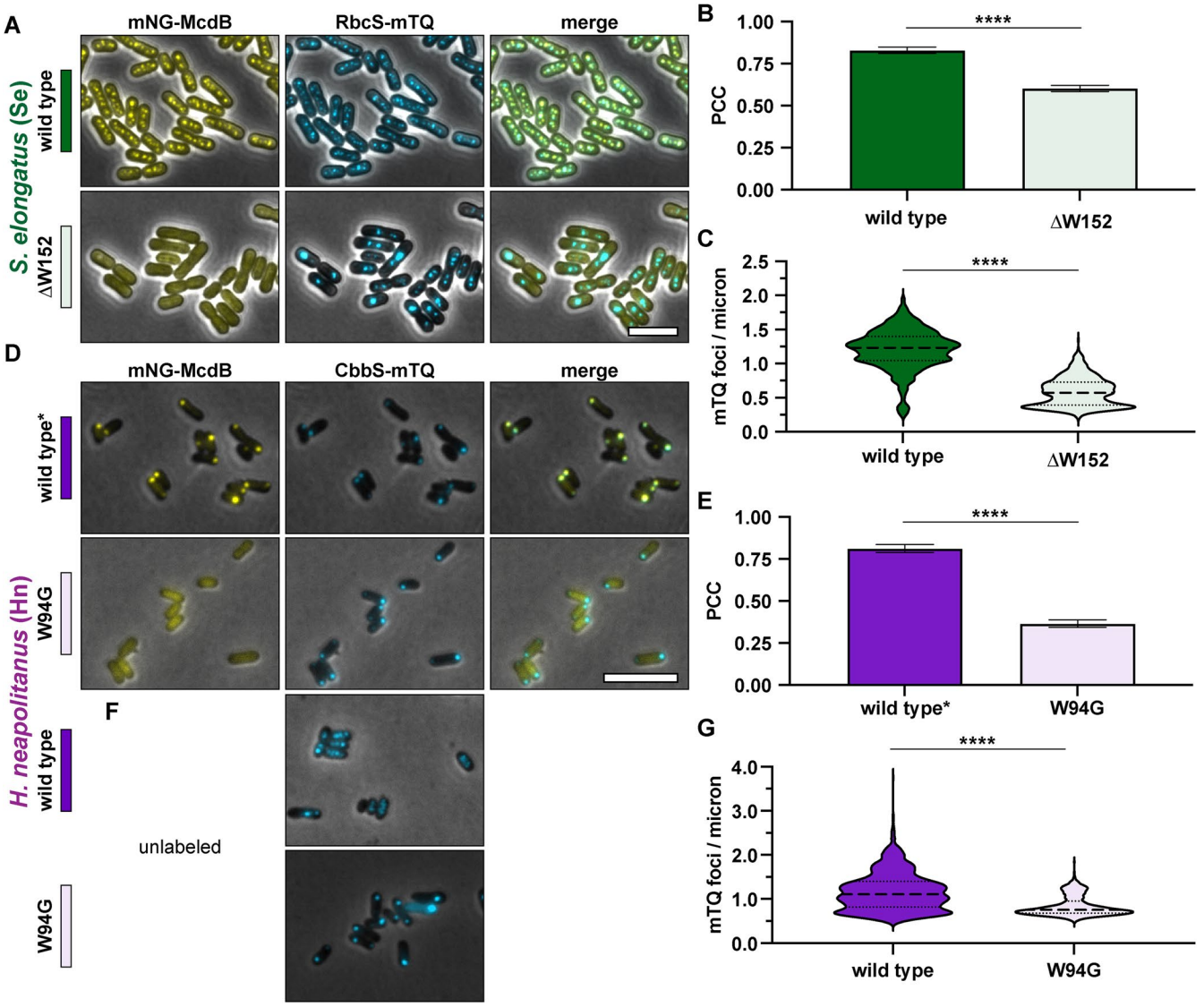

FIGURE 2: The invariant tryptophan for both α- and β-McdBs mediates carboxysome localization. (A) Representative microscopy images of the indicated *Se* strains. Phase-contrast images are shown in black and white and overlaid with the fluorescence channels: mNG-McdB proteins are yellow and RbcS-mTQ–labeled carboxysomes are cyan. Colored bars next to the strain names correspond to colors on the associated graphs. (B) PCCs quantified for the indicated *Se* strains. Graphs represent means and standard deviations from seven technical replicates. ****, $p < 0.001$ 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 three biological replicates each with $n > 500$ cells. ****, $p < 0.001$ from Mann–Whitney U-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–labeled carboxysomes are cyan. Colored bars next to the strain names correspond to colors on the associated graphs. (E). PCCs quantified for the indicated *Hn* strains. Graphs represent means and standard deviations from seven technical replicates ****, $p < 0.001$ 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 three biological replicates each with $n > 500$ cells. ****, $p < 0.001$ from Mann–Whitney U-test. For exact cell and foci counts for each strain and replicate, see Supplemental Table S1. Scale bars are 5 μm and apply to all images.

Other aromatic residues functionally replace the invariant tryptophan with varying activity

Our previous bioinformatic analyses identified putative McdB-like proteins associated with other BMCs (MacCready *et al.*, 2020), including the 1,2-propanediol utilization microcompartment (PDU) and the glycyl radical enzyme-containing microcompartment (GRM) (Kerfeld *et al.*, 2018). Intriguingly, the C-termini of these McdB-like proteins lack the invariant tryptophan (Figure 4A). Instead, we found tyrosine (Y) or phenylalanine (F) residues conserved within the last five amino acids (Figure 4B), suggesting other C-terminal aromatics could potentially fulfill the role of the invariant tryptophan.

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

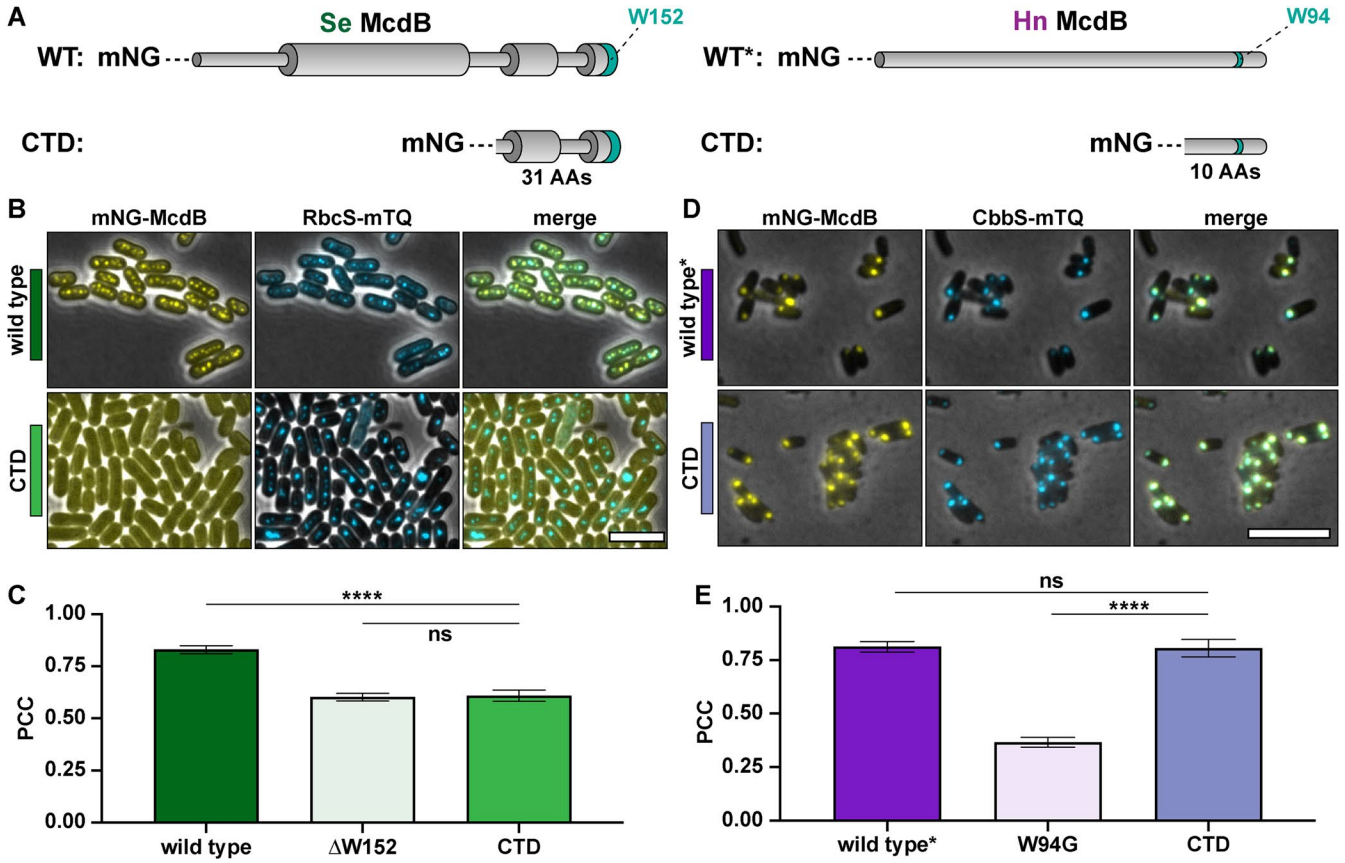

FIGURE 3: The C-termini of α- and β-McdBs show differences in their ability to localize to carboxysomes. (A) McdB protein models for *Se* and *Hn* wild type (WT) and CTDs. Wide cylinders represent α-helical regions and narrow cylinders represent region of intrinsic disorder. The invariant tryptophan (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–labeled carboxysomes are cyan. Colored bars next to the strain names correspond to colors on the associated graphs. (C). PCCs quantified for the indicated *Se* strains. Graphs represent means and standard deviations from seven technical replicates. ****, $p < 0.001$ and ns = nonsignificant 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–labeled carboxysomes are cyan. Colored bars next to the strain names correspond to colors on the associated graphs. (E) PCCs quantified for the indicated *Se* strains. Graphs represent means and standard deviations from seven technical replicates. ****, $p < 0.001$ and ns = nonsignificant from Welch's *t* test. Scale bars are 5 μm and apply to all images.

in the carboxysome-labeled strains. We found that *Se* McdB[W152Y] colocalized with carboxysomes to the same degree as wild type McdB (Figure 5, A and 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 5C; Supplemental Figure S5A). *Se* McdB[W152Y] still distributed carboxysomes, albeit with slightly perturbed spacing and higher foci intensities compared with that of wild type. *Se* McdB[W152F] showed an even lesser degree of carboxysome positioning compared with 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 5E). And none of the McdB mutants were capable of positioning carboxysomes in *Hn* (Figure 5F; Supplemental Figure S5B).

To summarize, we found a striking gradient of McdB variant colocalization with carboxysomes that followed similar trends: W152 ≈ W152Y > W152F > ΔW152 in *Se* (Figure 5B), and W94 > W94Y > W94F ≈ W94G in *Hn* (Figure 5E). Furthermore, in *Se*, the gradient of carboxysome association directly correlated with the carboxysome positioning function of the McdB mutants (Figure 5C; Supplemental Figure S5A). In *Hn*, however, none of the McdB mutants restored carboxysome positioning (Figure 5F; Supplemental Figure S5B). Additionally, these data provide supportive evidence that mutations to the invariant tryptophan do not cause destabilization or degradation of McdB proteins, as some mutations maintained proper McdB functionality.

Overall, in both *Se* and *Hn*, substituting the conserved tryptophan with tyrosine provided strong McdB localization to carboxysomes, compared with 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

A

Strain:	***Synechococcus elogatus* PCC 7942**
BMC:	β-carboxysome
CTD10:	DQLKR KLHGW
Strain:	***Halothiobacillus neapolitanus* c2**
BMC:	α-carboxysome
CTD10:	STHPR RVWPD
Strain:	***Eubacterium callanderi* KIST 612**
BMC:	1,2-Propanediol Utilization (PDU)
*CTD10:	KLHEL PDYLL
Strain:	***Roseburia inulinivorans* DSM 16841**
BMC:	Glycyl Radical enzyme-containing Microcompartment (GRM)
*CTD10:	VGDAM PIYFF

CTD10 = Last 10 C-terminal amino acids of the specified McdB-like protein

* These McdB-like proteins not been experimentally verified to be involved in BMC organization

B

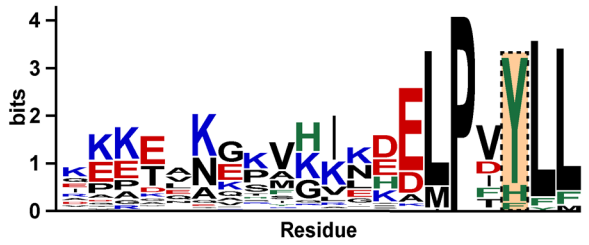

FIGURE 4: McdB-like proteins found near the operons of noncarboxysome BMCs have aromatics other than tryptophan at their C-termini. (A) Table displaying the last 10 C-terminal amino acids of McdBs from *Se* and *Hn*, as well as from putative McdBs from bacteria containing the indicated BMCs. Acidic residues 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 MSAs of the last 20 C-terminal amino acids of McdBs from noncarboxysome BMCs. The position corresponding to the invariant tryptophan as shown in A is boxed and highlighted.

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 (Gray *et al.*, 2017). We conclude that C-terminal aromatic residues can drive the localization of McdBs to both α- and β-carboxysomes, and suggests a similar role for the conserved C-terminal aromatic residues found in putative McdB-like proteins associated with other BMC types.

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 (Banani *et al.*, 2017; MacCready *et al.*, 2020; MacCready *et al.*, 2021; Basalla *et al.*, 2023). 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 (Alberti *et al.*, 2019). Our biochemical characterization also suggested these N-terminal positive residues may associate with the C-terminal regions 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 (Wang *et al.*, 2018; Das *et al.*, 2020). Interestingly, these studies have shown gradients of condensate formation by changing the type of aromatic residues involved (Wang *et al.*, 2018). We set out to determine whether 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 (Basalla *et al.*, 2023). 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 6A). 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 5, A–C).

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 6, B and C), which correlates with the inability of these *Hn* McdB variants to associate with and position carboxysomes in vivo (see Figure 5, D–F).

Finally, to investigate the condensate forming activity and relative expression levels of McdB variants in vivo, we expressed mCherry-tagged McdB constructs in *Escherichia coli* (*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 (Supplemental Figure S6A) are liquid-like condensates (Basalla *et al.*, 2023; Hoang *et al.*, 2024). Intriguingly, the tryptophan mutants once again displayed the same functional gradient in forming condensates in vivo (Supplemental Figure S6B). Importantly, all *Se* McdB variants showed the same levels of expression compared with wild type, and with no notable degradation (Supplemental Figure S6C). 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. Additionally, these data show that the presence of a fluorescence tag does not change the trend in condensate solubility seen in vitro. Together, we once again find a gradient of condensation activity in vivo that mirrors our in vitro condensation 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 (Supplemental Figure S6D). This is consistent with *Hn* McdB requiring significantly higher protein concentrations to form condensates in vitro, compared with *Se* McdB (see Figure 6, A and B). Because we could not form foci with wild type *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.

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

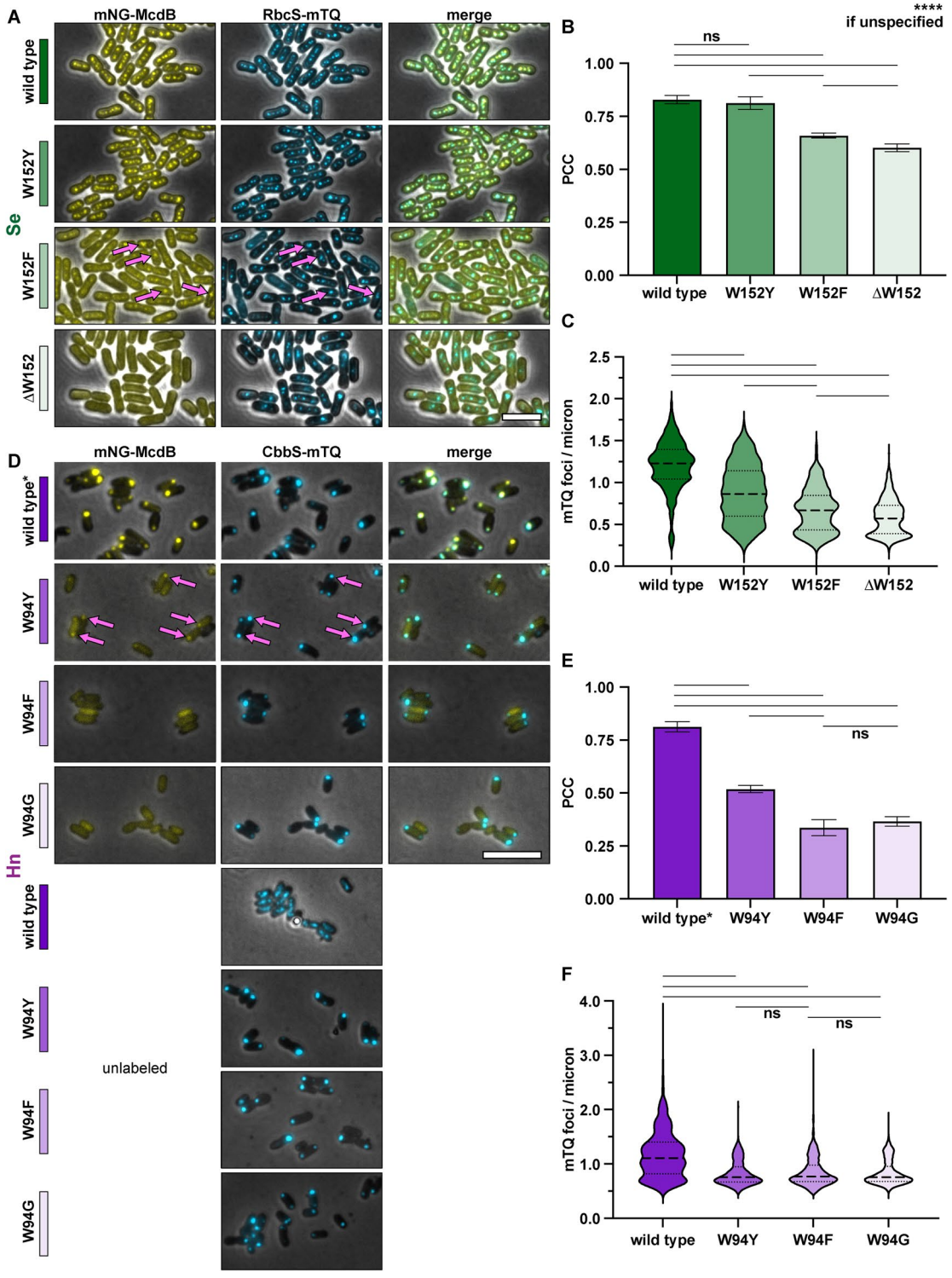

FIGURE 5: Changing the invariant tryptophan to other aromatic residues reveals a gradient of McdB colocalization with carboxysomes. (A) Representative microscopy images of the indicated *Se* strains. Phase-contrast images are shown in black and white and overlaid with the fluorescence channels: mNG-McdB proteins are yellow and RbcS-mTQ–labeled 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). PCCs quantified for the indicated *Se* strains. Graphs represent means and standard deviations from seven technical replicates. ****, $p < 0.001$ and ns = nonsignificant 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 three 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*. For exact cell and foci counts for each strain and replicate, see Supplemental Table S1. Scale bars are 5 μm and apply to all images.

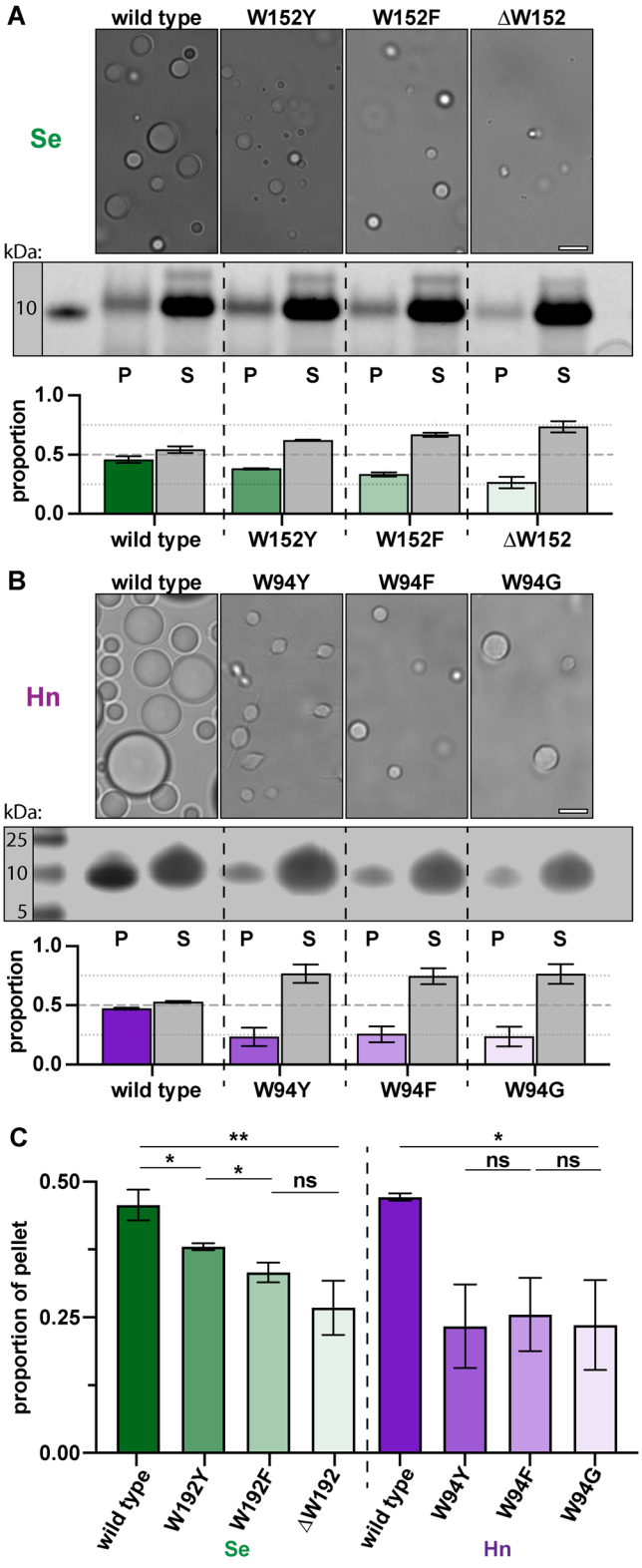

FIGURE 6: Changing the invariant tryptophan to other aromatic residues reveals a gradient of condensate formation for *Se* McdB but not *Hn* McdB. (A) (top) Representative DIC microscopy images of the indicated *Se* McdB variants at 50 μM (100 mM KCl; 20 mM HEPES, pH 7.2) after 1 h. (middle) Samples 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 SD from three biological 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 h. (C) Quantification summary of pellet fractions from *Se*- and *Hn*-McdB variants. **, $p < 0.01$; *, $p < 0.05$, ns = nonsignificant from Welch's *t* test. Scale bars are 5 μm and apply to all microscopy images.

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 1). We found that the invariant tryptophan is necessary for α- and β-McdBs to colocalize with α- and β-carboxysomes, respectively. With the invariant tryptophan removed, McdB became diffuse in the cytoplasm (Figure 2). Interestingly, the C-terminal motif containing the invariant tryptophan was necessary and sufficient for carboxysome localization by α-McdB of *Hn*, but not for β-McdB of *Se* (Figure 3). 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 (Supplemental Figure S4).

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). We therefore attempted to complement the removal of the invariant tryptophan in both α- and β-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 5). Lastly, this gradient of function in vivo directly correlated with the ability of these McdB variants to form condensates both in vitro (Figure 6) and in vivo (Supplemental Figure S6D). A summary of our findings is provided in Figure 7. 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.

McdB condensate formation in the context of carboxysome structure, function, and positioning

Several recent studies have shown that, in both prokaryotes and eukaryotes, Rubisco can form cocondensates with a scaffolding protein to aid in the assembly of mesoscale structures (Freeman *et al.*, 2017; Wang *et al.*, 2019; Oltrogge *et al.*, 2020; Zang *et al.*, 2021). These data have inspired a model where Rubisco-based structures may have evolved from condensates, and hints at the importance of condensate formation in the assembly of associated proteins (Long *et al.*, 2021). Importantly, while condensate formation has been implemented in the initial assembly of carboxysomes, the final structure of carboxysomes is still under active investigation to determine whether they are liquid-like and dynamic, more solid-like lattices, or can fluctuate based on environment (Niederhuber *et al.*, 2017; Sun *et al.*, 2019; Metskas *et al.*, 2022). Therefore, a current model is that condensate formation could underly the molecular assembly of carboxysome-associated molecules, but the final structure itself may not be reminiscent of a liquid-like condensate, a phenomenon which has been described for other systems (Yoo *et al.*, 2019).

Here, by mutating a critical tryptophan to other aromatic amino acids, we show correlations between McdB condensate formation and its ability to assemble with carboxysomes. An alternative explanation to this data is that the invariant tryptophan forms a hydrophobic pocket within the carboxysome shell that helps stabilize the

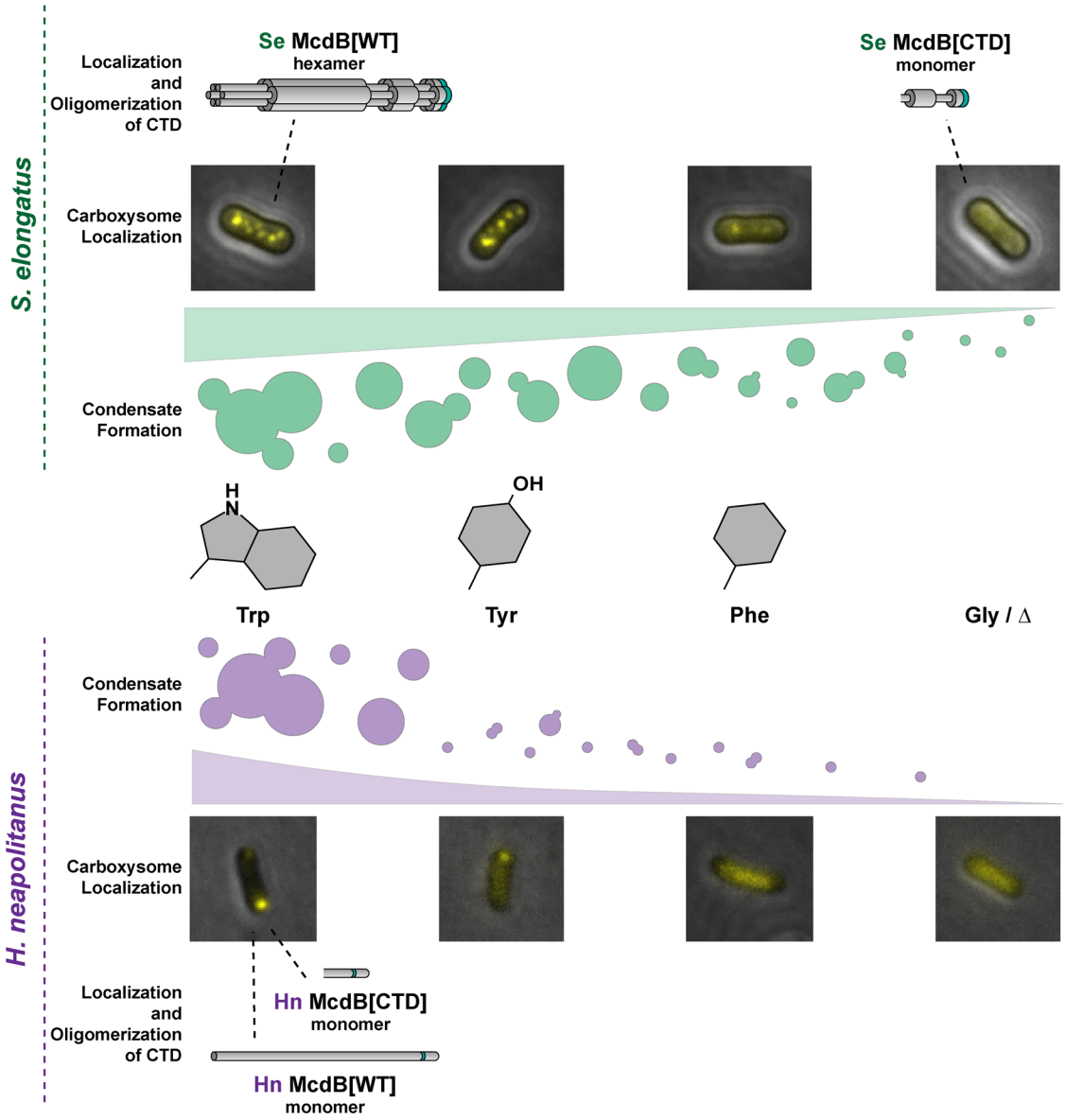

FIGURE 7: Summary of condensate formation and carboxysome localization for the aromatic substitution mutants of *Se* (top) and *Hn* (bottom) McdB. Wild type (WT) McdB proteins from *Se* and *Hn* have an invariant C-terminal tryptophan (trp), depicted as a green stripe in the cartoon protein models. *Se* 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 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).

association. However, it is important to note that the trend we see in McdB localization and condensate formation (Trp > Tyr > Phe) (Figures 5 and 6) does not directly follow the hydrophobicity of these amino acids (Trp > Phe > Tyr) (Wimley and White, 1996). Therefore, while hydrophobic interactions may help stabilize McdB–carboxysome associations, they cannot fully explain the trend in our data. We conclude that condensate formation, which does correlate with localization, contributes to the assembly of McdB with carboxysomes in congruence with the model described for Rubisco above.

Understanding the details of the molecular mechanism by which McdB localizes to carboxysomes is important for advancing our models of BMC assembly as well as advancing our understanding of the roles of condensate formation in biology. While the data provided here help advance these concepts, several follow-up studies are important for continued development of our proposed model. For instance, recent work has described how membranes can influence the condensate formation of associated proteins in a process termed two-dimensional phase separation (Ditlev, 2021). Whether the protein surface of the carboxysomes has a similar effect on McdB as lipid membranes have on other proteins will be an interesting comparison. Additionally, while we have characterized the impacts of different Se McdB domains on condensate formation (Basalla *et al.*, 2023), a similar characterization of Hn McdB is lacking. Given the differences seen between these proteins in this current study (e.g., Figure 3), it will be important to deepen our understanding of which regions of Hn McdB are sufficient for condensate formation.

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 (Kerfeld *et al.*, 2018). While the set of encapsulated core enzymes is highly diverse, the outer shell proteins of BMCs are well conserved in sequence and structure (Kerfeld *et al.*, 2018; Sutter *et al.*, 2021). For all BMCs, several different types of protein oligomers build the outer shell to give rise to the characteristic polyhedral shape of a BMC (Kerfeld *et al.*, 2018); 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 (Maccready *et al.*, 2018). 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 (Melnicki *et al.*, 2021; Sutter *et al.*, 2021). Case in point, the BMC-H shell proteins from α- and β-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 (Melnicki *et al.*, 2021; Sutter *et al.*, 2021). Therefore, α- and β-carboxysomes have distinct evolutionary histories, but converged on a functionally homologous BMC type (Rae *et al.*, 2013; Kerfeld and Melnicki, 2016; Melnicki *et al.*, 2021). It is intriguing that α- and β-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 β-McdB. Future studies such as this will help deepen our understanding of McdB localization to carboxysomes, and protein localization to BMCs in general.

Tryptophan mediates the assembly of several viral and phage capsids

BMC shells share several analogous features to viral capsids (Abrescia *et al.*, 2012). 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 (Krupovic and Koonin, 2017; Kerfeld *et al.*, 2018). Although structurally similar, BMC shells and viral capsids likely evolved independently, representing multiple convergent events (Abrescia *et al.*, 2012). 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 (Skoging and Liljestro, 1998; Tsuboi *et al.*, 2003; Komla-Soukha and Sureau, 2006). 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 (Cai *et al.*, 2015; Garcia-Alles *et al.*, 2019). 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 β-carboxysome shells to be used as tools in synthetic biology and biotechnological applications (Sutter *et al.*, 2019; Tan *et al.*, 2021). 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.

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 (Hirokawa *et al.*, 2009). How kinesin-1 recognizes and binds to a diverse set of cargo proteins to facilitate their transport has been under investigation for decades (Gindhart, 2006). 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 (Pernigo *et al.*, 2013).

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 β-McdBs tend to have acidic residues (often D) within five amino acids of the conserved tryptophan (see Figure 1). This is also true for the putative McdB-like proteins we identified for other BMCs (see Figure 4), although these contain aromatic 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 (Mahinthichaichan *et al.*, 2018). 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.

MATERIALS AND METHODS

Request a protocol through *Bio-protocol*.

Multiple sequence alignments

McdB amino acid sequences were obtained from our gene neighborhood analyses previously described for both α- (MacCready *et al.*, 2021) and β-McdBs (MacCready *et al.*, 2020). Multiple sequence alignments (MSAs) were performed using Clustal Omega (Madeira *et al.*, 2022) 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 nongap residues but excluding gap versus gap residues. Sequence logos were created using the above mentioned MSAs via WebLogo (v 2.8.2) (Crooks *et al.*, 2004).

Construct design

All constructs used in this study were generated using Gibson Assembly (Gibson *et al.*, 2009) and were sequence verified upon construction. 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 (Clerico *et al.*, 2007). Fluorescent fusions to proteins of interest were added to the indicated termini with a GSGSGS linker between the two proteins.

Growth and transformation of *Se* strains

All *Se* strains are listed in Supplemental Table S2. Cells 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 h (Clerico *et al.*, 2007). 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.

Growth and transformation of *Hn* strains

All *Hn* strains are listed in Supplemental Table S2. Cells were grown in ATCC Medium 290: S6 medium for Thiobacilli (Hutchinson *et al.*, 1965). 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 3,000 × *g* 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 h 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.

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 × 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).

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 (Schindelin *et al.*, 2012; Ducret *et al.*, 2016). 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. Pearson's correlation coefficients (PCCs) were calculated using ImageJ plugin JaCoP (Bolte and Cordelières, 2006). PCC values were determined for entire fields of view without cell segmentation, using the specified number of replicates. For cellular analysis, statistics were calculated on $n = 3$ biological replicates each with the number of cells or foci specified in Supplemental Table S1. 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).

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 isopropyl β-D-1-thiogalactopyranoside (IPTG) at 1 mM and L-arabinose at 0.2%. Cultures were grown for an additional 4 h, 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 Fisher Scientific]; protease inhibitor tablet [Thermo Fisher Scientific]) and sonicated with cycles of 10 s on, 20 s off at 50% power for 7 min. Lysates were clarified via centrifugation at 15,000 rcf for 30 min. 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 five 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.

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.

CD spectroscopy

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.

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). A total of 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.

Microscopy of protein condensates

Samples for imaging were set up in 16-well CultureWells (Grace Bio-Labs). Wells were passivated by overnight incubation in 5% (wt/vol) Pluronic acid (Thermo Fisher Scientific), and washed thoroughly with the corresponding buffer prior to use. All *Se* McdB samples were incubated for at least 30 min prior to imaging condensates, and all *Hn* McdB samples for at least 18 h 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.

Quantification of phase separation via centrifugation

Centrifugation was used to quantify the degree to which McdB and its variants condensed under certain conditions, as described previously (Alberti *et al.*, 2019). Briefly, 50 µl of sample was incubated at the conditions specified for 30 min, and then centrifuged at 16,000 × *g* for 10 min 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 h and then destaining in water for 14–16 h. 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).

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

All *Se* McdB constructs were expressed as N-terminal mCherry fusions (Shaner *et al.*, 2004) 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 (Shaner *et al.*, 2004) 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 postincubation for 3 h 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 h) or at the time of imaging (t = 3 h). 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.

ACKNOWLEDGMENTS

We would like to thank Dr. JK Nandakumar and Ritvija Agrawal for training and allowing us to use their SEC-MALS system. Dr. Joshua S. MacCready for providing access to previously curated lists of McdB protein sequences used here for the alignments. Claudia Mak for insightful conversations on the molecular nature of carboxysome shells and its contribution to our discussion here. And lastly, Dr. James Bardwell for allowing access to his CD spectrometer. This work was supported by the National Science Foundation to A.G.V. (NSF CAREER Award No. 1941966), Rackham Graduate Student Research Grant to J.L.B., and from research initiation funds provided by the MCDB Department to A.G.V.

REFERENCES

Abrescia NGA, Bamford DH, Grimes JM, Stuart DI (2012). Structure unifies the viral universe. Annu Rev Biochem 81, 795–822.

Alberti S, Gladfelter A, Mittag T (2019). Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. Cell 176, 419–434.

Banani SF, Lee HO, Hyman AA, Rosen MK (2017). Biomolecular condensates: Organizers of cellular biochemistry. Nat Rev Mol Cell Biol 18, 285–298.

Basalla JL, Mak CA, Byrne J, Ghalmi M, Hoang Y, Vecchiarelli AG (2023). Dissecting the phase separation and oligomerization activities of the carboxysome positioning protein McdB. Elife 12, e81362.

Bolte S, Cordelières FP (2006). A guided tour into subcellular colocalization analysis in light microscopy. J Microsc 224, 213–232.

Cai F, Sutter M, Bernstein SL, Kinney JN, Kerfeld CA (2015). Engineering bacterial microcompartment shells: Chimeric shell proteins and chimeric carboxysome shells. ACS Synth Biol 4, 444–453.

Clerico EM, Ditty JL, Golden SS (2007). Specialized techniques for site-directed mutagenesis in cyanobacteria. In: Circadian Rhythms: Methods and Protocols, ed. E. Rosato, New Jersey: Humana Press, 155–171.

Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004). WebLogo: A sequence logo generator. Genome Res 14, 1188–1190.

Das S, Lin YH, Vernon RM, Forman-Kay JD, Chan HS (2020). Comparative roles of charge, π, and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins. Proc Natl Acad Sci USA 117, 28795–28805.

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

Ducret A, Quardokus EM, Brun YV (2016). MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. Nat Microbiol 17(1), 16077.

Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (2006). The Prokaryotes, New York: Springer.

Flamholz AI, Dugan E, Blikstad C, Gleizer S, Ben-Nissan R, Amram S, Antonovsky N, Ravishankar S, Noor E, Bar-Even A, *et al.* (2020). Functional reconstitution of a bacterial CO_2 concentrating mechanism in E. coli. Elife 9, e59882.

Freeman Rosenzweig ES, Xu B, Kuhn Cuellar L, Martinez-Sanchez A, Schaffer M, Strauss M, Cartwright HN, Ronceray P, Plitzko JM, Förster F, *et al.* (2017). The eukaryotic CO2-concentrating organelle is liquid-like and exhibits dynamic reorganization. Cell 171, 148–162.e19.

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

Garcia-Alles LF, Root K, Maveyraud L, Aubry N, Lesniewska E, Mourey L, Zenobi R, Truan G (2019). Occurrence and stability of hetero-hexamer associations formed by β-carboxysome CcmK shell components. PLoS One 14, e0223877.

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6, 343–345.

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

Goedhart J, Von Stetten D, Noirclerc-Savoye M, Lelimousin M, Joosen L, Hink MA, Van Weeren L, Gadella TWJ, Royant A (2012). Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. Nat Commun 31(3), 751.
Gray VE, Hause RJ, Fowler DM (2017). Analysis of large-scale mutagenesis data to assess the impact of single amino acid substitutions. Genetics 207, 53–61.
Hakim P, Hoang Y, Vecchiarelli A (2021). Dissection of the ATPase active site of McdA reveals the sequential steps essential for carboxysome distribution. Mol Biol Cell 32, ar11.
Hill NC, Tay JW, Altus S, Bortz DM, Cameron JC (2020). Life cycle of a cyanobacterial carboxysome. Sci Adv 6, eaba1269.
Hirokawa N, Noda Y, Tanaka Y, Niwa S (2009). Kinesin superfamily motor proteins and intracellular transport. Nat Rev Mol Cell Biol 10, 682–696.
Hoang Y, Azaldegui CA, Dow RE, Ghalmi M, Biteen JS, Vecchiarelli AG (2024). An experimental framework to assess biomolecular condensates in bacteria. Nat Commun 15, 3222.
Hutchinson M, Johnstone KI, White D (1965). The taxonomy of certain Thiobacilli. J Gen Microbiol 41, 357–366.
Kerfeld CA, Aussignargues C, Zarzycki J, Cai F, Sutter M (2018). Bacterial microcompartments. Nat Rev Microbiol 16, 277–290.
Kerfeld CA, Melnicki MR (2016). Assembly, function and evolution of cyanobacterial carboxysomes. Curr Opin Plant Biol 31, 66–75.
Komla-Soukha I, Sureau C (2006). 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.
Krupovic M, Koonin EV (2017). Cellular origin of the viral capsid-like bacterial microcompartments. Biol Direct 12, 25.
Long BM, Förster B, Pulsford SB, Price GD, Badger MR (2021). Rubisco proton production can drive the elevation of CO2 within condensates and carboxysomes. Proc Natl Acad Sci USA 118, e201446118.
Long BM, Hee WY, Sharwood RE, Rae BD, Kaines S, Lim YL, Nguyen ND, Massey B, Bala S, von Caemmerer S, *et al.* (2018). Carboxysome encapsulation of the CO2-fixing enzyme Rubisco in tobacco chloroplasts. Nat Commun 9, 3570.
Lutkenhaus J (2012). The ParA/MinD family puts things in their place. Trends Microbiol 20, 411–418.
MacCready JS, Basalla JL, Vecchiarelli AG (2020). Origin and evolution of carboxysome positioning systems in cyanobacteria. Mol Biol Evol 37, 1434–1451.
Maccready JS, Hakim P, Young EJ, Hu L, Liu J, Osteryoung KW, Vecchiarelli AG, Ducat DC (2018). Protein gradients on the nucleoid position the carbon-fixing organelles of cyanobacteria. Elife 7, e39723.
MacCready JS, Tran L, Basalla JL, Hakim P, Vecchiarelli AG (2021). The McdAB system positions α-carboxysomes in proteobacteria. Mol Microbiol 116, 277–297.
Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A, Lopez R (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res 50, W276–W279.
Mahinthichaichan P, Morris DM, Wang Y, Jensen GJ, Tajkhorshid E (2018). Selective permeability of carboxysome shell pores to anionic molecules. J Phys Chem B 122, 9110–9118.
Melnicki MR, Sutter M, Kerfeld CA (2021). Evolutionary relationships among shell proteins of carboxysomes and metabolosomes. Curr Opin Microbiol 63, 1–9.
Metskas LA, Ortega D, Oltrogge LM, Blikstad C, Lovejoy DR, Laughlin TG, Savage DF, Jensen GJ (2022). Rubisco forms a lattice inside alpha-carboxysomes. Nat Commun 13, 4863.
Niederhuber MJ, Lambert TJ, Yapp C, Silver PA, Polka JK (2017). Superresolution microscopy of the β-carboxysome reveals a homogeneous matrix. Mol Biol Cell 28, 2734–2745.
Oltrogge LM, Chaijarasphong T, Chen AW, Bolin ER, Marqusee S, Savage DF (2020). Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation. Nat Struct Mol Biol 27, 281–287.
Pernigo S, Lamprecht A, Steiner RA, Dodding MP (2013). Structural basis for kinesin-1:cargo recognition. Science 340, 356–359.
Pulianmackal LT, Limcaoco JMI, Ravi K, Yang S, Zhang J, Tran MK, Ghalmi M, O'Meara MJ, Vecchiarelli AG (2023). Multiple ParA/MinD ATPases coordinate the positioning of disparate cargos in a bacterial cell. Nat Commun 14, 3255.
Rae BD, Long BM, Whitehead LF, Förster B, Badger MR, Dean Price G (2013). Cyanobacterial carboxysomes: Microcompartments that facilitate CO2 fixation. J Mol Microbiol Biotechnol 23, 300–307.
Rillema R, Hoang Y, Maccready JS, Vecchiarelli AG (2021). Carboxysome mispositioning alters growth, morphology, and Rubisco level of the cyanobacterium Synechococcus elongatus PCC 7942. mBio 12, e0269620.
Savage DF, Afonso B, Chen AH, Silver PA (2010). Spatially ordered dynamics of the bacterial carbon fixation machinery. Science 327, 1258–1261.
Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, *et al.* (2012). Fiji: An open-source platform for biological-image analysis. Nat Methods 9, 676–682.
Schumacher D, Harms A, Bergeler S, Frey E, Søgaard-Andersen L (2021). PomX, a ParA/MinD ATPase activating protein, is a triple regulator of cell division in Myxococcus xanthus. Elife 10, e66160.
Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE, Tsien RY (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat Biotechnol 22, 1567–1572.
Shaner NC, Lambert GG, Chammas A, Ni Y, Cranfill PJ, Baird MA, Sell BR, Allen JR, Day RN, Israelsson M, *et al.* (2013). A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. Nat Methods 10(5), 407–409.
Skoging U, Liljestro P (1998). Role of the C-terminal tryptophan residue for the structure-function of the alphavirus capsid protein. J Mol Biol 279, 865–872.
Sun Y, Wollman AJM, Huang F, Leake MC, Liu LN (2019). Single-organelle quantification reveals stoichiometric and structural variability of carboxysomes dependent on the environment. Plant Cell 31, 1648–1664.
Sutter M, Laughlin TG, Sloan NB, Serwas D, Davies KM, Kerfeld CA (2019). Structure of a synthetic β-carboxysome shell. Plant Physiol 181, 1050–1058.
Sutter M, Melnicki MR, Schulz F, Woyke T, Kerfeld CA (2021). A catalog of the diversity and ubiquity of bacterial microcompartments. Nat Commun 12, 3809.
Tan YQ, Ali S, Xue B, Teo WZ, Ling LH, Go MK, Lv H, Robinson RC, Narita A, Yew WS (2021). Structure of a minimal α-carboxysome-derived shell and its utility in enzyme stabilization. Biomacromolecules 22, 4095–4109.
Tsuboi M, Overman SA, Nakamura K, Rodriguez-Casado A, Thomas GJ (2003). Orientation and interactions of an essential tryptophan (Trp-38) in the capsid subunit of Pf3 filamentous virus. Biophys J 84, 1969–1976.
Vecchiarelli AG, Mizuuchi K, Funnell BE (2012). Surfing biological surfaces: exploiting the nucleoid for partition and transport in bacteria. Mol Microbiol 86, 513–523.
Wang H, Yan X, Aigner H, Bracher A, Nguyen ND, Hee WY, Long BM, Price GD, Hartl FU, Hayer-Hartl M (2019). Rubisco condensate formation by CcmM in β-carboxysome biogenesis. Nature 566, 131–135.
Wang J, Choi JM, Holehouse AS, Lee HO, Zhang X, Jahnel M, Maharana S, Lemaitre R, Pozniakovsky A, Drechsel D, *et al.* (2018). A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. Cell 174, 688–699.e16.
Wimley WC, White SH (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat Struct Biol 3, 842–848.
Yeates TO, Kerfeld CA, Heinhorst S, Cannon GC, Shively JM (2008). Protein-based organelles in bacteria: Carboxysomes and related microcompartments. Nat Rev Microbiol 6, 681–691.
Yoo H, Triandafillou C, Drummond DA (2019). Cellular sensing by phase separation: Using the process, not just the products. J Biol Chem 294, 7151–7159.
Zang K, Wang H, Hartl FU, Hayer-Hartl M (2021). Scaffolding protein CcmM directs multiprotein phase separation in β-carboxysome biogenesis. Nat Struct Mol Biol 28, 909–922.