Dissecting the phase separation and oligomerization activities of the carboxysome positioning protein McdB

Joseph L. Basalla[1], Claudia A. Mak[2], Miguel J. Limcaoco[3], Y Hoang[1], & Anthony G. Vecchiarelli[1#]

Affiliations

[1] Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI, 48109, USA.

[2] Department of Biological Chemistry, University of Michigan, Ann Arbor, MI, 48109, USA.

[3] Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI, 48109, USA.

[#]Correspondence

ave@umich.edu

ABSTRACT

Carboxysomes are protein-based organelles essential for efficient CO_2-fixation in cyanobacteria and some chemoautotrophic bacteria. Understanding carboxysome homeostasis has implications for both microbial physiology and engineering CO_2-fixing organisms. We recently identified the two-component system that spatially regulates carboxysomes, consisting of the proteins McdA and McdB. McdA is a member of the ParA/MinD-family of ATPases which position various structures across bacteria. McdB, however, represents a widespread but unstudied class of proteins. We previously found that McdB forms a hexamer and undergoes robust phase separation *in vitro*, but the sequence and structural determinants underlying these properties are unknown. Here, we define the domain architecture of McdB from the model cyanobacterium *S. elongatus* PCC 7942. We identify an N-terminal Intrinsically Disordered Region (IDR) that modulates condensate solubility, a central glutamine-rich dimerizing domain that drives phase separation, and a C-terminal domain that trimerizes McdB dimers. We also identify critical basic residues in the IDR which we mutate to fine-tune condensate solubility both *in vitro* and *in vivo*. Finally, we provide *in silico* evidence suggesting the N-terminus of McdB acts as a MoRF, folding upon interaction with McdA. The data advance our understanding and application of carboxysome homeostasis and the molecular grammar governing protein phase separation.

INTRODUCTION

Compartmentalization is a fundamental feature by which cells regulate their metabolism. Although lacking extensive lipid-membrane systems, recent reports have shown that proteinaceous bacterial microcompartments (BMCs) are a widespread strategy for compartmentalization in bacteria (1, 2). The best studied example of a 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 drastically increase the efficiency of carbon-fixation. Efforts to express functional carboxysomes in heterologous hosts to allow them to efficiently capture carbon are ongoing (4, 5), but gaps in our understanding on carboxysome homeostasis limit these efforts.

An important aspect of carboxysome homeostasis is the nature of carboxysome spatial regulation (6). Recently, the two-component system responsible for spatially regulating carboxysomes was characterized, which we named the maintenance of carboxysome distribution (Mcd) system (7, 8, 9). Both components of the McdAB system (i.e. both McdA and McdB) are necessary to prevent aggregation of carboxysomes and ensure their equal inheritance upon cell division (7, 9). Often in bacteria, subcellular structures are spatially regulated by the activity of ParA/MinD-family ATPases (10, 11). Generally, these ParA/MinD-family ATPases use the energy of ATP hydrolysis for positioning, and an adaptor protein mediates the interaction between the ATPase and the cargo being positioned (10, 11). This family positions a variety of genetic and proteinaceous cargos over specific cellular surfaces, with ParA-like ATPases using the nucleoid as a positioning matrix and MinD-like ATPases using the membrane (10, 11). We and others have shown that McdA is a ParA-like ATPase that uses the nucleoid to distribute carboxysomes down the cell length (6, 7, 12). However, we have only recently identified McdB

as the adaptor protein that connects McdA to the carboxysome (7, 9). Since identifying McdB as a necessary component for carboxysome spatial regulation, we have collected evidence suggesting that McdAB-like systems are also responsible for spatially regulating a variety of other BMCs (9). Thus, McdB represents a new class of uncharacterized adaptor proteins responsible for the spatial regulation of carboxysomes and possibly a variety of other BMCs. A biochemical understanding of partner proteins connecting ParA/MinD-family ATPases to their respective cargoes has been critical to understanding the spatial regulation of several fundamental processes in bacteria, including chromosome segregation (ParA and its partner protein ParB) and cell division positioning (MinD and its partner protein MinE) (13). However, little is known about McdB biochemistry.

From our initial studies in the model cyanobacterium *Synechococcus elongatus* PCC 7942 (*Se7942*), we found that McdB self-associates *in vitro* to form both a stable hexamer (9) and liquid-like condensates (8). However, the domain architecture and the regions required for McdB oligomerization and phase separation are unknown. It is also unknown how these two modes of McdB self-association influence each other. Briefly, phase separation of proteins in solution is a process that can occur under a specific set of conditions where protein-protein interactions are more favorable than protein-solvent interactions (14). In both eukaryotes and prokaryotes, phase separation has been demonstrated as an underlying self-organizing mechanism that allows for the spatiotemporal concentration of certain biomolecules under specific environmental conditions (15, 16). How McdB is recruited to carboxysomes has yet to be determined, but its phase separation activity has been suggested to be a contributing factor (17). Furthermore, characterizing the underlying chemistries for protein phase separation has led

to the development of phase separation as a synthetic tool to engineer cytoplasmic organization (18, 19, 20, 21).

In this report, we define a domain architecture of *Se7942* McdB and identify which domains contribute to oligomerization and phase separation, and discover an interplay between these two modes of self-association. We further create a series of point mutations that allow us to fine-tune the phase separation activity of McdB. Lastly, we use energy-based simulations to gain insight into the structure of the McdB-McdA interaction. Our findings will allow future investigations on how phase separation contributes to the functionality of McdB *in vivo*. Additionally, the ability to fine-tune the phase separation of McdB could have implications for its use as a tool for synthetic biology applications (19).

RESULTS

Structural predictions generate a low confidence α-helical model for Se7942 McdB

We first wanted to collect structural information for *Se7942* McdB and set out to determine its crystal structure. However, *Se7942* McdB displayed robust phase separation across a range of buffer conditions, making crystal trials thus far unsuccessful (Fig. S1). We next turned to I-TASSER (Iterative Threading ASSEmbly Refinement) (22, 23) to generate structural models for *Se7942* McdB. I-TASSER predicted *Se7942* McdB secondary structure to be predominantly α-helical with a largely disordered N-terminus (Fig. S2A). 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 (Fig. S2B). But ultimately, these top 10 PDB matches identified by I-TASSER align poorly with *Se7942* McdB, with each alignment showing low sequence identity (< 20% on average) and low-quality scores (Z-scores < 1 on average) (Fig. S2C). As a result, the top three final models generated by I-TASSER all have poor confidence scores (Fig. S2B). 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, 24). Together, these data provide low-confidence structural predictions for *Se7942* McdB. We therefore set out to validate these predictions with empirical approaches.

Defining the domain architecture of Se7942 McdB

We used circular dichroism (CD) to characterize the secondary structure of *Se7942* McdB. The spectrum showed a characteristic α-helical signature that remained stable even after incubation at 80°C (Fig. 1A), indicative of a stabilized coiled-coil (25). This is consistent with the helical predictions from I-TASSER, and with our previous bioinformatics data that predicted glutamine-rich, coiled-coil domains to be conserved across all cyanobacterial McdB homologs (8). We next sought to empirically identify these folded domains using limited proteolysis with trypsin (26, 27, 28) (Fig. 1B). Trypsin cuts at arginines and lysines, which are frequent throughout *Se7942* McdB, such that the largest fragment between any two basic residues is ~3 kDa (Fig. 1C). Therefore, any stably folded regions that are protected from trypsin would become apparent and could be resolved via this approach. The digestion yielded three major bands that showed stability over time, which we labeled A, B, and C (Fig. 1B). Band C was most stable, representing an ~11 kDa fragment that remained undigested for 12 hours. This strong

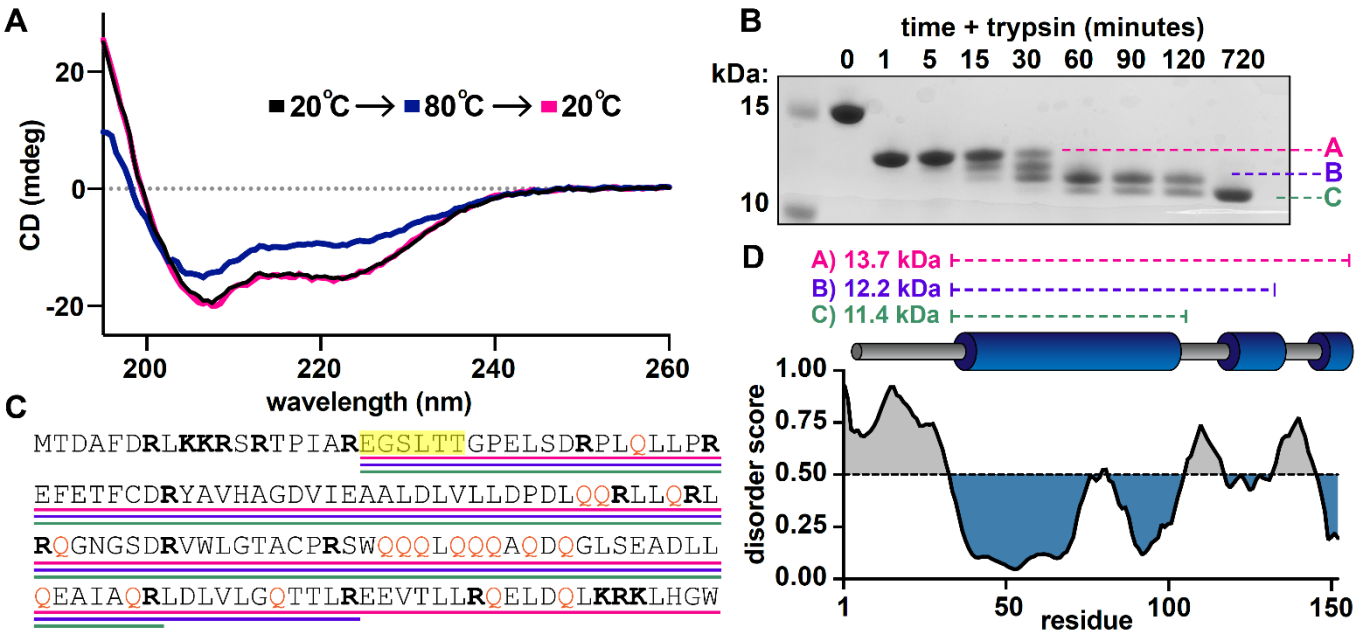


Figure 1: **Defining a domain architecture of *Se7942* McdB**. (**A**) Circular Dichroism (CD) spectra of McdB at 20ºC (black), 80ºC (blue), and then returned to 20ºC (magenta). Spectra show α-helical structure resilient to heat denaturation. (**B**) SDS-PAGE analysis of trypsin-digested McdB sampled 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 and Gln-Q in orange. Regions corresponding to bands A, B, and C from panel B are underlined in magenta, blue, and green, respectively. Amino acids determined through N-terminal sequencing of bands A, B, and C are highlighted yellow. (**D**) Structural model of *Se7942* 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.

protection against trypsin is consistent with our CD data, which showed high resilience to heat denaturation (see Fig. 1A). 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 (Fig. 1C). 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 (Fig. 1B, C). By combining the N-terminal sequencing results (Fig. 1C), the molecular weights of the three protected regions (Fig. 1B), the locations of all arginines and lysines (Fig. 1C), and the predicted disorder via PONDR VLXT

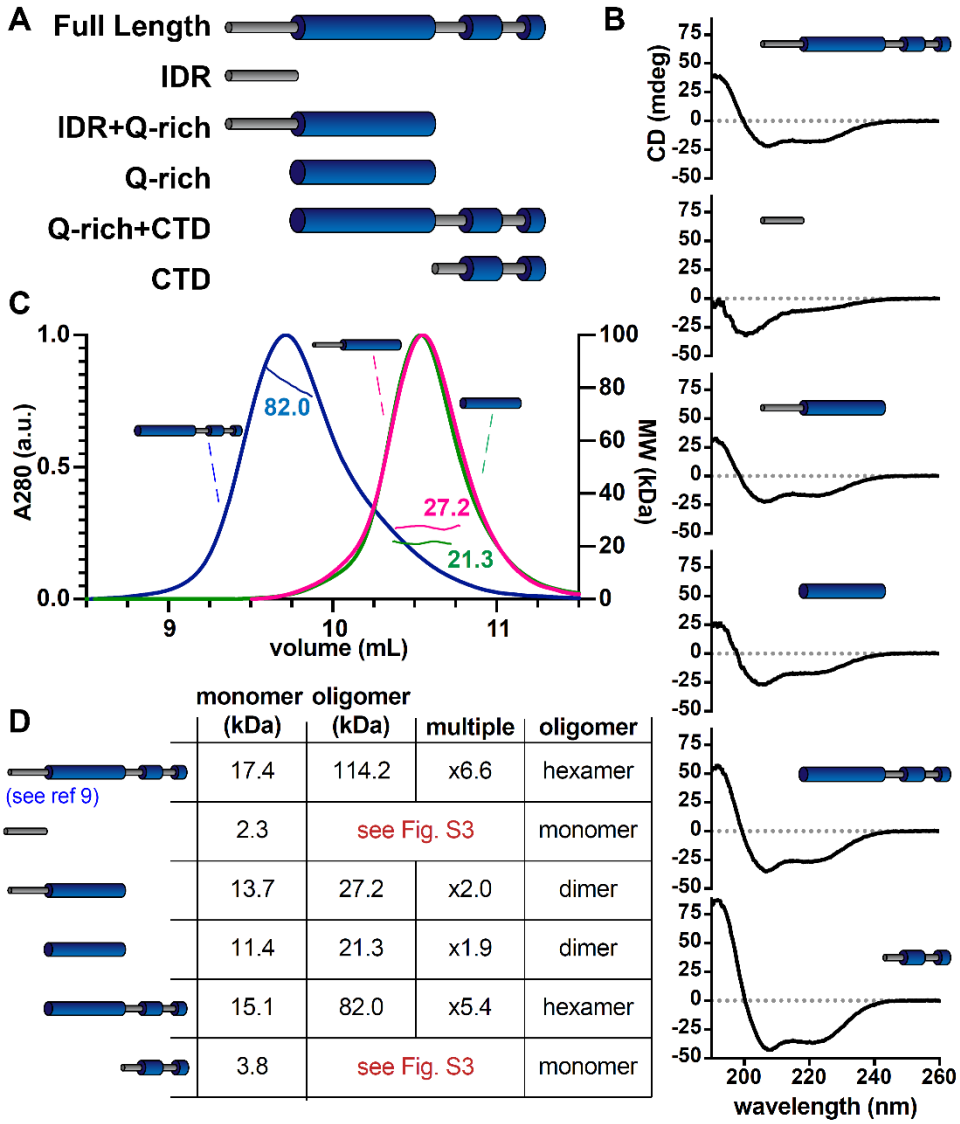
(29) (Fig. 1D), we developed a model for the domain architecture of *Se7942* McdB that was consistent with the I-TASSER predictions (Fig. 1D).

Se7942 McdB forms a trimer-of-dimers hexamer

From our structural model, we defined three major domains of *Se7942* McdB: 1) an intrinsically disordered region (IDR) at the N-terminus, 2) a highly stable glutamine-rich central domain (Q-rich), corresponding to band C (Fig. 1C), and 3) a C-terminal domain (CTD) with two short helical regions. We then used this model to design a series of truncation mutants, including each of these domains alone as well as the Q-rich domain with either the N-terminal IDR or the CTD (Fig. 2A). CD spectra of these truncations showed that the N-terminus was indeed disordered on its own, and both the Q-rich domain and CTD maintained α-helical signatures (Fig. 2B).

Having previously shown that full-length *Se7942* McdB forms a hexamer in solution (9), we next aimed to use 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 (Fig. S3A), we saw that only the Q-rich+CTD construct eluted at a volume similar to the full-length hexamer (Fig. S3B). Furthermore, the Q-rich domain alone, or with the IDR, appeared to elute between the expected monomer and hexamer peaks, suggesting that the Q-rich 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



monomer (kDa)	oligomer (kDa)	multiple	oligomer
17.4	114.2	x6.6	hexamer
2.3	see Fig. S3		monomer
13.7	27.2	x2.0	dimer
11.4	21.3	x1.9	dimer
15.1	82.0	x5.4	hexamer
3.8	see Fig. S3		monomer

Figure 2: **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 for the indicated McdB truncations 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 Fig. S3)**. (**D**) Summary of the oligomeric states of full-length McdB and truncation mutants.

constructs that appeared to oligomerize during SEC (Fig. 2C). We found that the Q-rich+CTD truncation was indeed hexameric, while the Q-rich domain, with or without the IDR, formed a dimer (Fig. 2D). These data suggest that the stable Q-rich domain of McdB contains a dimerization interface, while the CTD subsequently allows for trimerization of dimers. Therefore, we conclude that full-length *Se7942* McdB forms a hexamer as a trimer-of-dimers.

The Q-rich domain drives phase separation, the IDR modulates solubility, and the CTD further oligomerizes McdB to achieve full-length phase separation activity

We have previously shown that McdBs from *Se7942* and other cyanobacteria can form liquid-like condensates *in vitro* (8, 9). We therefore used our truncated mutants to determine which domains of *Se7942* McdB contribute to phase separation. Interestingly, all McdB truncations remained soluble at the buffer conditions that elicited phase separation of the full-length protein (Fig. 3A). This finding was unexpected given the wealth of data showing that single domains, often IDRs, of several phase-separating proteins are sufficient to drive phase separation (30, 31, 32, 33, 34). Our data show that no single domain of McdB is sufficient for full-length level phase separation activity. Rather, all domains, including those that are stably folded and provide oligomerization, must contribute to phase separation.

For some proteins, IDRs undergo phase separation only when fused to an oligomerizing domain (35). In these cases, oligomerization locally increases the effective IDR concentration, thus seeding and driving phase separation (35). When we added a crowding agent to increase the local concentration of our McdB truncations, the IDR alone was again insufficient for phase separation, even at concentrations up to 4 mM (Fig. 3B). Intriguingly, we did observe phase separation for all truncations containing the stably folded Q-rich domain (Fig. 3B). In fact, the Q-rich domain alone was found to be necessary and sufficient for driving phase separation, albeit in the presence of crowder and at higher concentrations.

Although the IDR was not required for McdB phase separation activity in the presence of crowder, the addition of the IDR to the Q-rich domain did produce condensates with size and morphology similar to that of the full-length protein (Fig. 3B). The data suggest that, while the Q-rich domain is necessary for phase separation, the IDR modulates the solubility of the

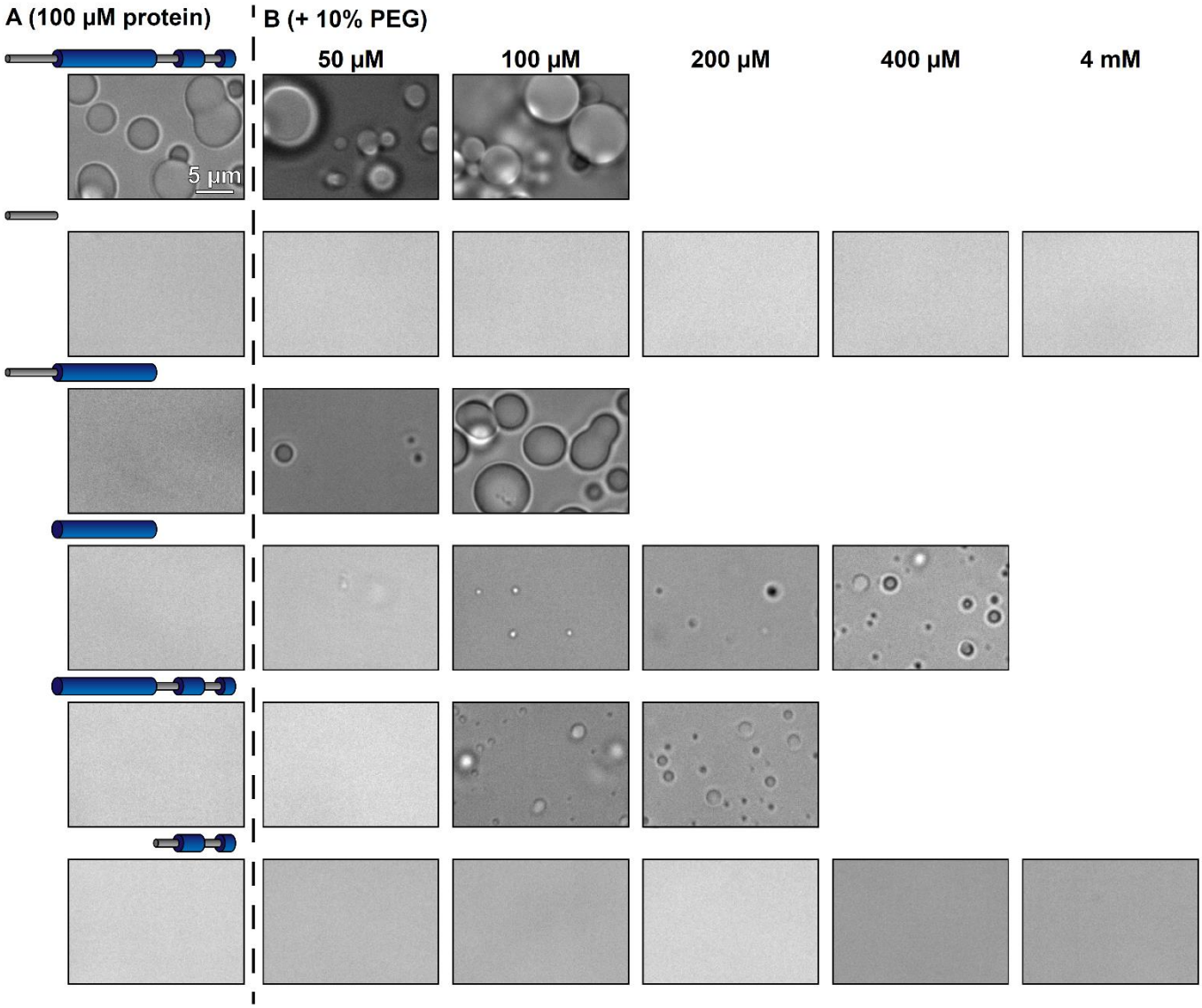


Figure 3: The Q-rich domain of McdB drives phase separation while the N-terminal IDR modulates solubility. (**A**) Representative DIC microscopy images of full-length and truncation mutants of McdB at 100 µM in 150 mM KCl and 20 mM HEPES pH 7.2 (**B**) As in (A), but with increasing protein concentration (as indicated) and with the addition of 10% PEG-8000. Scale bar: 5 µm for all images.

condensates, as has been described for other systems (19, 36). The CTD, on the other hand, had no effect on the Q-rich-domain condensates; Q-rich+CTD condensates formed at the same protein concentration and with identical droplet morphologies as the Q-rich domain alone. Also, the CTD alone was insufficient for phase separation at concentrations up to 4 mM, even in the presence of a crowder. Together, the findings lead us to a model where the Q-rich domain of

McdB drives its phase separation activity, the IDR modulates condensate solubility, and the CTD allows McdB dimers to trimerize thus increasing local concentrations. All three domains are therefore required for the robust phase separation activity observed for the full-length protein *in vitro*.

McdB phase separation is pH-sensitive and influenced by basic residues

We next sought to determine which residues facilitate McdB phase separation. Using turbidity as a measure of phase separation, we performed multi-dimensional phase diagrams across a range of protein concentrations, salt concentrations, and pH to determine which conditions favored protein-protein or protein-solvent interactions; an approach that has previously been used to gain insight into the critical residues for phase separation (37, 14). Over all concentrations of McdB tested, we saw that turbidity decreased at higher KCl concentrations (Fig. S4). This finding shows that McdB condensates are solubilized in solutions with higher ionic strength, and implicates electrostatic interactions. We also found that turbidity decreased at higher pH (Fig. S4), which shows that an increase in the solvation of basic residues specifically contributes to solubilization of McdB condensates.

To look at the effects of salt and pH on McdB phase separation in more detail, we used centrifugation to quantify the amount of McdB in the dense versus light phases as has been described previously (14). We used this approach across KCl and pH titrations while keeping McdB concentration constant. We saw that pH change had a significantly greater impact on McdB phase separation than changes in salt concentration. For instance, an increase in KCl from 50 mM to 1000 mM resulted in a ~20% decrease in the dense phase (Fig. 4A), whereas a change in pH from 6.3 to 8.6 resulted in nearly complete solubilization of condensates (Fig. 4B).

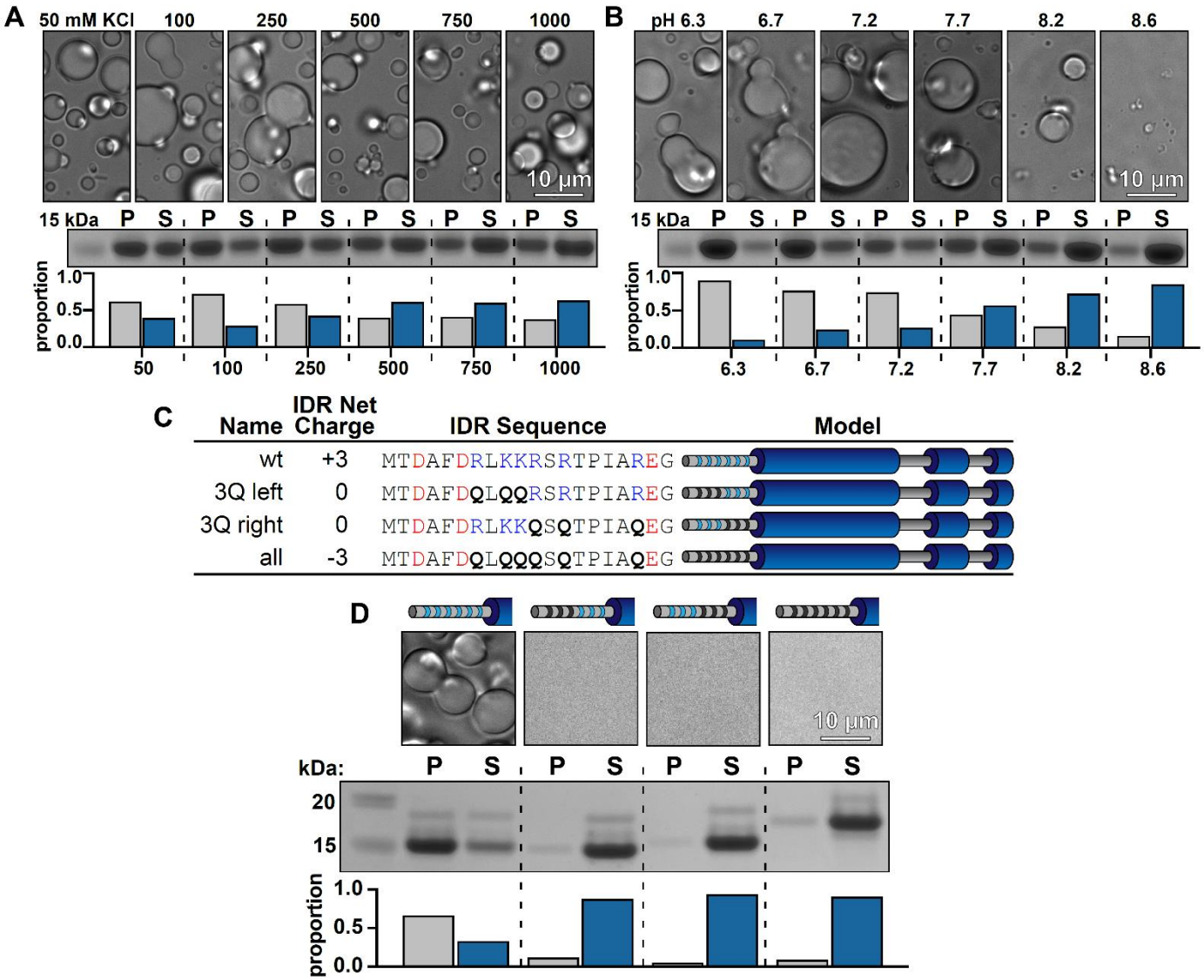


Figure 4: McdB phase separation is highly sensitive to pH and is influenced by basic residues in the N-terminal IDR. (**A**) Representative DIC microscopy images of 50 µM McdB in 20 mM HEPES pH 7.2 and increasing KCl concentration (*top*). McdB condensates were pelleted (P) and run on an SDS-PAGE gel along with the associated supernatant (S) (*middle*). Pellet and Supernatant band intensities were then quantified (*bottom*). Scale bar applies to all images. (**B**) As in (A), except salt was held constant at 100 mM KCl and the pH was increased as indicated. (**C**) Table showing the net charge and N-terminal IDR sequence of wildtype 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. (**D**) Representative DIC microscopy images for wild-type and the glutamine-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*). Pellet and Supernatant band intensities were then quantified (*bottom*).

204

Together, this leads us to conclude that basic residues specifically play a key role in the transition of McdB from a soluble to condensed phase.

The net charge of the IDR mediates solvation of Se7942 McdB condensates

Based on our findings that the IDR modulates the solubility of McdB condensates (see Fig. 3B) and that basic residues are critical for this process (Fig. 4B), we next performed glutamine-substitution mutagenesis targeting all six basic (Arg and Lys) residues in the IDR. Substituting basic residues with glutamines allows us to remove the positive charge while maintaining hydrophilicity. This is important because drastic changes to hydrophilicity, for example via alanine-substitution of basic residues, have been shown to result in aggregation of phase separating proteins (38). We first made two constructs that neutralized the net charge of the IDR, with one mutant having the first three basic residues substituted, and the other having the latter three substituted (Fig. 4C). We also made a mutant where all six basic residues in the IDR were substituted to glutamines, leaving a net charge of -3. Interestingly, none of these mutants underwent phase separation in buffer conditions that sufficed for wild-type McdB (Fig. 4D). This result is striking because it shows clearly that no specific residue is itself critical for phase separation, but instead the net charge of the IDR is what mediates the solubility of McdB condensates. Importantly, all glutamine mutants had CD spectra similar to that of wild-type McdB (Fig. S5), showing that the mutants still fold properly and solubilization is not due to any gross structural changes. Taken together, the data suggests that while the IDR is not necessary or sufficient for driving phase separation, its charge content mediates the transition between McdB self- and solvent- interactions, thus modulating the solubility of condensates.

McdB phase separation is tunable through changes to the charge content of the IDR

It was unexpected that mutating only three basic residues in the IDR was sufficient for complete solubilization of *Se7942* McdB condensates. We set out to determine if fewer mutations, which would have a lesser effect on IDR net charge, could elicit the same loss of phase separation. Therefore, we generated several mutants in which pairs of basic residues in the IDR were substituted with glutamines, leaving an IDR net charge of +1 (Fig. 5A). All of these +1 mutants still showed phase separation activity under standard buffer conditions used for the wild-type, albeit with resultant condensates that were smaller and fewer than wild-type (Fig. 5B). This further showed that IDR net charge underlies condensate solubility, and suggested that the triplet mutations with an IDR net charge of 0 may still be capable of phase separation at higher protein concentrations. Indeed, when we added a crowding agent, we saw that the IDR net-charge 0 mutants could still undergo phase separation (Fig. 5C). Moreover, this revealed a gradual shift of the proportion of McdB in the dense versus light phases as we incrementally changed the net charge of the IDR. Furthermore, we saw that position of these mutations had slight differences from one another, where not all +1 or 0 net charge mutants phase separated to the same extent (Fig. 5C). However, the mutant with no positive charges in the IDR (net charge -3) remained completely soluble even with crowder present (Fig. 5C). We therefore conclude that *Se7942* McdB phase separation can be modulated by adjusting the charge content of the IDR.

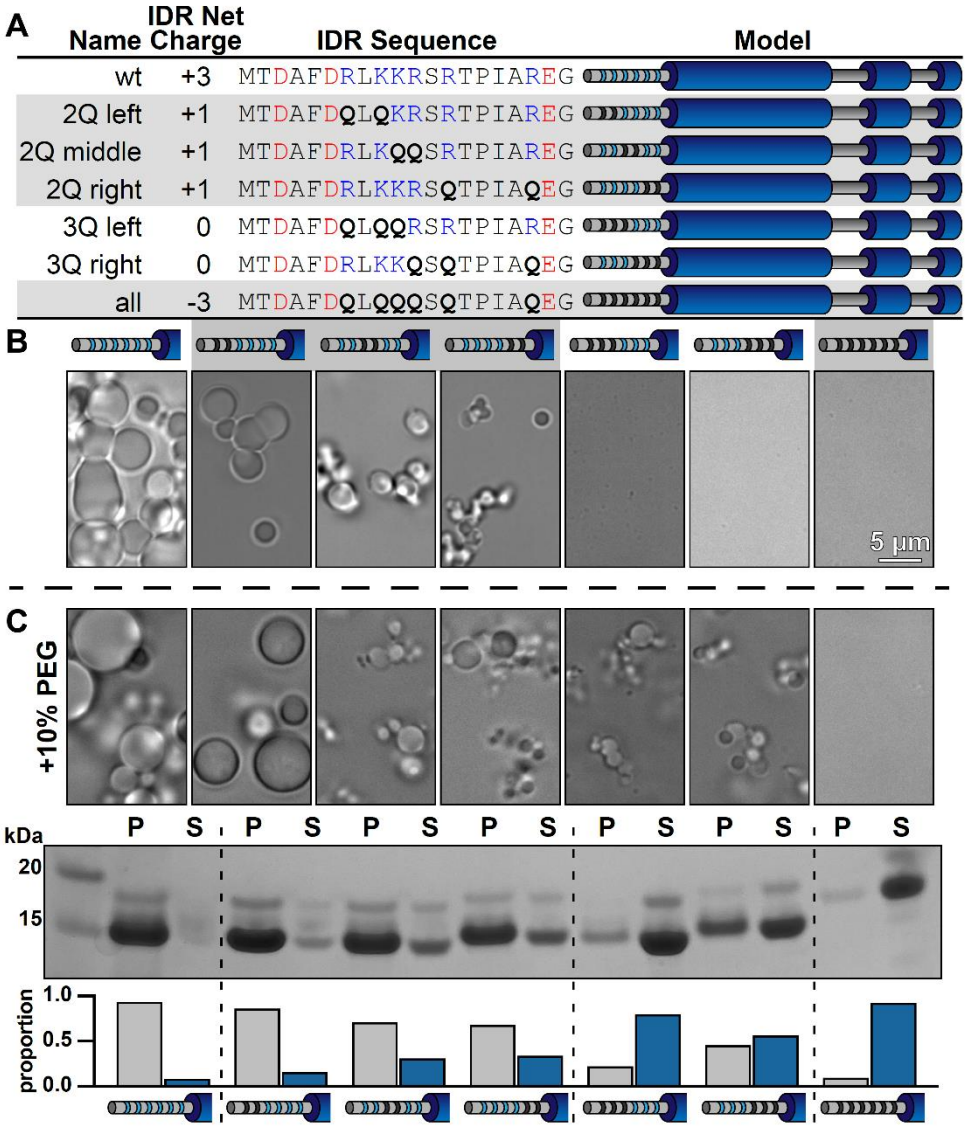


Figure 5: Net charge of the IDR tunes solubility of McdB condensates. (**A**) Table showing the net charge and N-terminal IDR sequence of wildtype McdB compared to the glutamine-substitution mutants. Acidic and basic residues in the IDR are colored red and blue, respectively. Glutamine-substitutions 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. (**B**) Representative DIC microscopy images for wild-type and the glutamine-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*). Pellet and Supernatant band intensities were then quantified (*bottom*).

Net charge of the IDR can be used to tune solubility of Se7942 McdB in E. coli

As a proof of principle on whether the IDR could also be used to tune McdB solubility within cells, we expressed fluorescent fusions of mCherry with both wild-type McdB and the full

glutamine-substitution mutant (with IDR net charge of -3) in *E. coli* MG1655. As the concentration of protein increased over time, wild-type McdB adopted two observable states: dense polar foci with high intensity, and a dilute cytoplasmic phase (Fig 6A). This was intriguingly similar to the dense and dilute phases we observed for McdB *in vitro* (e.g. see Fig. 4). After 3 hours of expression, nearly 70% of cells with wild-type McdB adopted this two-state regime (Fig. 6B). The foci were indeed driven by McdB, as the mCherry tag alone remained diffuse (Fig. 6A, B). The IDR mutant, on the other hand, was considerably more soluble than wild-type *in vivo*, where even after 3 hours of expression only ~10% of cells contained foci (Fig. 6A, B). Monitoring expression over time showed that this change in solubility was not due to significant differences in protein levels or due to cleavage of the fluorescent tag (Fig. 6C), 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 be used to tune McdB solubility *in vivo*.

McdA may stabilize the folding of an α-helix in the N-terminal IDR of McdB

Above, we empirically showed the presence of an N-terminal IDR that is highly susceptible to protease and gives disordered signature via CD. Intriguingly however, I-TASSER predicted that this N-terminal region encodes a short helix (see Fig. S2). A growing body of work describes what are known as molecular recognition features (MoRFs) (39, 40). MoRFs are functional regions of IDRs that interact with a structured partner and adopt a structured conformation upon interaction. Therefore, we hypothesized that the discrepancy between our empirical data and the I-TASSER predictions may represent a MoRF in the N-terminus of McdB that folds upon interaction with the McdA ATPase dimer.

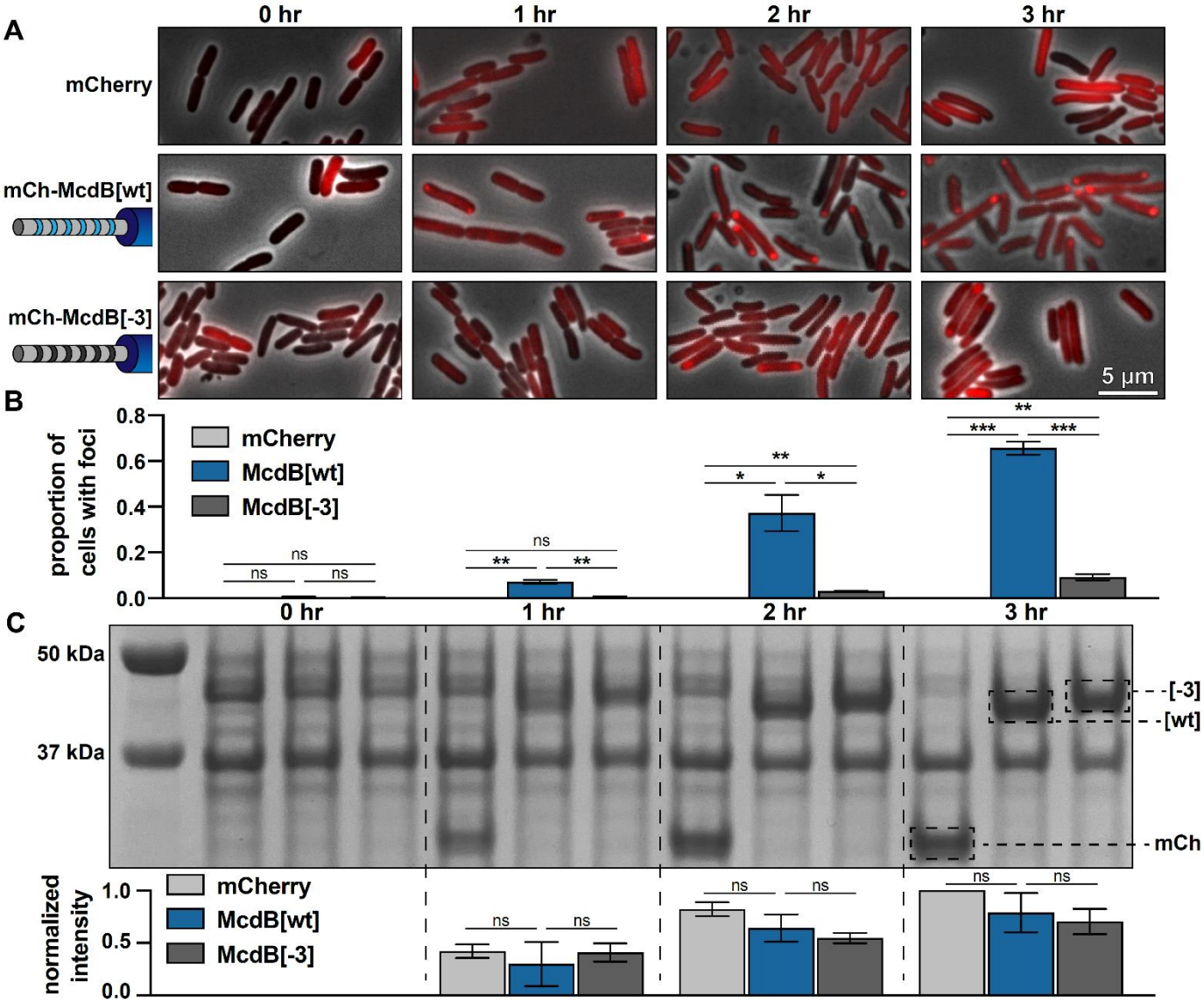


Figure 6: Net charge of the IDR affects McdB solubility *in vivo*. (**A**) Representative fluorescence microscopy images monitoring the expression of the indicated constructs over time. Scale bar applies to all images. (**B**) Quantification of the proportion of cells containing foci from the images represented in (A). All quantifications were done on > 300 cells and from n = 3 technical replicates. Reported values represent means with standard deviations. *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 standard deviations. Data were analyzed via Welch's *t* test.

To gain insight into this, we first used Rosetta to generate an ensemble of structural models for the N-terminus of *Se7942* McdB alone. Using the energies from these models, we predicted disorder propensity for the N-terminus of McdB via Rosetta ResidueDisorder (41).

This approach predicted the first 20 amino acids of McdB to be disordered (Fig. 7A), consistent with our empirical data and the PONDR scores. We then compared these results to simulations of the N-terminal region of McdB docking onto the McdA dimer. To do this, we first used AlphaFold2 (AF2) (42, 43) to generate a structural model of the *Se7942* McdA dimer (Fig. S6A). The AF2 model strongly overlaps with the crystal structure for an McdA homolog from the cyanobacterium *Cyanothece* PCC7427 (Fig. S6B; Fig. 7B) (44). We next generated an ensemble of 10,000 docking simulations in Rosetta between the N-terminus of McdB and the AF2 McdA dimer (Fig. S6C). The top 50 most energetically favorable models had notably lower energy scores compared to the rest of the ensemble (Fig. S6D). Extracting secondary structure from these top 50 models showed that the N-terminus of McdB was no longer fully disordered, but instead contained α-helical content often within the first 10 amino acids (Table S1), overlapping with the I-TASSER secondary structure prediction (see Fig. S2A). This suggests that while the N-terminus of McdB is disordered on its own, the formation of an α-helix may be the more stable conformation upon its interaction with McdA, representative of a MoRF within this region of McdB.

Although the N-termini of McdB adopted α-helices upon docking, their orientation on the McdA dimer varied, as demonstrated in the top four most energetically favorable models (Fig. 7C). However, overlaying the McdB N-termini from these models revealed that the orientation of certain basic residues was conserved in space (Fig. S6E). Looking more closely at the interactions between McdB and McdA in these models showed that several salt bridges form between basic residues in McdB and acidic residues in McdA (Fig. 7D). Taken together, the results suggest that the N-terminus of McdB folds into an α-helix upon interaction with the McdA dimer, and the helix becomes stabilized in part by salt bridges between the two proteins.

Consideration of the N-terminus of McdB as a MoRF explains the discrepancy between our empirical data and I-TASSER predictions of its secondary structure.

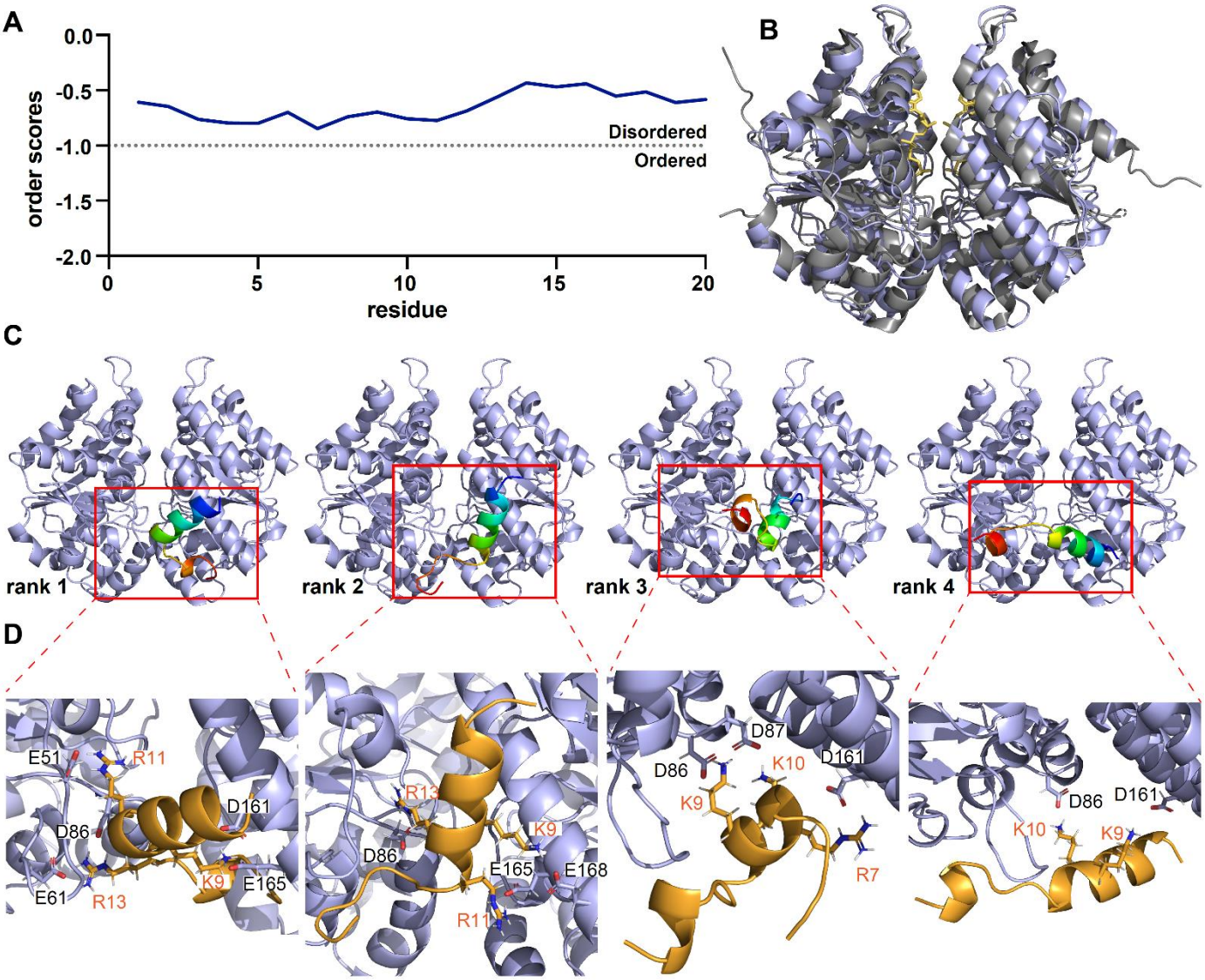


Figure 7: Evidence for a MoRF in the N-terminus of *Se7942* McdB. (**A**) Disorder prediction of the N-terminal 20 amino acids of *Se7942* McdB using Rosetta ResidueDisorder (41). Values with order scores greater than -1.0 are predicted to be disordered. (**B**) Overlay between the homodimer of *Se7942* McdA generated using AF2 (purple) and the crystal structure of *Cyanothece* McdA[D38A]-ATP (grey; PDB 6NOP) (44). ATP molecule is colored yellow. (**C**) Representative structures of the N-terminus of *Se7942* McdB (spectral) docked onto the homodimer of *Se7942* McdA (purple). Models represent the top 4 most energetically favorable ranks out of an ensemble of 10,000 generated through Rosetta. For the McdB N-termini, amino end is colored blue and carboxyl end is colored red. (**D**) Zoomed in representation of (C) with McdB N-termini colored orange. Residues involved in stabilizing salt bridges are shown and labeled, with residues from the McdB N-termini labeled in orange.

DISCUSSION

In this report, we generate a structural model of *Se7942* McdB based on several empirical and predictive approaches. We define a tripartite domain architecture with an N-terminal IDR, a highly stable α-helical Q-rich domain, and a CTD consisting of several smaller α-helices (Fig. 1). Using this model, we created truncations to dissect which domains were required for full-length McdB to form a hexamer. We found that the Q-rich domain provides a dimerization interface, and the CTD trimerizes the dimer (Fig. 2). We further used these truncations to determine which regions of McdB are required for its robust phase separation activity *in vitro*. Intriguingly, the Q-rich dimerization domain drove phase separation, whereas the N-terminal IDR, while not required for phase separation, modulated solvation of the condensates. The CTD was also not required for phase separation, but its trimerization of McdB dimers was needed to achieve wild-type levels of McdB phase separation (Fig. 3). We then performed multi-dimensional buffer screens and quantified the degree of phase separation across solution conditions to identify the importance of basic residues in mediating condensate solubility. Together, this information led us to show that positive charge, specifically in the IDR, is critical for mediating protein-protein versus protein-solvent interactions for McdB (Fig. 4). These findings allowed us to design a series of McdB mutants, where we used the net charge of the IDR to tune McdB solubility both *in vitro* (Fig. 5) and *in vivo* (Fig. 6). Lastly, we provide evidence that the N-terminal IDR may be a MoRF, forming an α-helix upon interaction with McdA (Fig. 7). In summary, we determined McdB domain architecture, its oligomerization domains, the regions required for phase separation, and how to fine-tune its phase separation activity (Fig. 8). The findings will allow us to now probe the biological significance of these various modes of McdB self-association in the

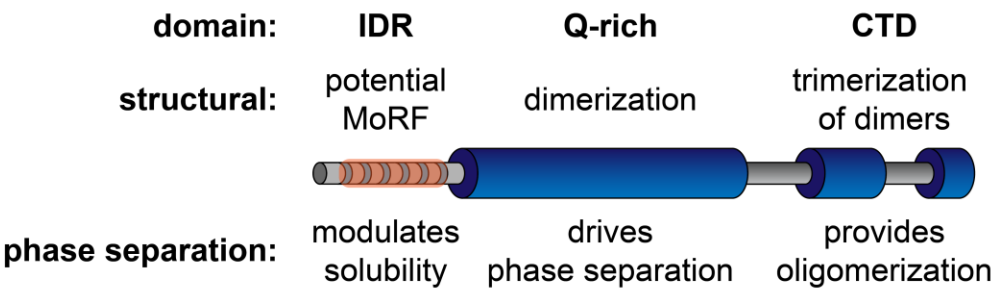


Figure 8: Proposed domain structure and model of *Se7942* McdB self-association. The central Q-rich domain (long blue bar) is necessary and sufficient for dimerization and driving phase separation. The α-helical domains of the C-terminus (short blue bars) trimerize McdB dimers which influences the propensity for phase separation. While the N-terminal IDR (long grey bar) is not required for phase separation activity, its basic residues (blue stripes) modulate the solvation of McdB condensates. A putative MoRF within the N-terminus of the IDR (orange bar) is predicted to fold into an α-helix upon interaction with McdA.

carboxysome positioning reaction with McdA, as well as utilize our ability to tune its phase separation activity in heterologous systems for synthetic biology applications.

Glutamine rich domains and their involvement in protein-protein interactions and phase separation

We have previously shown that all McdBs from β-cyanobacteria have central Q-rich regions (8). Here, we show that the central Q-rich domain of *Se7942* McdB forms a stable α-helical dimer that is both necessary and sufficient for driving phase separation, albeit to a lesser extent than the full-length protein. The ability of Q-rich regions to form both stable structures and liquid-like assemblies has been seen for other proteins, and an appreciation of the interplay between these two modes of interaction is developing. For instance, Q-rich domains are involved in several stable protein-protein interactions, such as in amyloids and coiled-coils (25, 45), and in some instances can lead to the formation of protein aggregates and disease (46, 47). Extreme examples include Huntington's disease and spinocerebellar ataxia 1 (SCA1), where expansion of polyQ tracts in the disease-related proteins leads to aggregate formation and disease onset (48, 49). More recently, aggregate formation from extended polyQ proteins has been shown to occur

from initial phase separation into liquid-like condensates followed by a transition into gel- and solid-like assemblies (48, 49). A similar role has been shown for Q-rich regions in the FUS family of proteins, whereby glutamines stabilize liquid-like droplets through hydrogen bond networks (50) and result in solidification into more gel-like structures (51). Thus, while Q-rich regions can drive the formation of liquid-like condensates, a lack of favorable protein-solvent interactions in these systems can lead toward more solid-like aggregates over time. Intriguingly, while we see that the IDR of *Se7942* McdB is not necessary or sufficient for driving phase separation like the Q-rich region, IDR charge content can mediate protein-solvent interactions to influence condensate solubility in response to pH. Thus, for *Se7942* McdB, there is a fine interplay between a Q-rich domain underlying protein-protein interactions and an IDR mediating the solubility of higher order assemblies.

The role of IDR-hydration in mediating condensate solubility

For many phase separating proteins, IDRs are necessary and sufficient for driving phase separation to a degree that is comparable to the full-length protein (30, 31, 32, 33, 34). In contrast, we show here that each domain of McdB provides a specific aspect necessary for full-length phase separation activity as described above. Interestingly, while deleting the IDR did not prevent phase separation, substituting six basic residues in the IDR could substantially solubilize condensates both *in vitro* and *in vivo*. These findings underscore the importance of the IDR not as a driver of phase separation, but rather a modulator of solubility. In line with this, recent models have described how solvation of charged residues within IDRs can serve as a key modulator of protein phase separation (52, 53, 54). Indeed, our data show that substituting basic residues in the IDR for polar, hydrophilic residues incrementally shifts the protein from the

dense- to soluble-phase. Thus, we conclude that hydration of positive charges within the IDR modulates McdB solubility, which helps explain the high sensitivity of McdB phase separation to changes in pH.

As the phase separation field develops, similar nuanced mechanisms are being reported that describe combinations of phase separation-driver domains, solubility modulators, and influences of oligomerization. For example, some proteins require oligomerization to provide the multivalency needed for phase separation (36, 55, 56), where some IDRs can only induce phase separation when fused to an oligomerizing domain (35). Interestingly, a recent report on the bacterial protein PopZ describes a model similar to what we propose here for McdB in which a folded domain serves as the driver of phase separation, and an IDR modulates phase separation to tune solubility of the condensates (19). This is intriguing because PopZ and McdB are functionally and evolutionarily distinct, thus representing a potential instance of convergence onto a similar mode of condensate regulation.

MoRFs related to ParA/MinD-family ATPases and phase separation

Our previous bioinformatics data have shown that there is a conserved enrichment of basic residues within the N-termini of McdB homologs (8). Here, we provide evidence for a potential MoRF in the N-terminus of *Se7942* McdB that folds upon its interaction with the cognate ATPase McdA. Intriguingly, we see that the conserved basic residues in McdB contribute to the stability of this structure. For several other ParA/MinD-family ATPases, the associated partner proteins have also been shown, or suggested, to interact with their cognate ATPase via positively charged and disordered N-termini (57, 58, 59, 60, 61, 62). More recent data suggest the N-termini from some of these partner proteins can also fold and stabilize as α-

helices upon interaction with their cognate ATPase (63, 64). Thus, N-terminal MoRFs stabilized through critical basic residues may be a conserved means of mediating the interaction between partner proteins and their cognate ParA/MinD-family ATPases.

Notably, the basic residues in the N-terminus of *Se7942* McdB that help stabilize the potential MoRF also play a role in the solvation of McdB condensates. It has been shown for some proteins, conformational shifts can facilitate the transition from soluble to dense phase (65, 66), consistent with a recent model describing how protein conformation can regulate exchange of molecules between dense and dilute phases (67). Thus, conformational heterogeneity can affect protein phase separation, reminiscent of a MoRF. Whether and how the phase separation activity of McdB is related to its structural dynamics and interactions with McdA is a focus for future study.

McdB pH sensitivity may be related to the efficiency of cyanobacterial carbon capture

In carbon-fixing organisms, the collection of processes that contribute to the efficient fixation of carbon are referred to as the carbon concentrating mechanism (CCM). The development of a model for the cyanobacterial CCM has provided insight into how different processes, such as membrane permeability and the presence of carboxysomes, affect overall carbon capture (68). It has been shown that incorporation of a pH flux into CCM models provides values more consistent with experimentation (69). 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 carboxysome lumen being ~ pH 7.5 (69, 70).

Intriguingly, one of our major findings here is the identification of pH as a major regulator of McdB phase separation, with lower pH values favoring phase separation. If and how carboxysomes maintain a lower pH than the cytoplasm has yet to be determined experimentally. However, it is attractive to speculate that coating carboxysomes with a viscous, pH-sensitive McdB phase may influence the diffusivity of carboxysome shells to protons, which is an important aspect of current models for an efficient CCM (68, 69). Indeed, in a previous study, we have shown that cells deleted for McdB show additional growth deficits compared to cells lacking McdA, suggesting that McdB plays additional roles in the CCM beyond spatially regulating carboxysomes with McdA (71). Moreover, a separate study has shown that the shells of purified carboxysomes, which lack McdB, are freely permeable to protons (72). Thus, it is attractive to speculate that McdB is required as a proton barrier to maintain efficient carbon capture and cell growth. We expect that the McdB mutants and mechanistic insights outlined in this report will aid in future attempts to address this potential role.

Tunable protein condensates as useful tools for synthetic biology

One widespread function of biomolecular condensates is their ability to regulate enzyme activity (73). The specific chemistries of condensates can affect the degree to which certain metabolites and enzymes are soluble within the dense phase. Thus, condensates can serve as specific reaction centers that regulate the overall metabolism of a cell by transiently altering the activities of key reactions. In line with this, it has been shown that certain scaffolding proteins can form phase separated condensates with RuBisCO (74, 75). It has been speculated that these RuBisCO containing condensates were the original CCM, which then led to the evolution of carboxysomes and the modern CCM (70). An exciting future direction for the phase separation

field is the prospect of designing these enzyme-containing condensates and implementing them in synthetic cells to engineer metabolism (18, 19, 73). Here, we have identified how to tune phase separation of a glutamine-rich condensate through alterations to the pI of the N-terminal IDR of McdB (Fig. 5C). As mentioned above, both a driver-domain and a tuner-domain have also been identified for the phase separation activity of the bacterial protein PopZ (19). Importantly, the identification of how to fine-tune PopZ phase separation allowed this group to then design PopZ variants that could form condensates in human cells (19). Using this strategy, a future direction will be to tune the properties of McdB condensates in heterologous systems so as to expand the repertoire of tunable protein condensates that can be used to manipulate the metabolisms of synthetic cells (19).

ACKNOWLEDGEMENTS:

We would like to thank Drs. Jeanne Stuckey and Jennifer Meagher at the University of Michigan Proteomics Resource Facility for help with SEC-MALS. Dr. Henriette Remmer at the University of Michigan Proteomics & Peptide Synthesis Core for help with N-terminal sequencing. Dr. JK Nandakumar for advice on SEC-MALS experiments and access to ASTRA software. The National Crystallization Center at the Hauptman-Woodward Medical Research Institute for performing crystallization buffer screens. 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.

DECLARATION OF INTERESTS:

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT:

The data that supports the findings of this study are available in the supplementary material of this article.

EXPERIMENTAL PROCEDURES

Protein expression and purification

Wild-type and mutant variants of 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 37ºC unless otherwise stated. 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 -80ºC.

Pellets were resuspended in 30 mL lysis buffer [300 mM KCl; 50 mM Tris-HCl pH 8.4; 5 mM BME; 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]. 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]. 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 to a final concentration of 1:100 protease:sample, and incubated overnight at 23ºC 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 [150 mM KCl; 20 mM CAPS pH 10.2; 5 mM BME; 10% glycerol]. Peak fractions were pooled, concentrated, and stored at -80ºC.

Proteolysis and N-terminal sequencing

Proteolysis was performed on *Se7942* 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 30°C 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.

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.

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.

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.

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. 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.

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 (14). 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.

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 (14). 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).

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 37ºC to an OD600 = 0.3 and induced with 1 mM IPTG. Following induction, cultures were grown at 37ºC 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.

AlphaFold 2 modelling of Se7942 McdA and predictions of McdB NTD with Rosetta.

The McdB(N-terminus)-McdA docking models were generated using the CollabFold implementation of AlphaFold2 (76, 42). We generated five structures with the default CollabFold/AlphaFold2 hyperparameters, save for the number of recycles being increased to 12. The structures were energetically minimized with AMBER using the Amber99sb force field. We

selected docked models based on the pLDDT scores of the binding interface residues and similarity to previously resolved ParA-like ATPase/partner-protein crystal structures (ref). In addition to the AlphaFold2 modeling, we also generated docked models using Rosetta's FlexPepDock protocol (77). First, we equilibrated the McdA homodimer structures generated from AlphaFold2 to the ref2015_cart_cst Rosetta force field with the FastRelax full-atom refinement protocol with cartesian coordinate space minimization using the lbfgs_armijo_nonmonotone minimizer. To preserve the position of the backbone atoms predicted by AlphaFold2, a backbone atom coordinate constraint was added. For this initial step, we generated 20 trajectories with the lowest scoring structure being used for the McdB(N-terminus) docking step. We used the score3 with the docking_cen.wts_patch and the REF2015 force fields for the low-resolution and high-resolution docking steps of the FlexPepDock protocol respectively. We did this to simulated 10,000 docking trajectories. For the top 4 docked models, we validated key binding residues through an in-silico alanine mutation scan and ΔΔG calculations. We calculated the ΔΔG using the FlexDDG protocol in Rosetta with the talaris2014 forcefield (78). In FlexDDG, for each mutation trajectory, the backbone and side chain conformations were sampled 35000 times using Rosetta's monte carlo backrub method. At every 2500 sample interval the ΔΔG of mutation was calculated. The final reported ΔΔG is the average of 35 such trajectories.

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. 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).
13. B. E. Funnell, ParB partition proteins: Complex formation and spreading at bacterial and plasmid centromeres. *Front. Mol. Biosci.* **3**, 44 (2016).
14. S. Alberti, A. Gladfelter, T. Mittag, Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).
15. 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).
16. 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).
17. J. S. MacCready, A. G. Vecchiarelli, Positioning the model bacterial organelle, the carboxysome. *MBio* **12** (2021).
18. W. Peeples, M. K. Rosen, Mechanistic dissection of increased enzymatic rate in a phase-separated compartment. *Nat. Chem. Biol.* **17**, 693–702 (2021).

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

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

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

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

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

24. 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).

25. 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).

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

27. 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).

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

29. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker, V. N. Uversky, PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta - Proteins Proteomics* **1804**, 996–1010 (2010).

30. 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).

31. 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).

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

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

34. 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).

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

36. 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).

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

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

39. R. Sharma, A. Sharma, A. Patil, T. Tsunoda, Discovering MoRFs by trisecting intrinsically disordered protein sequence into terminals and middle regions. *BMC Bioinformatics* **19**, 155–163 (2019).

40. Mohan, *et al.*, Analysis of Molecular Recognition Features (MoRFs). *J. Mol. Biol.* **362**, 1043–1059 (2006).

41. S. S. Kim, J. T. Seffernick, S. Lindert, Accurately Predicting Disordered Regions of Proteins Using Rosetta ResidueDisorder Application. *J. Phys. Chem. B* **122**, 3920–3930 (2018).

42. J. Jumper, *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

43. J. Ko, J. Lee, Can AlphaFold2 predict protein-peptide complex structures accurately? *bioRxiv*, 2021.07.27.453972 (2021).

44. 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).

45. 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).

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

47. 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).

48. S. Zhang, E. Hinde, M. Parkyn Schneider, D. A. Jans, M. A. Bogoyevitch, Nuclear bodies formed by polyQ-ataxin-1 protein are liquid RNA/protein droplets with tunable dynamics. *Sci. Rep.* **10**, 1–18 (2020).

49. T. R. Peskett, *et al.*, A Liquid to Solid Phase Transition Underlying Pathological Huntingtin Exon1 Aggregation. *Mol. Cell* **70**, 588–601 (2018).

50. C. Murthy, *et al.*, Molecular interactions underlying liquid−liquid phase separation of the FUS low-complexity domain. *Nat. Struct. Mol. Biol.* **26**, 637–648 (2019).

51. 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).

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

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

57. 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).

58. 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).

59. 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).

60. 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).

61. 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).

62. 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).

63. H. Zhang, M. A. Schumacher, Structures of partition protein ParA with nonspecific DNA and ParB effector reveal molecular insights into principles governing Walker-box DNA segregation. *Genes Dev.* **31**, 481–492 (2017).

64. K. T. Park, M. T. Villar, A. Artigues, J. Lutkenhaus, MinE conformational dynamics regulate membrane binding, MinD interaction, and Min oscillation. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 7497–7504 (2017).

65. S. Lüdeke, *et al.*, Dynamic Structural Changes and Thermodynamics in Phase Separation Processes of an Intrinsically Disordered–Ordered Protein Model. *Angew. Chemie Int. Ed.* **61**, e202112738 (2022).

66. E. Conicella, G. H. Zerze, J. Mittal, N. L. Fawzi, ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure* **24**, 1537–1549 (2016).

67. Y. Zhang, *et al.*, Interface resistance of biomolecular condensates. *bioRxiv* (2022) https:/doi.org/10.1101/2022.03.16.484641 (April 14, 2022).

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

69. 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).

70. 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).

71. 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).

72. B. B. Menon, S. Heinhorst, J. M. Shively, G. C. Cannon, The carboxysome shell is permeable to protons. *J. Bacteriol.* **192**, 5881–5886 (2010).

73. 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).

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

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

76. M. Mirdita, *et al.*, ColabFold - Making protein folding accessible to all. *bioRxiv*, 2021.08.15.456425 (2022).

77. B. Raveh, N. London, L. Zimmerman, O. Schueler-Furman, Rosetta FlexPepDock ab-initio: Simultaneous Folding, Docking and Refinement of Peptides onto Their Receptors. *PLoS One* **6**, e18934 (2011).

78. K. A. Barlow, *et al.*, Flex ddG: Rosetta Ensemble-Based Estimation of Changes in Protein-Protein Binding Affinity Upon Mutation. *J. Phys. Chem. B* **122**, 5389 (2018).

SUPPLEMENTAL MATERIALS

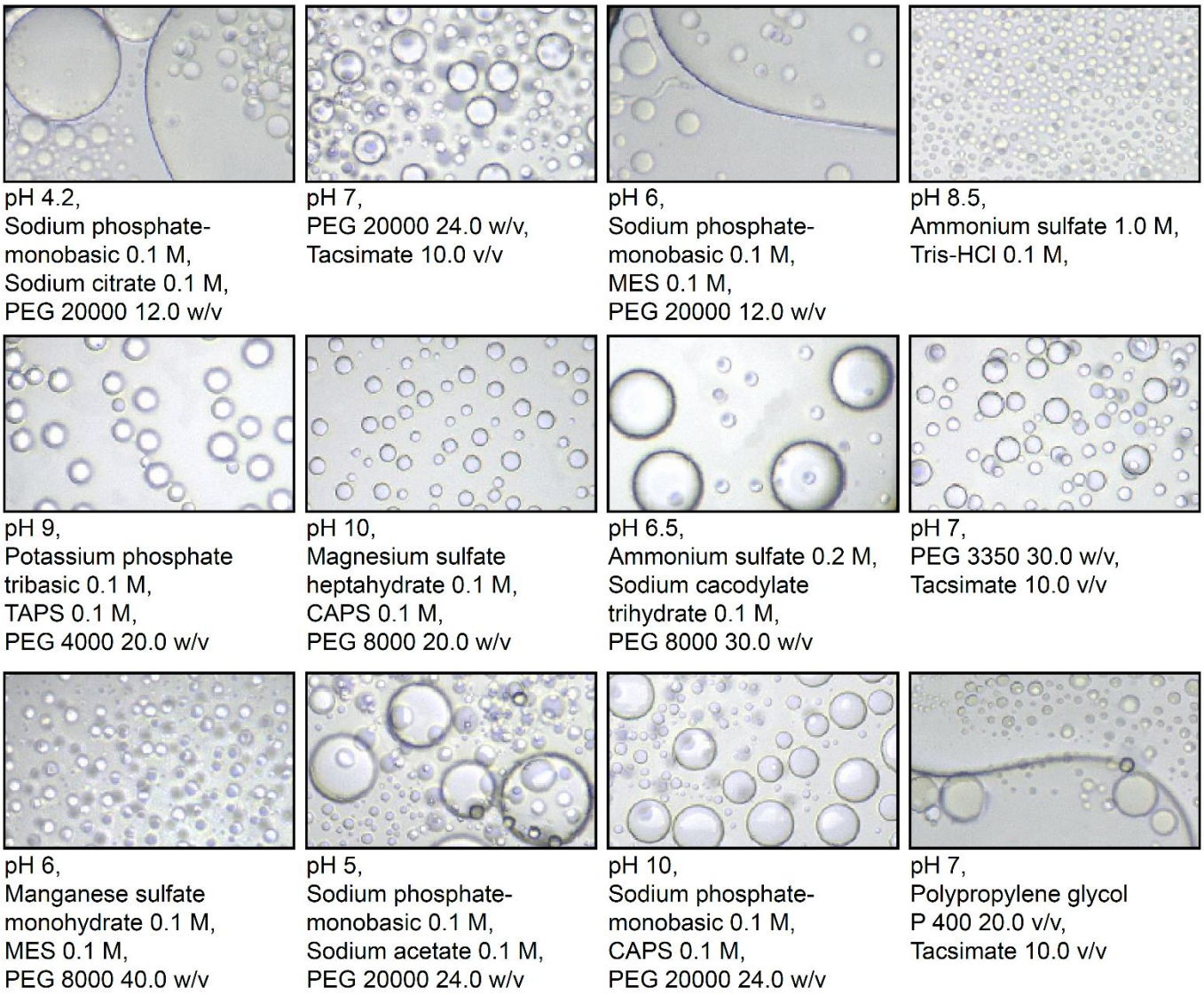


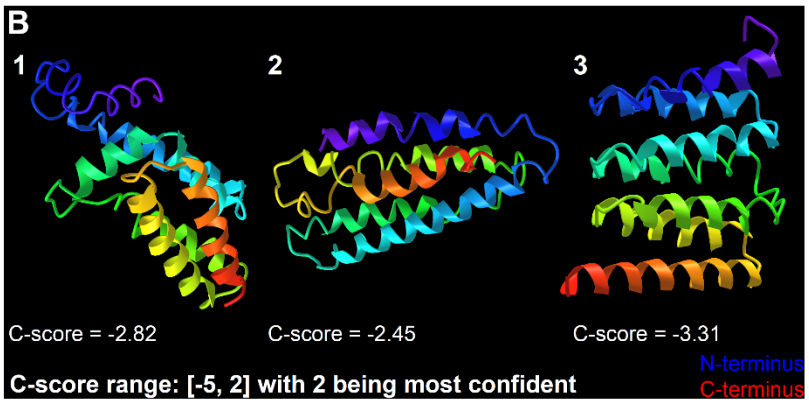
Figure S1: Phase separation of *Se7942* McdB across a range of buffer conditions during crystal screens. Images taken during buffer screens for crystallography. McdB at 10 mg/mL in (50 mM KCl; 10 mM CAPS pH 10.2) was diluted into the buffers indicated below each image. All images shown are at the same final concentration and magnification. Images were taken after 24-36 hours post dilution.

A

```
                        20                  40                  60
                         |                   |                   |
Seq.   MTDAFDRLKKRSRTPIAREGSLTTGPELSDRPLQLLPREFETFCDRYAVHAGDVIEAALDLVLLDPDLQQRLLQRL
Pred.  CCCHHHHHHHCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCHHHHHHHHHHH
Conf.  960778888624886201477676856576716662388998889887610149999766761087799999999

         80                  100                 120                 140
          |                   |                   |                   |
Seq.   RQGNGSDRVWLGTACPRSWQQQLQQQAQDQGLSEADLLQEAIAQRLDLVLGQTTLREEVTLLRQELDQLKRKLHGW
Pred.  HHCCCCCCSSSCCCCCHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCC
Conf.  707897633344759668999999999876865888999999999889871137999999999999999874289
   H:Helix; S:Strand; C:Coil
```

B



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 S2: I-TASSER predictions for *Se7942* 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.

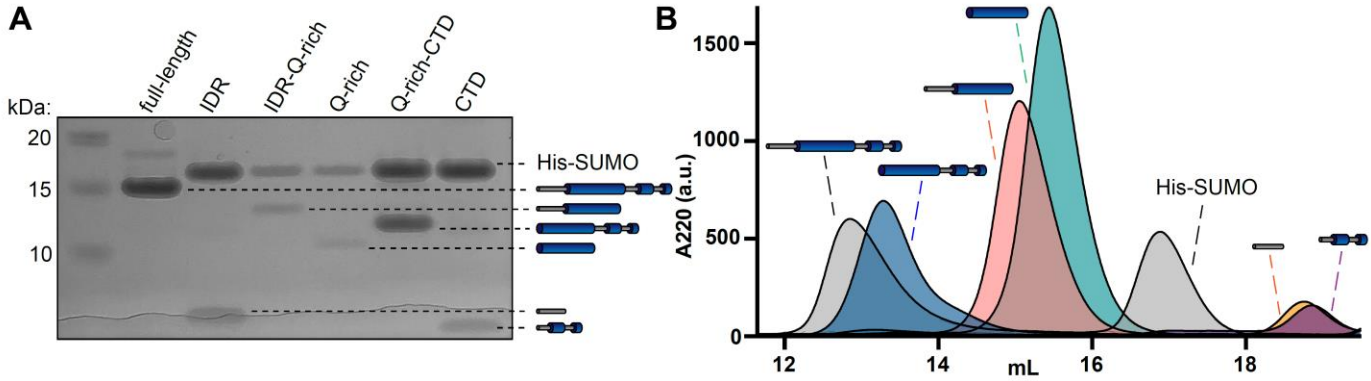


Figure S3: **McdB truncations display different oligomeric states.** (**A**) SDS-PAGE analysis shows that wild-type McdB and all truncation mutants run at a lower molecular weight compared to the His-SUMO solubility tag. (**B**) Size exclusion chromatography (SEC) showed that full-length McdB and the Q-rich+CTD domain have similar elution profiles, suggesting similar oligomeric forms. The Q-rich domain with and without the IDR also have similar elution profiles, but after full-length and before His-SUMO, suggesting an intermediate oligomer. The IDR and CTD mutants, on the other hand, eluted after the His-SUMO tag, showing they remain monomeric.

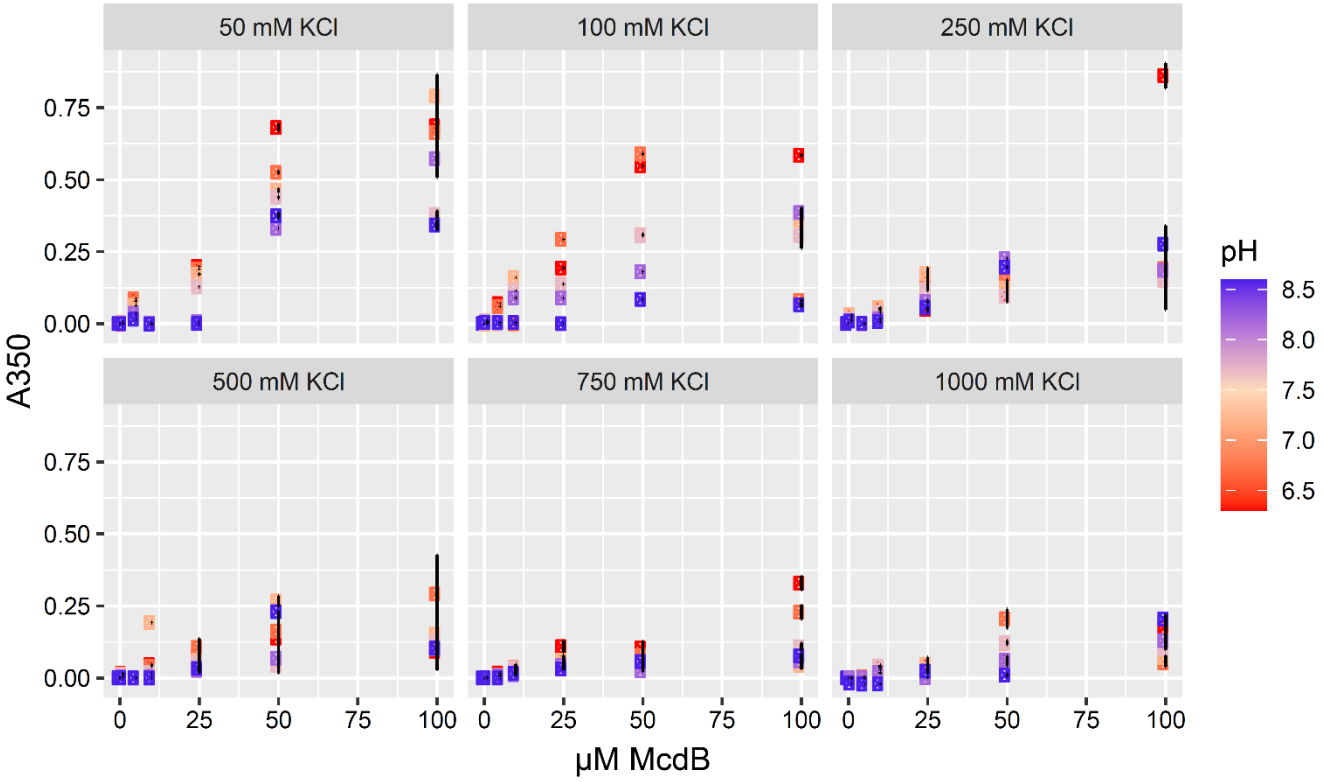


Figure S4: **Multi-dimensional phase diagrams for *Se7942* McdB phase separation.** Turbidity-based phase diagrams for McdB phase separation across varying protein concentration, KCl concentration, and pH. Data points represent the mean and error bars represent SD from at least three technical replicates. Turbidity monitored at A = 350 nm.

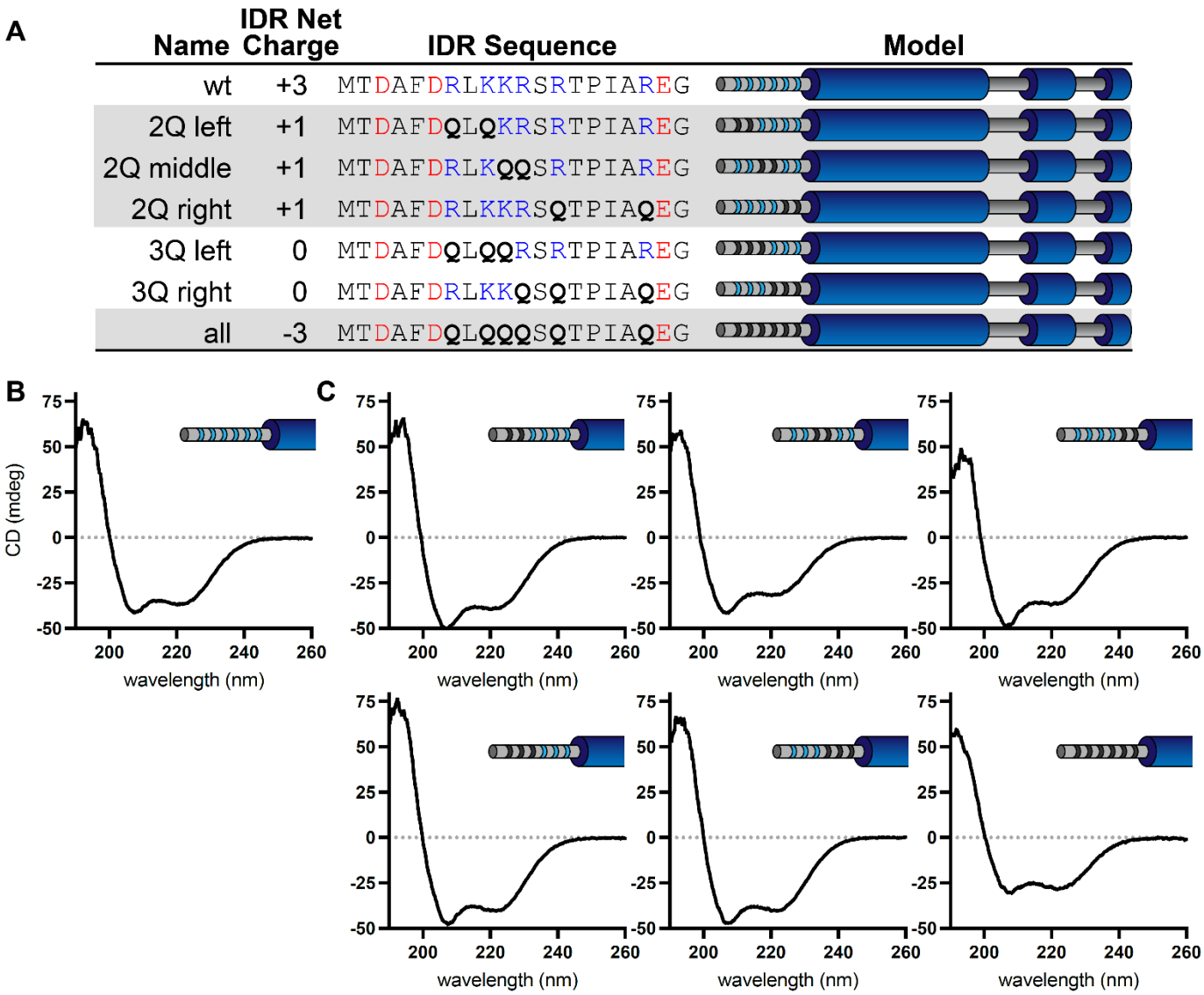


Figure S5: CD spectra of wild-type McdB and N-terminal glutamine substitution mutants. (**A**) Table showing the net charge and N-terminal IDR sequence of wildtype McdB compared to the glutamine-substitution mutants. Acidic- and basic-residues in the IDR are colored red and blue, respectively. Glutamine-substitutions 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**) wildtype McdB and (**C**) mutants with the indicated glutamine substitutions in the N-terminal IDR of McdB.

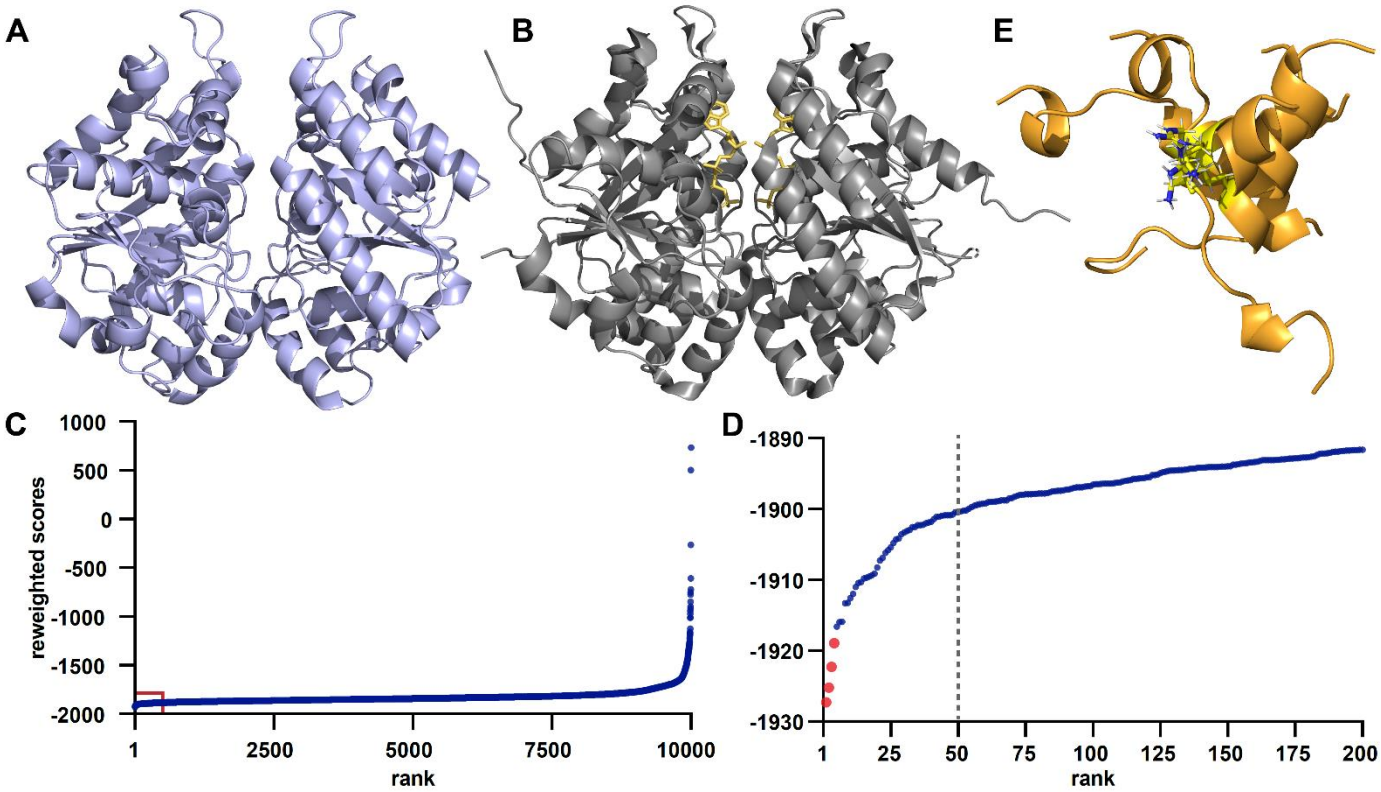


Figure S6: McdA structures, similarities between favorable docked McdBs, and energy levels of the docked ensemble. (**A**) Structural model of the homodimer of *Se7942* McdA generated using AF2. (**B**) Model of the crystal structure of *Cyanothece* McdA[D38A]-ATP (PDB 6NOP) (44). Shown in the same orientation as model in (A). ATP molecules are colored yellow. (**C**) Plot of the energy scores for each docked model out of an ensemble of 10,000 generated in Rosetta. Models are plotted from lowest to highest energy level, with rank 1 = lowest score. Region outlined in red is expanded in (**D**), with vertical line added at rank 50. The top 4 ranks are bolded red. These models are shown in Fig. 8C. (**E**) Overlay of the McdB N-termini from models shown in Fig. 8C, D. Orientation of the peptides differ, but spatial orientation of key basic residues (highlighted yellow) is well conserved.

Table S1. Secondary structures of McdB N-termini from the top 50 docked Rosetta models.

	MTDAFDRLKKRSRTPIAREG		MTDAFDRLKKRSRTPIAREG
1.	CHHHHHHHHHHCCCCHHHHC	26.	CCCHHHHHHHHCCCCCCCCC
2.	CCHHHHHHHHHHHCCCCCCC	27.	CCHHHHHHHHHHHCCHHHHC
3.	CCCCHHHHHHCCCCHHHHHC	28.	CCHHHHHCCCCCCCCCCCCC
4.	CHHHHHHHHHCCCCCHHHHC	29.	CCCHHHHHHHCCCCCCCCCC
5.	CCHHHHHCHHHHHCCHHHHC	30.	CCCHHHHHHHHCCCCCCCCC
6.	CCHHHHHHHHHHHCCCCCCC	31.	CCCHHHHHCCCCCCCHHHHC
7.	CHHHHHHHCCCCCCCCCCCC	32.	CHHHHHHHHHCCCCCCCCCC
8.	CCCHHHHHCCCCCCCCCCCC	33.	CHHHHHHHHHHHCCCHHHHC
9.	CCCCCCCCCCCCCCHHHHHC	34.	CHHHHHHHHCCCCCCCCCCC
10.	CCCCCHHHHHCCCCHHHHHC	35.	CHHHHHHHHCCCCCCCCCCC
11.	CHHHHHHHHCCCCCCCCCCC	36.	CHHHHHHHHHHCCCCHHHHC
12.	CCHHHHHHHHHCCCCCCCCC	37.	CHHHHHHHHHHHCCHHHHHC
13.	CCHHHHHHHCCCCCCCCCCC	38.	CHHHHHHHHHCCCCCHHHHC
14.	CHHHHHHHHHHHCCCCCCCC	39.	CHHHHHHHHCCCCCCCCCCC
15.	CCHHHHHHHCCCCCHHHHHC	40.	CHHHHHHHHHHHHCCCCCCC
16.	CHHHHHHHHHHHHCCCCCCC	41.	CCHHHHHHHHHHCCCCCCCC
17.	CCHHHHHHHHHHCCCCCCCC	42.	CCHHHHHCCCCCCCCHHHHC
18.	CCHHHHHHCCCCCCCCCCCC	43.	CCHHHHHHHCCCCCHHHHHC
19.	CCHHHHHHHHHCCCCHHHHC	44.	CCCCHHHHCCCCCCHHHHHC
20.	CCHHHHHHHHHHHCCCCCCC	45.	CHHHHHHHHHCCCCCCCCCC
21.	CCCHHHHCCCCCCCCCCCCC	46.	CCCHHHHHHHHHCCCCCCCC
22.	CHHHHHHHHHHHHCCCCCCC	47.	CHHHHHHHHHHCCCCCCCCC
23.	CHHHHHHHHHHCCCCHHHHC	48.	CCHHHHHHHCCCCCHHHHCC
24.	CHHHHHHHHHCCCCCCCCCC	49.	CHHHHHHHHCCCCCCCCCCC
25.	CCHHHHHHHHCCCCHHHHHC	50.	CCHHHHHHHHCCCCCCCCCC

C = random coil

H = α-helix