Title

Molecular Design Principles of Bacterial Carbon Fixation: Investigations into Carboxysome Assembly and Permeability

Permalink

https://escholarship.org/uc/item/3b79f862

Author

Turnsek, Julia Borden

Publication Date

2023

Peer reviewed|Thesis/dissertation

Molecular Design Principles of Bacterial Carbon Fixation: Investigations into Carboxysome Assembly and Permeability

By

Julia Borden Turnšek

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cellular Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Savage, Chair
Professor Sabeeha Merchant
Professor Arash Komeili
Professor John Dueber

Fall 2023

Abstract

Molecular Design Principles of Bacterial Carbon Fixation: Investigations into Carboxysome Assembly and Permeability

by

Julia Borden Turnšek

Doctor of Philosophy in Molecular and Cellular Biology

University of California, Berkeley

Professor David Savage, Chair

All life on Earth relies on biological carbon fixation, the process by which organisms convert inorganic carbon, primarily in the form of carbon dioxide (CO_2), into longer chain compounds to fuel cellular processes. To enhance the efficiency of CO_2 fixation, certain types of bacteria, specifically cyanobacteria and some proteobacteria, evolved specialized proteinaceous microcompartments called carboxysomes. Carboxysomes encapsulate the enzymes carbonic anhydrase and Rubisco inside a polyhedral layer of shell proteins. This molecular architecture serves to concentrate CO_2 around Rubisco, allowing it to operate at its maximum catalytic rate.

Correct carboxysome assembly is essential to the survival of the organism in CO_2 concentrations found in today's atmosphere (~0.04%). In the α-carboxysome lineage, the disordered scaffold protein CsoS2 links Rubisco and shell proteins, and is absolutely required for carboxysome formation and cell growth at ambient CO_2 levels. This work examines how the sequence of CsoS2 scales from a disordered amino acid chain to directing the ordered self-assembly of thousands of proteins. It investigates how cells utilize specific chemistries, such as redox reactions, to assist in this assembly pathway. The result of this molecular design and coordinated construction is to build a carboxysome with a precise permeability, yet this permeability has never been measured. Results presented here address these fundamental questions.

I interrogated highly conserved and repetitive residues in CsoS2 to determine their role in carboxysome assembly. Through *in vivo* mutagenesis and *in vitro* biochemical assays I discovered that the residues VTG and Y are necessary for carboxysome assembly, and bind weakly yet multivalently to shell proteins. Conserved cysteine doublets, which hinted at a role for redox in assembly, showed no effect when mutated *in vivo*, but displayed biochemical phenotypes *in vitro*. In a major step towards reconstituting carboxysomes *in vitro*, I demonstrated formation of carboxysome-like phase-separated condensates with Rubisco, CsoS2, and shell, thereby showing that key carboxysome proteins can self-associate in a cell-free environment.

Once assembled correctly, the carboxysome establishes a permeability barrier and selectivity filter, allowing entry of essential metabolites such as ribulose bisphosphate and bicarbonate while restricting leakage of CO_2. To measure carboxysome permeability, we developed two parallel methods, one based on a bulk plate assay and one on single-particle microscopy. Both methods utilized the redox sensitive reporter protein roGFP to simultaneously measure both the permeability of reducing agents and the internal carboxysome redox environment. Data from both approaches revealed that purified carboxysomes were permeable to the reducing agent TCEP, which reduced encapsulated roGFP over time.

Carboxysomes are the bacterial domain's solution to the problem of capturing dilute CO_2 from air and water, concentrating it, and converting it into sugars. Carboxysome functionality depends on the robust self-assembly of thousands of proteins, establishment of a specific internal chemical environment, and control over metabolite permeability. Insights from this work augment our understanding of these processes, and will aid future efforts to engineer carboxysomes into alternative organisms or cell-free systems for enhanced biological carbon capture.

Table of Contents

Preliminary Pages

Chapter 1: Introduction

Chapter 2: Conserved and repetitive motifs in an intrinsically disordered protein drive α-carboxysome assembly

Chapter 3: Measuring permeability and redox in the carboxysome

Chapter 4: Conclusion

Appendix

List of Figures

Chapter 1: Introduction

Chapter 2: Conserved and repetitive motifs in an intrinsically disordered protein drive α-carboxysome assembly

Chapter 3: Measuring permeability and redox in the carboxysome

Dedicated to my parents.
Thank you for everything.

Quotes

Information imprisoned in old cliché patterns can often come together in a new way of its own accord once the pattern is disrupted.

- Edward de Bono, Lateral Thinking

The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka' but 'That's funny...'

- Isaac Asimov

Doesn’t everything reduce in the end to a poetic image – one that encapsulates an entire experience into one stroke?

- Frances Mayes

Suspend disbelief.

- Rob Phillips

Acknowledgements

To my family – To my Mom and Dad, a heartfelt thank you for nurturing my love of science, letting me choose my own path, and supporting me unwaveringly wherever I go. You both show me every day the meaning of lifelong curiosity, learning, and practicing your craft. I'm lucky to have such role models in my life from day 1. To my sister Sharon, thank you for being there for me, for listening, and for being the amazing sister that you are. I feel incredibly fortunate to have you by my side.

To my AP Biology teacher Ms. Ward – Your inspirational teaching sparked my future career as a molecular biologist. It was in your class where I first learned about the power and possibilities of genetic engineering. I'll always remember our conversation a few years ago about biological phase separation – I hope one day it makes the high school biology textbooks!

To my early academic mentors Matt Knope and Liz Hambleton – Now that I'm reaching the end of my PhD I have a deep appreciation for how you took a chance on a young, inexperienced scientist. Thank you for introducing me to the world of research.

To my manager and mentor at Modern Meadow Lixin Dai – Thank you for showing me the scientist I wanted to be. From you, I learned the power of biological intuition, how to approach and cross seemingly intractable obstacles, and how to be a kind and effective leader. Thank you for everything you taught me, and for believing in me.

To Yuly Fuentes and Suzanne Lee – Thank you for not just nurturing, but expanding my creative side. You knew I straddled the two worlds of art and science and showed me that a third world exists where the two merge in a beautiful blend of possibilities.

To friends – Lupe, thank you for your friendship throughout this whole wild journey. You know how to keep it real and always see the humorous side of things. Thanks for sharing in all the ups and downs and for the lunchtime gossip. To Amreet and Marcela, thank you for always being by my side, no matter where in the world we are. And to Sheila and Paul, for your friendship and shared interest in learning Slovene. It was always a welcome break from science, especially during the pandemic.

To Woods Hole, the Physical Biology of the Cell course leaders (Rob Phillips, Hernan Garcia, Jane Kondev, and Julie Theriot), and all my fellow PBoC coursemates – There is no freer, more uninhibited place to ask crazy questions and chase answers with creativity and gusto. Thanks for teaching us to be open in our approaches to science and life. "Few" is the scale of everything, the power of god can be estimated, and equilibrium is a state of mind!

To Melinda Kliegman – Thank you for recommending me to be a group leader for the IGI/CGIAR collaboration. Thank you for seeing and promoting my interests beyond the bench and giving me an opportunity to take this leadership role.

To Antonio – Thank you for trusting me as your teacher. It was fun to learn with you and from you. I can't wait to see where you end up!

To my committee – Thank you for your feedback and questions over the years, and for your many suggestions when I was facing obstacles.

To my PI, Dave – Thank you for always being a supportive mentor and giving me the freedom to define my own scientific story. You're leading a lab of great people doing fascinating science. The wealth of knowledge I've gained extends far beyond the content printed on these pages. From plants, CRISPR, libraries, grant writing, leadership, and more, it has been a truly holistic and enriching training experience.

To my labmates – A big thank you both to the people who joined before me (Cissi, Rob, Tom, Avi, Arik, Jack, Emeric, Eli, Luke) and to the people who joined after (Naiya, Maria, Brittney, Julia, Noam, Evan, David, Jorge, Rachel, Flora, Andrew, Luis, Muntathar, Cynthia). Cissi, thank you for teaching me how to purify proteins, basically the biochemist's version of learning how to ride a bike. It's safe to say you revolutionized my PhD by giving me this skillset. Back in the days when the lab was all guys, having you as another female voice and presence was a godsend. It's been amazing to see how many women have joined after us! Rob, the lab truly felt empty after you left. I always enjoyed our fun conversations on encapsulin evolution and biochemistry. Arik, your sarcastic humor is hilarious. Thanks for being there since day -1, when I interviewed for Berkeley! Brittney, Rachel, and Eli, thanks for keeping it fun and funky in Bay 3. Evan, it was always fun to brainstorm the next big artisanal biotech with you! Naiya, you are the best baymate I could have asked for. Thanks for listening to all my random commentary over the years. You are a role model for patience and persistence with experiments. Lastly, Luke: Thank you for being my partner in carboxysome research until the very end. You're one of the greatest teachers I've known, and a brilliant scientist. I also love how you personify things; is CsoS2 a protein, or an acrobat / double agent / flaky friend?

And finally, a profound thank you to my husband, Jernej. Veš, da je življenje veliko več kot le znanost. Na celotnem potovanju si bil ob meni in vedno si vedel, da je samo to – le potovanje, majhen, a pomemben del naših življenj. Tvoja ljubezen in podpora, pa tudi kanček tiste trme, ki jo imam zdaj kot Turnšek, so me ponesli do te točke. Verjameva drug v drugega, se spodbujava in navijava drug za drugega. You mean more to me than anything else, I thank you, and I love you!

Chapter 1:
Introduction

1.1 Chapter Summary

Carboxysomes are CO_2-fixing protein compartments present in all cyanobacteria and some proteobacteria. These structures are attractive candidates for carbon assimilation bioengineering because they concentrate carbon, allowing the fixation reaction to occur near its maximum rate, and because they self-assemble in diverse organisms with a set of standard biological parts. Recent discoveries, including work presented in this thesis, have expanded our understanding of how the carboxysome assembles, distributes itself, and sustains its metabolism. These studies have already led to substantial advances in engineering the carboxysome and carbon concentrating mechanism into recombinant organisms, with an eye towards establishing the system in industrial microbes and plants. Future studies may also consider the potential of *in vitro* carboxysomes for both discovery and applied science.

The following text and figures are adapted from Borden, J. S., and Savage, D. F. (2021) New discoveries expand possibilities for carboxysome engineering. Curr. Opin. Microbiol. 61, 58–66. (1)

1.2 Introduction

All cyanobacteria and many chemoautotrophic proteobacteria use specialized proteinaceous organelles called carboxysomes to facilitate CO_2 fixation. Carboxysomes have fascinated researchers and biotechnologists for both their icosahedral structure and ability to enable efficient carbon fixation kinetics. It's estimated that ~10-25% of CO_2 fixed globally passes through these compartments annually (2, 3). Since they were first purified in 1973 (4), researchers have sought to both understand and engineer carboxysomes. Although the presence of Rubisco signaled a critical role in CO_2 fixation, studies on carboxysomes continue to reveal new and unexpected components, structures, and potential applications.

Carboxysomes are icosahedral protein assemblies ranging from 100-500 nm in diameter, depending on the species (5). They have a proteinaceous shell composed primarily of hexameric proteins and are capped with pentameric proteins at the icosahedral vertices. In general, they enclose Rubisco, carbonic anhydrase (CA), a Rubisco nucleating protein, and, likely, Rubisco activase. There are two lineages of carboxysomes, α and β, which evolved convergently (6, 7). Remarkably, both lineages arrived at the same general carboxysome structure and function, though they differ in gene organization and protein sequences.

Carboxysomes function within a broader metabolic network called the Carbon Concentrating Mechanism, or CCM (Figure 1a). Inorganic carbon transporters in the cell membrane pump HCO_3^- into the lumen, raising its concentration to about 30x the equilibrium concentration in water (3, 8). The disequilibrium between HCO_3^- and CO_2 is advantageous because it stockpiles a charged, and therefore membrane-impermeable, form of carbon in the cell. This preferences the dehydration reaction in the carboxysome, concentrating CO_2 near Rubisco. Rubisco is thus poised to operate near

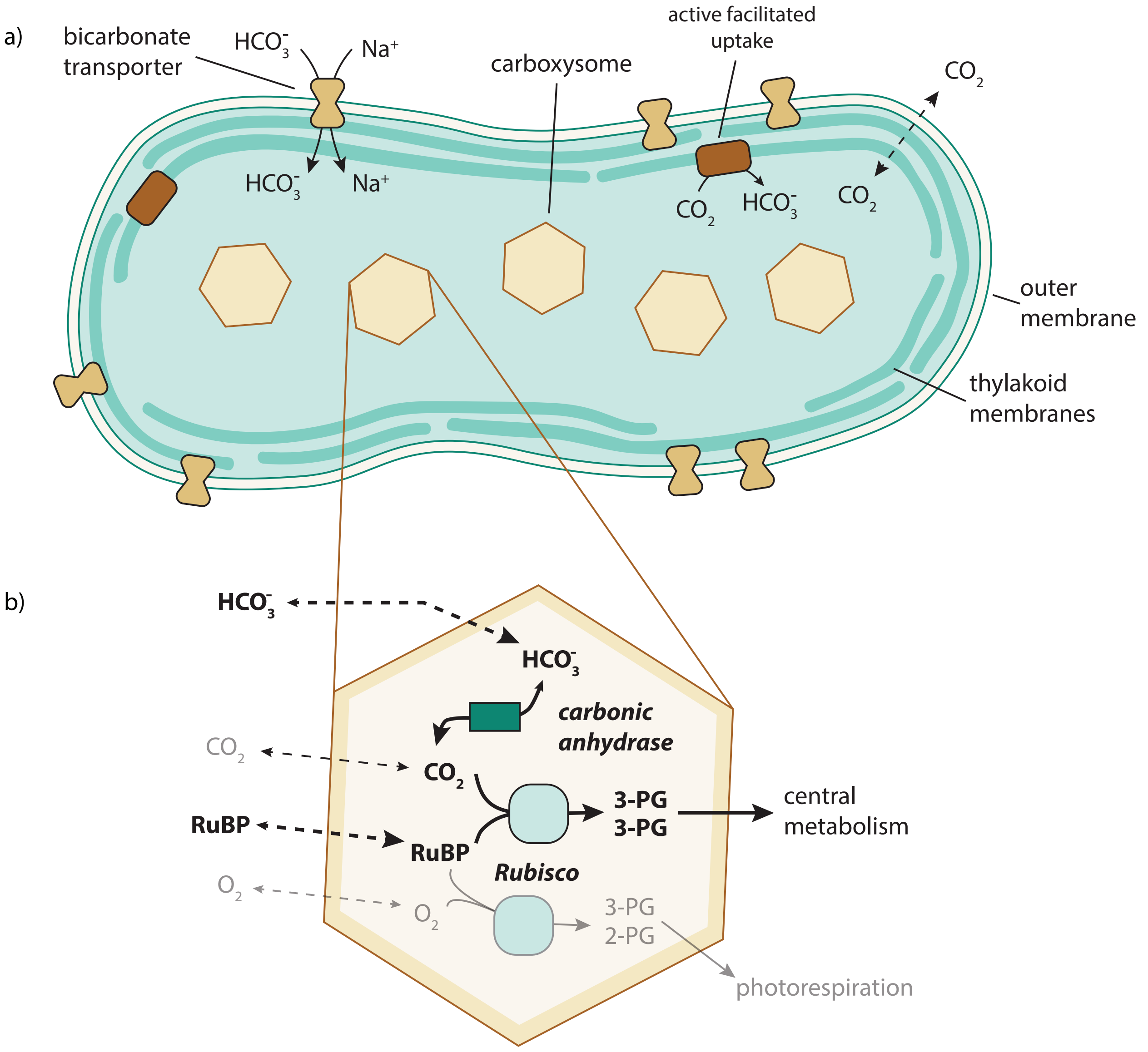


Figure 1. a) The CCM in a cyanobacterial cell. Bicarbonate transporters and facilitated CO_2 uptake proteins raise the intracellular HCO_3^- concentration while CO_2 flows freely across the plasma membrane. b) Carboxysome metabolism. HCO_3^- enters the carboxysome along its concentration gradient, where it is converted to CO_2 via a carbonic anhydrase. CO_2 and RuBP serve as substrates for Rubisco, which produces two molecules of 3-PG. O_2 may occasionally serve as a Rubisco substrate, though at a minimal level.

its V_{max} when carboxylating ribulose-1,5-bisphosphate (RuBP) to produce two molecules of 3-phosphoglycerate (3-PG) (Figure 1b). In addition to carboxylation, Rubisco can also oxygenate RuBP; the product of this off-target reaction must be recycled via wasteful photorespiration pathways. The high CO_2 environment of the carboxysome therefore competitively inhibits oxygenation, and it remains an open question as to whether exclusion of O_2 by the shell is necessary for CCM function (3). Finally, this unique environment enabled Rubisco evolution to maximize for carboxylation activity, and carboxysomal Rubiscos are among some of the fastest known Rubiscos, despite having low specificity for CO_2 over O_2 (9, 10).

Knocking out various components of the CCM renders cells incapable of growing at the atmospheric CO_2 concentration (~0.04%), and they must be grown in high CO_2 (~1-10%) (11, 12). In particular, carbonic anhydrase must be active only inside of the carboxysome; knocking it out or expressing it in the cytosol destroys the CCM (13). Carboxysomes must also limit CO_2 permeability so that it doesn't diffuse away from Rubisco. Pentamer deletion strains, which produce carboxysomes with pores at the icosahedral vertices, only grow in high CO_2 (14). A precise understanding of how shell proteins limit CO_2 diffusion while allowing entry and exit of other intermediates such as 3-PG and RuBP is still not well understood. Despite speculation that the carboxysome shell is selectively permeable, promoting uptake of HCO_3^- while blocking O_2, no direct evidence has been experimentally measured. Mathematical models show that the CCM does not require O_2 impermeability to function (3, 15), though both O_2 and CO_2 may encounter an increased resistance at the hexamer pore compared to HCO_3^- and 3-PG (15). In total, these results show that concentrating CO_2 near Rubisco by limiting CO_2 leakage from the carboxysome is essential to the function of the CCM and is an important principle in the development of biotechnological tools to concentrate CO_2.

1.3 From carboxysome structure to functional CCM reconstitution

Reconstituting functional carboxysomes, i.e. those that can concentrate carbon, into an alternative host organism has been a major academic and bioengineering goal. However, it has been difficult to do because structure alone cannot recapitulate the CCM. The first recombinantly produced carboxysomes were α-carboxysomes from the model proteobacterium *Halothiobacillus neapolitanus*, expressed in *E. coli*. Expressing the native 10-gene operon was sufficient to produce wild-type looking carboxysomes (16). Similarly, a synthetic operon of 12 genes from cyanobacterium *Synechococcus elongatus* PCC7942 produced wild-type looking β-carboxysomes in *E. coli* (17). The engineered heterologous systems in both studies possessed active Rubisco, but it remained unclear whether recombinant carboxysomes could concentrate carbon, arguably the carboxysome's essential feature. A transposon mutagenesis screen of *H. neapolitanus* under high CO_2 vs. low CO_2 conditions revealed dozens of new genes responsible for the functioning of the α-carboxysome CCM, including several uncharacterized proteins in a secondary operon (18). Characterizing unknown hits, as well as undertaking a systematic study in β-carboxysomes beyond the previous screens (11, 12), will be crucial to uncovering what components are necessary to fully reconstitute the CCM. The following sections highlight recently discovered proteins and protein activities and how they may translate into using CCMs for enhancing metabolism.

1.4 New discoveries in carboxysome assembly and CCM function

Researchers made early progress in identifying and characterizing major players in the carboxysome and CCM such as Rubisco, CA, shell proteins, and carbon transporters, reviewed in greater detail in references (6, 19). This section reviews proteins discovered or characterized in recent years that have greatly increased our understanding of the carboxysome and CCM, and which are important new entries in the carboxysome biotechnological toolbox.

CsoS2 & CcmM – CsoS2, from α-carboxysomes, and CcmM, from β-carboxysomes, are essential for carboxysome assembly and structure. Though they have no sequence or domain homology (Figure 2a & b), they share many striking similarities. Both are conserved, essential proteins located in the core carboxysome locus of their respective lineages (19). Both are highly abundant in the carboxysome, with numbers roughly equal to Rubisco holoenzyme (6).

Notably, both CsoS2 and CcmM bind Rubisco and facilitate carboxysome nucleation. Both α- and β-carboxysome Rubiscos evolved binding sites that bridge two large subunits while making contacts with the small subunit (Figure 2c) (20, 21). This likely ensures that only the 16-subunit Rubisco holoenzyme is encapsulated during carboxysome assembly. Both CsoS2 and CcmM Rubisco-binding domains contain 3-5 repeat motifs separated by predicted disordered sequences (Figure 2a & b). Despite these shared features, the binding domains differ in their secondary structure. In CsoS2, the Rubisco-binding N-terminal domain (NTD) repeats are alpha helical, while the C-terminal CcmM repeats have structural similarity to the Rubisco small subunit (termed small subunit-like or SSUL) (20–22). In CsoS2, a single repeat binds with low affinity, but multivalent interactions could promote high affinity binding across multiple Rubiscos, thus nucleating carboxysome assembly (Figure 2d) (21). Likewise, the three CcmM repeats together bind Rubisco with micromolar affinity (20, 22). Following a common theme for repetitive, multivalent proteins, both Rubisco-binding domains of CsoS2 and CcmM were shown to undergo liquid-liquid phase separation (LLPS) with Rubisco, though it should be noted both studies required salt concentrations below that of physiological 150 mM (20, 21).

CsoS2 and CcmM both have a short and long isoform, and the significance of this is not fully understood (Figure 2a & b), though work from Oltrogge et al. 2023 suggests that the ratio of long to short determines the intrinsic curvature and size of the carboxysome (23). The isoforms are produced in CsoS2 by ribosomal frameshifting and by an internal ribosome entry site (IRES) in CcmM [21,22]. In wild-type α-carboxysomes, both the short (Csos2A) and long (CsoS2B) forms are present at a roughly equimolar ratio (24). When the frameshifting site is mutated, CsoS2B is sufficient to reconstitute carboxysomes on its own, but CsoS2A cannot (24). Relatedly, CcmM has a short (M35) and long (M58) form. Both M35 and M58 are needed for functional β-carboxysomes (25).

Both CsoS2 and CcmM may possess redox-regulated intrinsic flexibility. Suggestively, in *Thermosynechococcus elongatus* BP-1, the CcmM C-terminal γ-CA is only active under disulfide-forming oxidizing conditions (26). Most repeat segments of both CsoS2 and CcmM contain 1-2 cysteines. Cells with β-carboxysomes with mutated CcmM cysteines grew 2-3 times slower than wild-type, and many carboxysomes were irregularly shaped (20). The effect in α-carboxysomes is not yet fully understood, and is investigated in this thesis. In biochemical studies, the reduced form of CcmM repeats bound Rubisco with higher affinity but showed less mobility under LLPS conditions (20). This suggests a model in which the carboxysome nucleates under reducing cytosolic conditions and, upon complete assembly, matures into a liquid-like oxidizing environment (Figure 2d). Microscopy of developing β-carboxysomes using a redox-sensitive GFP suggested that this model may be true *in vivo* (27). The effect of redox regulation in these compartments remains an understudied, yet potentially highly significant, aspect of their assembly and function, and is further examined in Chapters 2 & 3 of this thesis (28, 29).

McdA & McdB – Cells with β-carboxysomes arrange them linearly along a central longitudinal axis throughout growth and equally distribute carboxysomes to daughter cells during division (30). This organization is driven by a pair of proteins, McdA and McdB (31). McdA is a ParA-type ATPase that binds the nucleoid and shows a characteristic oscillatory behavior between cell poles. McdB, by analogy to plasmid partitioning systems (32), is thus thought to engage both the carboxysome and McdA. This facilitates an even distribution of carboxysomes, and those that lack either or both proteins show carboxysome clumping at a polar end. Expression of carboxysomes in a strain that lacks the proper positioning and partitioning machinery results in carboxysome aggregation, and loss of carbon fixation function in descendants without carboxysomes (30, 33). However, cells with McdA/B knockouts do not require high CO_2 to grow, likely because carboxysome-less cells can simply produce new ones, though their doubling time is significantly longer (30).

DabA & DabB – Inorganic carbon (C_i) transporters are essential to the CCM as active C_i accumulation powers the downstream action of the carboxysome (3, 8). A review by Price et al. summarizes five of the C_i uptake systems (8). In 2019, the DAB complex joined this list. DabA and DabB form a membrane-bound complex that appears to couple CO_2 transport into the cell to a cation gradient (18, 34, 35).

Rubisco activases – Rubisco is prone to inhibition by its substrate, RuBP, and other sugar derivatives. Rubisco activases catalyze release of this inhibitor. These enzymes are essential in plants and algae, but do not appear to be essential in carboxysome-containing bacteria (36, 37). The activases are divergent in the two carboxysomal lineages: α-lineages contain activase CbbQ and associated protein CbbO, while β-lineages contain β-Rca. Through convergent but different mechanisms, both activases bind Rubisco and are likely targeted to the carboxysome (37–40).

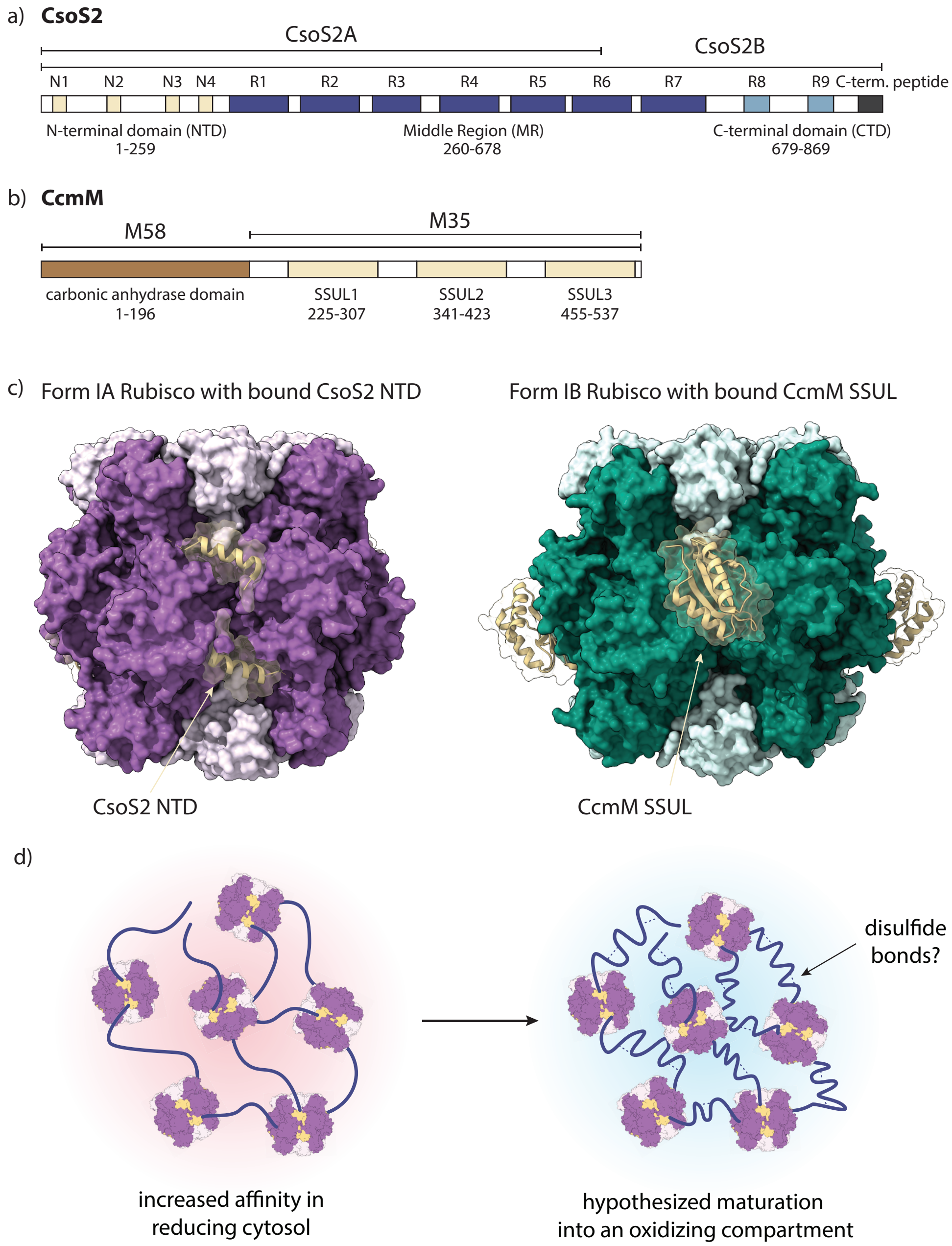


Figure 2. a) Domain structure of CsoS2 from *H. neapolitanus* (uniprot ID: O85041; CSOS2_HALNC), with marked short (CsoS2A) and long (CsoS2B) forms. b) Domain structure of CcmM from *S. elongatus* PCC7942 (uniprot ID: Q03513; CCMM_SYNE7), with marked short (M35) and long (M58) forms. SSUL stands for "small subunit-like" domain. c) Structures of the CsoS2 NTD bound to Form 1A Rubisco (PDB: 6UEW) and CcmM SSUL bound to Form 1B Rubisco (PDB: 6HBC). Structures were rendered in ChimeraX. d) Hypothesized model of carboxysome nucleation. The carboxysome nucleating protein binds Rubisco with high avidity and affinity in the reducing cytosol. Maturation may involve oxidation (or exclusion of reducing agents) and disulfide-bond induced conformational changes.

Though the biochemistry of these activases is increasingly understood, more research needs to be done to understand their role in carboxysomal carbon fixation.

1.5 Bioengineering the carboxysome and CCM

1.5.1 Optimizing the bacterial CCM

CCM-enhanced microbes could serve many bioindustrial applications seeking to take advantage of CO_2-dependent metabolism. New discoveries suggest optimization could start with Rubisco (Figure 3c). Fixation flux could, in theory, be improved via encapsulation of a faster Rubisco, many of which were recently discovered and characterized (10). In contrast to carboxysomal Form I Rubiscos, most of the fastest Rubiscos are Form II, and would need to be engineered for carboxysome targeting likely using CsoS2, CcmM, or other encapsulation peptides (41, 42). The carboxysome appears to be sensitive to the type of Rubisco it encapsulates - cells with an orthologous Form Iα Rubisco expressed in an α-carboxysome did not grow well in air, and replacement with a Form II lacked carboxysomes and required high CO_2 for growth (43, 44). A recombinant Rubisco may also require its cognate Rubisco activase to be expressed in the carboxysome.

To engineer a heterologous bacterial host to utilize a carboxysomal CCM, more genes are needed than just those in the major carboxysomal operon, which typically contains Rubisco, a carboxysomal nucleating protein, CA, and shells. A complete reconstitution of the *H. neapolitanus* CCM in *E. coli* required expression of a secondary operon alongside the major operon, thus enabling Rubisco-dependent *E. coli* to grow at atmospheric CO_2 (45). This secondary operon included the DAB inorganic carbon transporter, the CbbO and CbbQ Rubisco activase complex, and acRAF, a proposed Rubisco chaperone (46), along with several other unknown ORFs. Of these, both the DAB and acRAF were shown to be essential CCM components in the native organism *H. neapolitanus* (18). The DAB complex and bicarbonate transporter SbtA are the only transporters that have been demonstrated to be active when expressed in a heterologous system (18, 47), making them useful candidates for CCM engineering. This successful reconstitution marks a substantial progression in our knowledge from carboxysome structure to CCM function.

Now that a first-principles study in *E. coli* has shown heterologous reconstitution of a carboxysomal CCM to be possible, expansion into industrial hosts is a logical next step (Figure 3a). Baumgart et al. expressed the *H. neapolitanus* carboxysome operon in the biotechnologically relevant bacterium *Corynebacterium glutamicum*, but carboxysomes were small and malformed (48). Despite this, there is precedent for recombinant bacterial microcompartments to express in many diverse bacterial species, as was demonstrated with expression of the Pdu compartment in over 6 different hosts (49). Introducing the CCM into eukaryotic hosts such as yeast may enable efforts to engineer bioindustrial strains that are better equipped to utilize carboxylation as part of a metabolic engineering strategy (50). In other cases, the CO_2-dependent growth behavior of autotrophic strains, such as *C. necator,* could be improved through introduction of a CCM (51). Finally, lower DNA payloads are advantageous for recombinant CCM engineering, and researchers are testing the limits of minimal carboxysome systems by eliminating unnecessary proteins or creating fusions (Figure

3d). Many of these minimal systems show structural integrity and Rubisco activity (52–55).

1.5.2 Optimizing the plant CCM

Much effort has gone towards creating bacterial CCMs in plants. Many agriculturally important C3 plants such as wheat and rice lack CCMs, and instead devote ~5% of leaf biomass to Rubisco (56), consuming large amounts of nitrogen in the process. Plants with engineered carboxysomal CCMs could theoretically increase yield while consuming far less nitrogen (57). Many groups have proposed how to engineer carboxysomal CCMs into plants, and readers are directed to cited papers for more in-depth details (58, 59). In general, the engineering milestones are as follows: (1) Insert bicarbonate transporters into the chloroplast inner membrane to raise the concentration of bicarbonate in the stroma, (2) Express carboxysomes in the

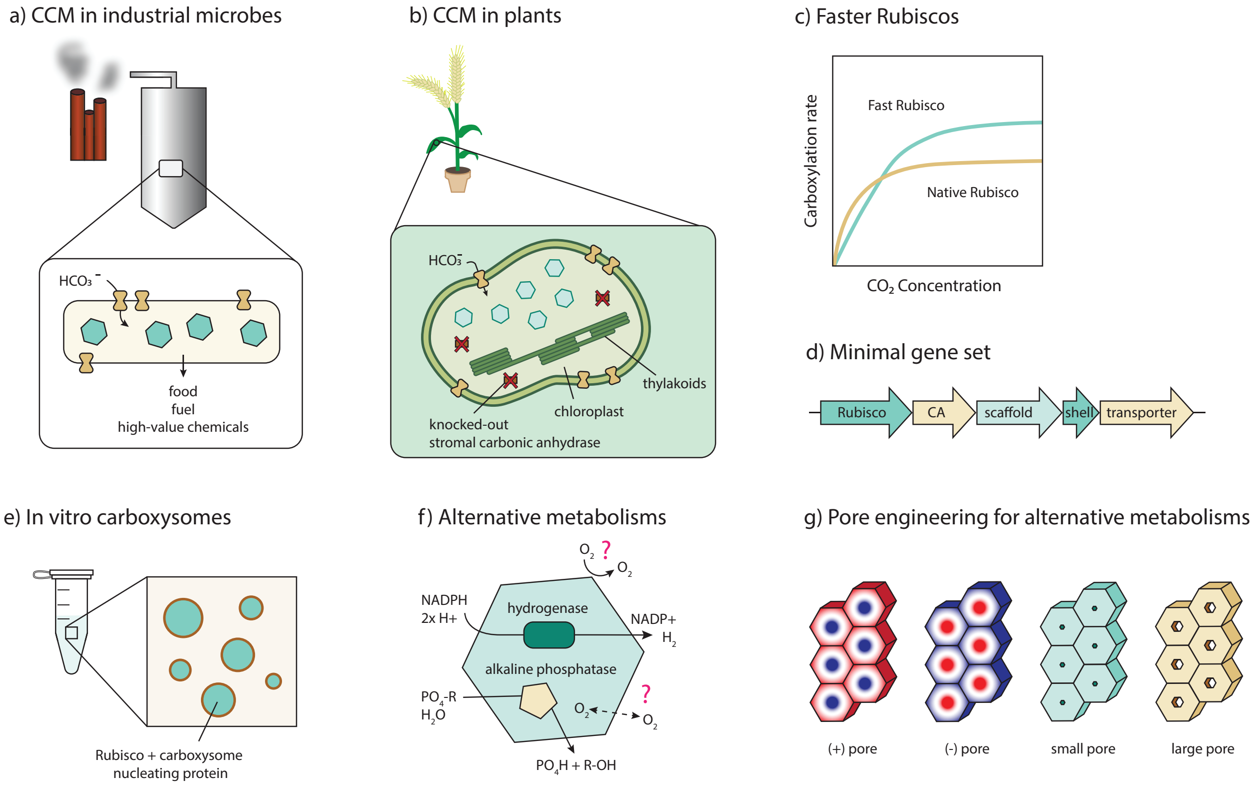


Figure 3. a) Engineering a carboxysomal CCM into industrial microbes could convert atmospheric CO_2 into high value products. b) Engineering a carboxysomal CCM into plants could increase plant CO_2 efficiency and promote growth and biomass yield. c) Faster Rubiscos could enable more efficient carbon fixation. d) A minimal gene set lowers the DNA payload when engineering the CCM into new host organisms. e) *In vitro* carboxysomes are a novel platform to study carboxysome assembly, and could act as *in vitro* catalytic reactors. f) Carboxysomes can be repurposed for alternative metabolisms. Enzymatic activity may depend on whether or not the carboxysome is an oxygen privileged environment, which remains unknown. g) Pore engineering, such as changing the charge or size of the pore, may aid development of alternative metabolisms.

chloroplast, and (3) Knock out stromal carbonic anhydrases in order to maintain a high ratio of HCO_3^- to CO_2 (Figure 3b).

Initial efforts towards this ultimate goal are already underway. Minimal α- and β-carboxysomes have been expressed in chloroplasts, and studies showed the formation of carboxysome-like structures (54, 60, 61). Cyanobacterial Rubiscos expressed in plants maintained kinetic properties equivalent to those of their native host (54, 61, 62). As expected due to lack of bicarbonate transporters, plants only grew under high CO_2 conditions, though with severe growth deficiencies compared to wild-type. Single-gene bicarbonate transporters BicA and SbtA have been expressed in the chloroplast inner envelope membrane, though it is unclear if they had activity (63, 64). Future efforts will thus need to focus on identifying, characterizing, and testing transporters that are capable of functional heterologous expression. A recent survey of dissolved inorganic carbon transporters in bacteria may provide useful candidates (65). It is also possible that additional components such as the partitioning proteins McdA and McdB will improve growth by ensuring even carboxysome distribution among dividing chloroplasts in leaf cells.

1.6 Future Directions

The relative simplicity of carboxysomes opens up the possibility of creating *in vitro* structures capable of performing carbon concentration, fixation, and other activities in order to understand and engineer function (Figure 3e). Rubisco and CsoS2 or CcmM readily form liquid-separated droplets *in vitro*, achieving the first step of cargo nucleation (20, 21). Chapter 2 of this thesis demonstrates the most complete reconstitution of an α-carboxysome as of this writing, with Rubisco, CsoS2, and shell forming liquid condensates under physiological salt concentrations (28). A logical next step is to show partitioning of other cargo proteins such as a CA and Rubisco activase into the droplets, followed by shell encapsulation. Experiments to test preferential partitioning of metabolites such as RuBP, HCO_3^-, or CO_2 into the light or dense phase could probe whether or not LLPS plays a role in metabolite transfer and CO_2 concentration.

New insights on carboxysome structure and assembly are now enabling them to be re-engineered for alternative metabolisms, a domain which has mostly been limited to other types of bacterial microcompartments (Figure 3f) (66). Li et al. expressed an [FeFe]-hydrogenase and ferredoxin in the α-carboxysome shell in *E. coli* to enhance H_2 production while shielding the hydrogenase from inactivating O_2. They observed an increase in H_2 in an aerobic environment compared to unencapsulated enzyme (42). This kind of study opens doors for exciting new biotechnological applications of carboxysomes and structures engineered from them, while continuing to shed light on basic carboxysome biology. In particular, this study suggests that the carboxysome is an O_2-excluding environment, a theory which has generated significant discussion (3, 15). In addition, shell protein pore engineering (Figure 3g) may continue to further enable novel metabolism, including even redox-based reactions, while also providing exciting new insights into how carboxysomes permit entry of substrates, exit of products, and restrict loss of intermediates.

1.7 Conclusion

Carboxysomes are unique among protein microcompartments for their ability to concentrate CO_2 and turn it into a useful cellular product. They are a biotechnologist's dream: they self-assemble in diverse organisms with a set of standard biological parts. They are, however, deceptively simple structures. Recent studies, including this thesis, reveal that we are still discovering many of the proteins and cellular chemistries necessary to build functional carboxysome-based CO_2-concentrating systems. These and future discoveries will prove crucial to making meaningful engineering advances.

1.8 Acknowledgements

The authors thank Cecilia Blikstad and Luke Oltrogge for thoughtful comments. The work was supported by a grant from the U.S. Department of Energy (DE-SC00016240).

1.9 References

1. Borden, J. S., and Savage, D. F. (2021) New discoveries expand possibilities for carboxysome engineering. *Curr. Opin. Microbiol.* **61**, 58–66
2. Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., Karl, D. M., Li, W. K. W., Lomas, M. W., Veneziano, D., Vera, C. S., Vrugt, J. A., and Martiny, A. C. (2013) Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 9824–9829
3. Mangan, N. M., Flamholz, A., Hood, R. D., Milo, R., and Savage, D. F. (2016) pH determines the energetic efficiency of the cyanobacterial CO_2 concentrating mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5354-62
4. Shively, J. M., Ball, F., Brown, D. H., and Saunders, R. E. (1973) Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of Thiobacillus neapolitanus. *Science*. **182**, 584–586
5. Espie, G. S., and Kimber, M. S. (2011) Carboxysomes: cyanobacterial RubisCO comes in small packages. *Photosynth. Res.* **109**, 7–20
6. Rae, B. D., Long, B. M., Badger, M. R., and Price, G. D. (2013) Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO_2 fixation in cyanobacteria and some proteobacteria. *Microbiol. Mol. Biol. Rev.* **77**, 357–379
7. Cabello-Yeves, P. J., Scanlan, D. J., Callieri, C., Picazo, A., Schallenberg, L., Huber, P., Roda-Garcia, J. J., Bartosiewicz, M., Belykh, O. I., Tikhonova, I. V., Torcello-Requena, A., De Prado, P. M., Millard, A. D., Camacho, A., Rodriguez-Valera, F., and Puxty, R. J. (2022) α-cyanobacteria possessing form IA RuBisCO globally dominate aquatic habitats. *ISME J.*
8. Price, G. D. (2011) Inorganic carbon transporters of the cyanobacterial CO_2 concentrating mechanism. *Photosynth. Res.* **109**, 47–57
9. Flamholz, A. I., Prywes, N., Moran, U., Davidi, D., Bar-On, Y. M., Oltrogge, L. M.,

Alves, R., Savage, D., and Milo, R. (2019) Revisiting Trade-offs between Rubisco Kinetic Parameters. *Biochemistry*. **58**, 3365–3376

10. Davidi, D., Shamshoum, M., Guo, Z., Bar-On, Y. M., Prywes, N., Oz, A., Jablonska, J., Flamholz, A., Wernick, D. G., Antonovsky, N., de Pins, B., Shachar, L., Hochhauser, D., Peleg, Y., Albeck, S., Sharon, I., Mueller-Cajar, O., and Milo, R. (2020) Highly active rubiscos discovered by systematic interrogation of natural sequence diversity. *EMBO J.* **39**, e104081
11. Marcus, Y., Schwarz, R., Friedberg, D., and Kaplan, A. (1986) High CO(2) Requiring Mutant of Anacystis nidulans R(2). *Plant Physiol.* **82**, 610–612
12. Price, G. D., and Badger, M. R. (1989) Isolation and Characterization of High CO(2)-Requiring-Mutants of the Cyanobacterium Synechococcus PCC7942 : Two Phenotypes that Accumulate Inorganic Carbon but Are Apparently Unable to Generate CO(2) within the Carboxysome. *Plant Physiol.* **91**, 514–525
13. Price, G. D., and Badger, M. R. (1989) Expression of Human Carbonic Anhydrase in the Cyanobacterium Synechococcus PCC7942 Creates a High CO(2)-Requiring Phenotype : Evidence for a Central Role for Carboxysomes in the CO(2) Concentrating Mechanism. *Plant Physiol.* **91**, 505–513
14. Cai, F., Menon, B. B., Cannon, G. C., Curry, K. J., Shively, J. M., and Heinhorst, S. (2009) The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO2 leakage barrier. *PLoS One*. **4**, e7521
15. Faulkner, M., Szabó, I., Weetman, S. L., Sicard, F., Huber, R. G., Bond, P. J., Rosta, E., and Liu, L.-N. (2020) Molecular simulations unravel the molecular principles that mediate selective permeability of carboxysome shell protein. *Sci. Rep.* **10**, 17501
16. Bonacci, W., Teng, P. K., Afonso, B., Niederholtmeyer, H., Grob, P., Silver, P. A., and Savage, D. F. (2012) Modularity of a carbon-fixing protein organelle. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 478–483
17. Fang, Y., Huang, F., Faulkner, M., Jiang, Q., Dykes, G. F., Yang, M., and Liu, L.-N. (2018) Engineering and Modulating Functional Cyanobacterial CO2-Fixing Organelles. *Front. Plant Sci.* **9**, 739
18. Desmarais, J. J., Flamholz, A. I., Blikstad, C., Dugan, E. J., Laughlin, T. G., Oltrogge, L. M., Chen, A. W., Wetmore, K., Diamond, S., Wang, J. Y., and Savage, D. F. (2019) DABs are inorganic carbon pumps found throughout prokaryotic phyla. *Nature Microbiology*. **4**, 2204–2215
19. Kerfeld, C. A., and Melnicki, M. R. (2016) Assembly, function and evolution of cyanobacterial carboxysomes. *Curr. Opin. Plant Biol.* **31**, 66–75
20. Wang, H., Yan, X., Aigner, H., Bracher, A., Nguyen, N. D., Hee, W. Y., Long, B. M., Price, G. D., Hartl, F. U., and Hayer-Hartl, M. (2019) Rubisco condensate formation by CcmM in β-carboxysome biogenesis. *Nature*. **566**, 131–135
21. Oltrogge, L. M., Chaijarasphong, T., Chen, A. W., Bolin, E. R., Marqusee, S., and Savage, D. F. (2020) Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation. *Nat. Struct. Mol. Biol.* **27**, 281–287
22. Ryan, P., Forrester, T. J. B., Wroblewski, C., Kenney, T. M. G., Kitova, E. N., Klassen, J. S., and Kimber, M. S. (2019) The small RbcS-like domains of the β -

carboxysome structural protein CcmM bind RubisCO at a site distinct from that binding the RbcS subunit. *J. Biol. Chem.* **294**, 2593–2603

23. Oltrogge, L. M., Chen, A. W., Chaijarasphong, T., Turnšek, J. B., and Savage, D. F. (2023) α-carboxysome size is controlled by the disordered scaffold protein CsoS2. *bioRxiv*. 10.1101/2023.07.07.548173
24. Chaijarasphong, T., Nichols, R. J., Kortright, K. E., Nixon, C. F., Teng, P. K., Oltrogge, L. M., and Savage, D. F. (2016) Programmed Ribosomal Frameshifting Mediates Expression of the α-Carboxysome. *J. Mol. Biol.* **428**, 153–164
25. Long, B. M., Tucker, L., Badger, M. R., and Price, G. D. (2010) Functional cyanobacterial beta-carboxysomes have an absolute requirement for both long and short forms of the CcmM protein. *Plant Physiol.* **153**, 285–293
26. Peña, K. L., Castel, S. E., de Araujo, C., Espie, G. S., and Kimber, M. S. (2010) Structural basis of the oxidative activation of the carboxysomal gamma-carbonic anhydrase, CcmM. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2455–2460
27. Chen, A. H., Robinson-Mosher, A., Savage, D. F., Silver, P. A., and Polka, J. K. (2013) The bacterial carbon-fixing organelle is formed by shell envelopment of preassembled cargo. *PLoS One*. **8**, e76127
28. Turnšek, J. B., Oltrogge, L. M., and Savage, D. F. (2023) Conserved and repetitive motifs in an intrinsically disordered protein drive α-carboxysome assembly. *bioRxiv*. 10.1101/2023.07.08.548221
29. Carpenter, W. B., Lavania, A. A., Borden, J. S., Oltrogge, L. M., Perez, D., Dahlberg, P. D., Savage, D. F., and Moerner, W. E. (2022) Ratiometric Sensing of Redox Environments Inside Individual Carboxysomes Trapped in Solution. *J. Phys. Chem. Lett.* **13**, 4455–4462
30. Savage, D. F., Afonso, B., Chen, A. H., and Silver, P. A. (2010) Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science*. **327**, 1258–1261
31. MacCready, J. S., Hakim, P., Young, E. J., Hu, L., Liu, J., Osteryoung, K. W., Vecchiarelli, A. G., and Ducat, D. C. (2018) Protein gradients on the nucleoid position the carbon-fixing organelles of cyanobacteria. *Elife*. 10.7554/eLife.39723
32. Garner, E. C., Campbell, C. S., and Mullins, R. D. (2004) Dynamic instability in a DNA-segregating prokaryotic actin homolog. *Science*. **306**, 1021–1025
33. MacCready, J. S., Tran, L., Basalla, J. L., Hakim, P., and Vecchiarelli, A. G. (2021) The McdAB system positions α-carboxysomes in proteobacteria. *Mol. Microbiol.* 10.1111/mmi.14708
34. Scott, K. M., Leonard, J. M., Boden, R., Chaput, D., Dennison, C., Haller, E., Harmer, T. L., Anderson, A., Arnold, T., Budenstein, S., Brown, R., Brand, J., Byers, J., Calarco, J., Campbell, T., Carter, E., Chase, M., Cole, M., Dwyer, D., Grasham, J., Hanni, C., Hazle, A., Johnson, C., Johnson, R., Kirby, B., Lewis, K., Neumann, B., Nguyen, T., Nino Charari, J., Morakinyo, O., Olsson, B., Roundtree, S., Skjerve, E., Ubaldini, A., and Whittaker, R. (2019) Diversity in CO2-Concentrating Mechanisms among Chemolithoautotrophs from the Genera Hydrogenovibrio, Thiomicrorhabdus, and Thiomicrospira, Ubiquitous in Sulfidic Habitats Worldwide. *Appl. Environ. Microbiol.* 10.1128/AEM.02096-18
35. Mangiapia, M., Mcb4404l, U., Brown, T.-R. W., Chaput, D., Haller, E., Harmer, T. L.,

Hashemy, Z., Keeley, R., Leonard, J., Mancera, P., Nicholson, D., Stevens, S., Wanjugi, P., Zabinski, T., Pan, C., and Scott, K. M. (2017) Proteomic and Mutant Analysis of the CO2 Concentrating Mechanism of Hydrothermal Vent Chemolithoautotroph Thiomicrospira crunogena. *J. Bacteriol.* 10.1128/JB.00871-16

36. Sutter, M., Roberts, E. W., Gonzalez, R. C., Bates, C., Dawoud, S., Landry, K., Cannon, G. C., Heinhorst, S., and Kerfeld, C. A. (2015) Structural Characterization of a Newly Identified Component of α-Carboxysomes: The AAA+ Domain Protein CsoCbbQ. *Sci. Rep.* **5**, 16243
37. Lechno-Yossef, S., Rohnke, B. A., Belza, A. C. O., Melnicki, M. R., Montgomery, B. L., and Kerfeld, C. A. (2020) Cyanobacterial carboxysomes contain an unique rubisco-activase-like protein. *New Phytol.* **225**, 793–806
38. Tsai, Y.-C. C., Lapina, M. C., Bhushan, S., and Mueller-Cajar, O. (2015) Identification and characterization of multiple rubisco activases in chemoautotrophic bacteria. *Nat. Commun.* **6**, 8883
39. Tsai, Y.-C. C., Ye, F., Liew, L., Liu, D., Bhushan, S., Gao, Y.-G., and Mueller-Cajar, O. (2020) Insights into the mechanism and regulation of the CbbQO-type Rubisco activase, a MoxR AAA+ ATPase. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 381–387
40. Flecken, M., Wang, H., Popilka, L., Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2020) Dual functions of a rubisco activase in metabolic repair and recruitment to carboxysomes. *Cell*. **183**, 457-473.e20
41. Kinney, J. N., Salmeen, A., Cai, F., and Kerfeld, C. A. (2012) Elucidating essential role of conserved carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly. *J. Biol. Chem.* **287**, 17729–17736
42. Li, T., Jiang, Q., Huang, J., Aitchison, C. M., Huang, F., Yang, M., Dykes, G. F., He, H.-L., Wang, Q., Sprick, R. S., Cooper, A. I., and Liu, L.-N. (2020) Reprogramming bacterial protein organelles as a nanoreactor for hydrogen production. *Nat. Commun.* **11**, 5448
43. Menon, B. B., Dou, Z., Heinhorst, S., Shively, J. M., and Cannon, G. C. (2008) Halothiobacillus neapolitanus carboxysomes sequester heterologous and chimeric RubisCO species. *PLoS One*. **3**, e3570
44. Pierce, J., Carlson, T. J., and Williams, J. G. (1989) A cyanobacterial mutant requiring the expression of ribulose bisphosphate carboxylase from a photosynthetic anaerobe. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 5753–5757
45. Flamholz, A. I., Dugan, E., Blikstad, C., Gleizer, S., Ben-Nissan, R., Amram, S., Antonovsky, N., Ravishankar, S., Noor, E., Bar-Even, A., Milo, R., and Savage, D. F. (2020) Functional reconstitution of a bacterial CO2 concentrating mechanism in Escherichia coli. *Elife*. 10.7554/eLife.59882
46. Wheatley, N. M., Sundberg, C. D., Gidaniyan, S. D., Cascio, D., and Yeates, T. O. (2014) Structure and identification of a pterin dehydratase-like protein as a ribulose-bisphosphate carboxylase/oxygenase (RuBisCO) assembly factor in the α-carboxysome. *J. Biol. Chem.* **289**, 7973–7981
47. Du, J., Förster, B., Rourke, L., Howitt, S. M., and Price, G. D. (2014) Characterisation of cyanobacterial bicarbonate transporters in E. coli shows that SbtA homologs are functional in this heterologous expression system. *PLoS One*. **9**,

e115905
48. Baumgart, M., Huber, I., Abdollahzadeh, I., Gensch, T., and Frunzke, J. (2017) Heterologous expression of the Halothiobacillus neapolitanus carboxysomal gene cluster in Corynebacterium glutamicum. *J. Biotechnol.* **258**, 126–135
49. Graf, L., Wu, K., and Wilson, J. W. (2018) Transfer and analysis of Salmonella pdu genes in a range of Gram-negative bacteria demonstrate exogenous microcompartment expression across a variety of species. *Microb. Biotechnol.* **11**, 199–210
50. Xia, P.-F., Zhang, G.-C., Walker, B., Seo, S.-O., Kwak, S., Liu, J.-J., Kim, H., Ort, D. R., Wang, S.-G., and Jin, Y.-S. (2017) Recycling Carbon Dioxide during Xylose Fermentation by Engineered Saccharomyces cerevisiae. *ACS synthetic biology [electronic resource].* **6**, 276–283
51. Ahrens, W., and Schlegel, H. G. (1972) Carbon dioxide requiring mutants of Hydrogenomonas eutropha strain H 16. I. Growth and Co2-fixation. *Arch. Mikrobiol.* **85**, 142–152
52. Gonzalez-Esquer, C. R., Shubitowski, T. B., and Kerfeld, C. A. (2015) Streamlined Construction of the Cyanobacterial CO2-Fixing Organelle via Protein Domain Fusions for Use in Plant Synthetic Biology. *Plant Cell.* **27**, 2637–2644
53. Frey, R., Mantri, S., Rocca, M., and Hilvert, D. (2016) Bottom-up Construction of a Primordial Carboxysome Mimic. *J. Am. Chem. Soc.* **138**, 10072–10075
54. Long, B. M., Hee, W. Y., Sharwood, R. E., Rae, B. D., Kaines, S., Lim, Y.-L., Nguyen, N. D., Massey, B., Bala, S., von Caemmerer, S., Badger, M. R., and Price, G. D. (2018) Carboxysome encapsulation of the CO2-fixing enzyme Rubisco in tobacco chloroplasts. *Nat. Commun.* **9**, 3570
55. Cai, F., Bernstein, S. L., Wilson, S. C., and Kerfeld, C. A. (2016) Production and Characterization of Synthetic Carboxysome Shells with Incorporated Luminal Proteins. *Plant Physiol.* **170**, 1868–1877
56. Bar-On, Y. M., and Milo, R. (2019) The global mass and average rate of rubisco. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 4738–4743
57. McGrath, J. M., and Long, S. P. (2014) Can the cyanobacterial carbon-concentrating mechanism increase photosynthesis in crop species? A theoretical analysis. *Plant Physiol.* **164**, 2247–2261
58. Price, G. D., Pengelly, J. J. L., Forster, B., Du, J., Whitney, S. M., von Caemmerer, S., Badger, M. R., Howitt, S. M., and Evans, J. R. (2013) The cyanobacterial CCM as a source of genes for improving photosynthetic CO2 fixation in crop species. *J. Exp. Bot.* **64**, 753–768
59. Hanson, M. R., Lin, M. T., Carmo-Silva, A. E., and Parry, M. A. J. (2016) Towards engineering carboxysomes into C3 plants. *Plant J.* **87**, 38–50
60. Lin, M. T., Occhialini, A., Andralojc, P. J., Devonshire, J., Hines, K. M., Parry, M. A. J., and Hanson, M. R. (2014) β-Carboxysomal proteins assemble into highly organized structures in Nicotiana chloroplasts. *Plant J.* **79**, 1–12
61. Lin, M. T., Occhialini, A., Andralojc, P. J., Parry, M. A. J., and Hanson, M. R. (2014) A faster Rubisco with potential to increase photosynthesis in crops. *Nature.* **513**, 547–550

62. Occhialini, A., Lin, M. T., Andralojc, P. J., Hanson, M. R., and Parry, M. A. J. (2016) Transgenic tobacco plants with improved cyanobacterial Rubisco expression but no extra assembly factors grow at near wild-type rates if provided with elevated CO2. *Plant J.* **85**, 148–160
63. Pengelly, J. J. L., Förster, B., von Caemmerer, S., Badger, M. R., Price, G. D., and Whitney, S. M. (2014) Transplastomic integration of a cyanobacterial bicarbonate transporter into tobacco chloroplasts. *J. Exp. Bot.* **65**, 3071–3080
64. Uehara, S., Adachi, F., Ito-Inaba, Y., and Inaba, T. (2016) Specific and efficient targeting of cyanobacterial bicarbonate transporters to the inner envelope membrane of chloroplasts in arabidopsis. *Front. Plant Sci.* **7**, 16
65. Scott, K. M., Harmer, T. L., Gemmell, B. J., Kramer, A. M., Sutter, M., Kerfeld, C. A., Barber, K. S., Bari, S., Boling, J. W., Campbell, C. P., Gallard-Gongora, J. F., Jackson, J. K., Lobos, A., Mounger, J. M., Radulovic, P. W., Sanson, J. M., Schmid, S., Takieddine, C., Warlick, K. F., and Whittaker, R. (2020) Ubiquity and functional uniformity in CO2 concentrating mechanisms in multiple phyla of Bacteria is suggested by a diversity and prevalence of genes encoding candidate dissolved inorganic carbon transporters. *FEMS Microbiol. Lett.* 10.1093/femsle/fnaa106
66. Kirst, H., and Kerfeld, C. A. (2019) Bacterial microcompartments: catalysis-enhancing metabolic modules for next generation metabolic and biomedical engineering. *BMC Biol.* **17**, 79

Chapter 2:

Conserved and repetitive motifs in an intrinsically disordered protein drive α-carboxysome assembly

2.1 Chapter Summary

All cyanobacteria and some chemoautotrophic bacteria fix CO_2 into sugars using specialized proteinaceous compartments called carboxysomes. Carboxysomes enclose the enzymes Rubisco and carbonic anhydrase inside a layer of shell proteins to increase the CO_2 concentration for efficient carbon fixation by Rubisco. In the α-carboxysome lineage, a disordered and highly repetitive protein named CsoS2 is essential for carboxysome formation and function. Without it, the bacteria require high CO_2 to grow. How does a protein lacking structure serve as the architectural scaffold for such a vital cellular compartment? In this chapter, we identify key residues present in the repeats of CsoS2, VTG and Y, which are necessary for building functional α-carboxysomes *in vivo*. These highly conserved and repetitive residues contribute to the multivalent binding interaction and phase separation behavior between CsoS2 and shell proteins. We also probe the role of conserved C doublets, which display an *in vitro*, but not *in vivo*, phenotype. Finally, we demonstrate 3-component reconstitution of CsoS2, Rubisco, and shell proteins into spherical condensates, and show the utility of reconstitution as a biochemical tool to study carboxysome biogenesis. The precise self-assembly of thousands of proteins is crucial for carboxysome formation, and understanding this process could enable their use in alternative biological hosts or industrial processes as effective tools to fix carbon.

Some text and figures in this chapter are adapted from the following manuscript in preparation: Turnšek, J. B., Oltrogge, L. M. and Savage, D. F. (2024 (likely)) Conserved and repetitive motifs in an intrinsically disordered protein drive α-carboxysome assembly. BioRxiv. (In review) (1).

2.2 Introduction

Carboxysomes are proteinaceous cellular microcompartments that are the metabolic centerpieces of the bacterial CO_2 concentrating mechanism. Each structure is >100 nm in diameter and encloses the enzymes carbonic anhydrase and Rubisco in an polyhedral-like shell, raising the luminal CO_2 concentration and driving Rubisco to operate at its maximum rate and specificity (2–5). There are two carboxysomal lineages that evolved convergently: α-carboxysomes, which emerged in proteobacteria and were horizontally transferred to α-cyanobacteria, and β-carboxysomes, which originated in β-cyanobacteria (6). In this work we focus on the α-carboxysomal lineage, using the proteobacterium *Halothiobacillus neapolitanus* as our model system (7–10).

All carboxysomes require five essential protein components: Rubisco, carbonic anhydrase, hexameric shell proteins, pentameric shell proteins, and a scaffold protein. Much is known about how the enzymatic and shell proteins function in the metabolism and structure of the carboxysome (4, 6, 11, 12), as well as how the β-carboxysome scaffolding protein CcmM drives carboxysome biogenesis within the β-lineage (13–16). Although both carboxysome lineages contain scaffolding proteins, these proteins are related in function alone; they have no sequence or structural similarity. In contrast to the β-lineage, how the α-carboxysome scaffolding protein directs α-carboxysome assembly is far less understood.

The α-carboxysome scaffolding protein, CsoS2, is highly conserved and is required for carboxysome assembly. CsoS2 knockouts cannot produce carboxysomes, rendering the bacteria incapable of growing at atmospheric CO_2 levels (0.04% CO_2) (Fig. S1) (17, 18). CsoS2 is a large ~90 kDa protein with three distinct domains (Fig. 1A) (19). The N-terminal domain (NTD) contains 4 alpha-helical repeats that bind Rubisco in a low-affinity and multivalent manner (20). The middle region (MR) has 7 distinct repeats, followed by the C-terminal domain (CTD) with 2 repeats and a highly conserved C-terminal peptide (CTP).

The CTD of CsoS2 binds to shell proteins, and has been successfully used as an encapsulation peptide for heterologous cargo (17, 21, 22). Recently, Ni and Jiang et al. obtained a cryo-EM structure of synthetic mini-carboxysomes with resolved CTD density, showing how it spans across shell-shell interfaces like a staple, reinforcing shell interactions as well as overall mini-carboxysome curvature and T number (23). Interestingly, CsoS2 has a short (CsoS2A) and long (CsoS2B) form produced by a ribosomal frameshifting site in the 6th MR repeat, effectively cutting off the CTD in the short form (24). While the long form is essential for carboxysome formation, the short form is not, further establishing the critical role of the CTD (25). In *H. neapolitanus* both CsoS2A and B are found in equimolar ratios within the carboxysome (17, 26).

Less well studied is the MR domain, which makes up almost 50% of the CsoS2 sequence. The 7 MR repeats have a number of intriguing highly conserved residues and sequence motifs. These were described previously (17, 27) and remain prominent in an up-to-date consensus sequence compiled from 272 de-replicated CsoS2 sequences in which each MR repeat was classified, extracted, and re-aligned against all other individual repeats. Four residues and motifs stand out in particular: (1) (V/I)(T/S)G triplets spaced 8 amino acids apart (hereafter VTG repeats), (2) cysteine pairs, (3) a highly conserved lysine, and (4) a highly conserved tyrosine (Fig. 1B and C).

In addition to this repeated motif structure, CsoS2 has a number of intrinsically disordered regions (IDRs) as identified by computational disorder predictors and corroborated by circular dichroism spectroscopy (20). CsoS2, like most IDPs, stymies AI structure prediction programs - AlphaFold yields a disordered coil and low confidence scores (Fig. 1D) (28). While it accurately depicted the known NTD alpha helices, it had medium to poor performance predicting the CTD structure, and the MR has never been resolved in either cryoEM or cryo-electron tomography (23, 29, 30).

How an intrinsically disordered protein directs the assembly of the α-carboxysome, and the role of the MR's highly conserved residues, remains unknown. In this study, we show that some, but not all, of these residues are essential for the growth of *H. neapolitanus* in air. These residues bind to shell proteins in a weak yet highly multivalent fashion and also facilitate the formation of biological condensates when mixed with shell *in vitro*, which may mimic *in vivo* assembly. The *in vitro* condensates show variable liquid properties between shell and CsoS2 under reducing conditions, suggesting a redox-modulated assembly strategy optimized for shell positioning and localization to the exterior of the carboxysome while minimizing the escape of CsoS2 and Rubisco throughout carboxysome biogenesis.

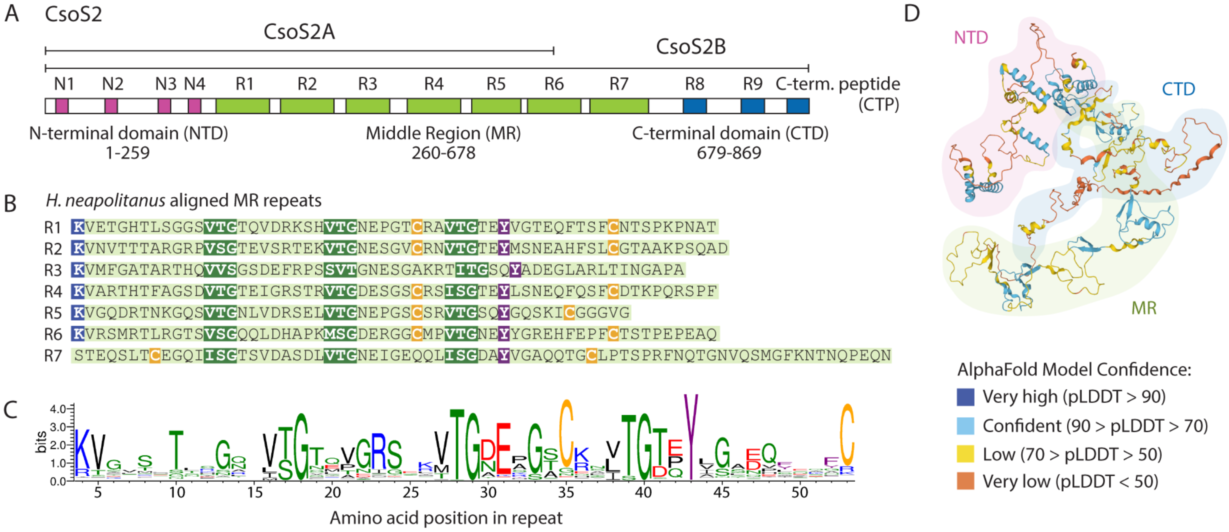


Figure 1. CsoS2 middle region (MR) contains highly conserved and repeated motifs with no known function. (A) Domain architecture of CsoS2. Repeats within domains are indicated by colored blocks. Amino acid numbering is specific to *H. neapolitanus* CsoS2. (B) Alignment of all MR repeats in *H. neapolitanus* CsoS2, with highly conserved motifs highlighted. (C) Sequence logo of the MR repeat generated from an alignment of 1662 MR repeats identified across 272 dereplicated CsoS2 sequences. Blue is basic, red is acidic, green is polar/small, black is hydrophobic, yellow is cysteine, purple is aromatic. (D) AlphaFold model of *H. neapolitanus* CsoS2 (UniProt O85041). pLDDT is AlphaFold's per-residue confidence score, which scales from 0 to 100.

2.3 Results

2.3.1 Identification of highly conserved motifs in the CsoS2 Middle Region that are essential for carboxysome assembly.

To probe the function of the highly conserved motifs in the MR repeats, we mutated these residues and assayed the growth of *H. neapolitanus* in air. *H. neapolitanus* needs carboxysomes to grow in air (0.04% CO_2), but not at higher CO_2 concentrations, enabling a selection system for deleterious CsoS2 variants. Because the MR has 7 repeats and binding may display complicated behavior, a series of

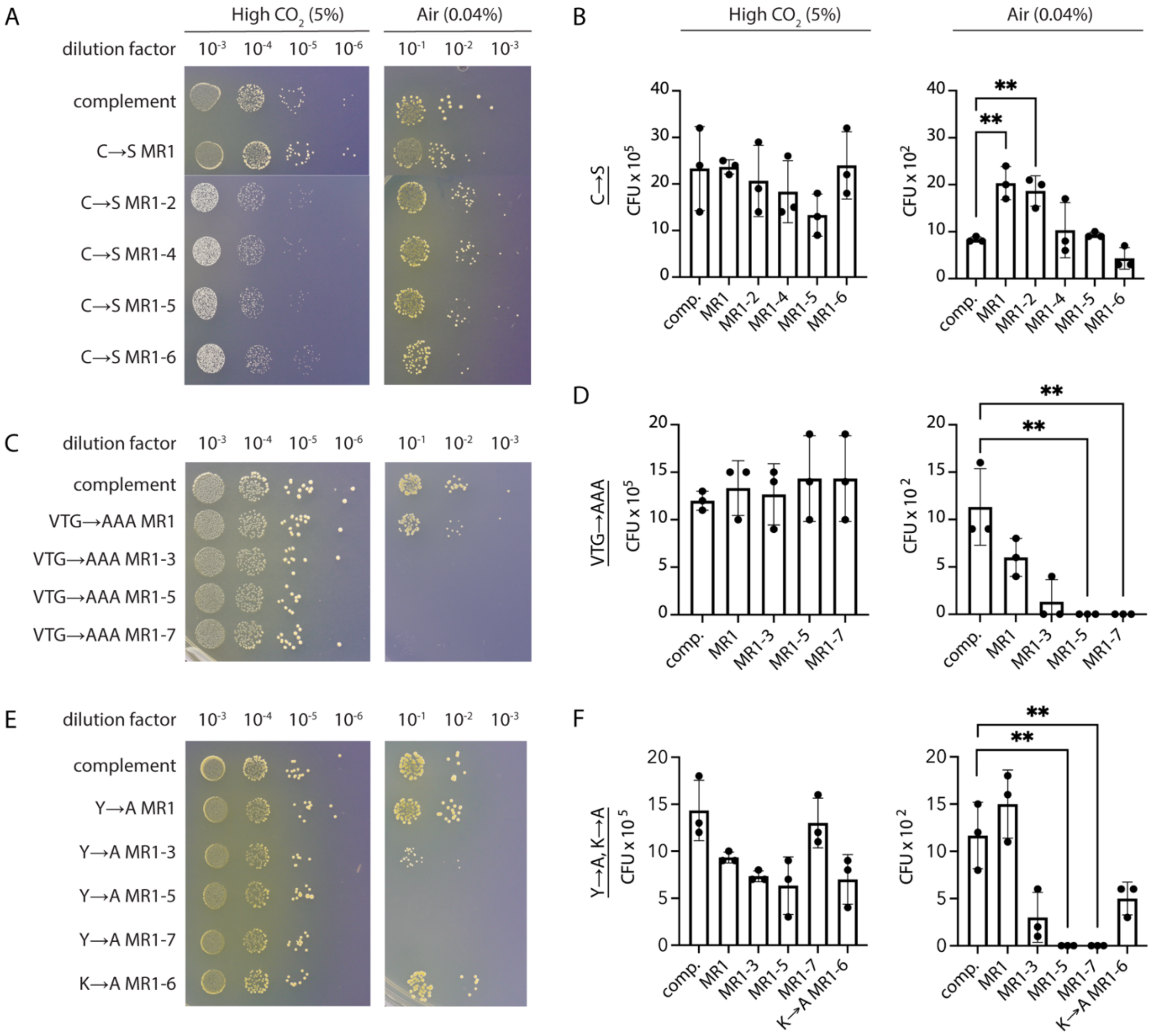


Figure 2. VTG and Y mutations significantly affect growth of *H. neapolitanus* in air. (A, C, E) Representative plates from dilution spotting assays of CsoS2 mutants in *H. neapolitanus*, grown in high CO_2 (5%) and air (0.04%). (B, D, F) Quantification of spotting assay results. Significance of ** is P ≤ 0.01 in an unpaired t-test. (A, B) C→S mutants, (C, D) VTG→AAA mutants, and (E, F) Y→A and K→A mutants. MR, Middle Region. CFU, colony forming units. Comp, complement.

mutants was generated until the entire MR was disrupted (Fig. S2). VTGs were mutated to AAA, Y to A, K to A, and C to S. All strains were generated by knocking out the genomic copy of CsoS2 and re-inserting a complement or mutated copy into a neutral site on the genome (Fig. S1 and S16). All strains expressed similar amounts of CsoS2 (Fig. S3), though it should be noted that only CsoS2B was detected; it is likely that expression from the neutral site instead of the native operon reduced ribosomal frameshifting responsible for the production of non-essential CsoS2A.

CsoS2 cysteine and lysine deletion strains showed no loss of growth in air (Fig. 2A, B and E, F), while VTG and tyrosine mutants showed a dramatic loss of growth (Fig. 2C, D and E, F). Growth was dependent on the number of VTG or Y motifs mutated, showing greater attenuation with more mutated repeats. No loss of growth among the cysteine mutants was quite unexpected due to the purported role of redox in carboxysome formation (14, 19) and the seemingly obvious disulfide-bonding function of the conserved cysteine pairs. Despite the growth of CsoS2 cysteine mutant strains of *H. neapolitanus*, CsoS2 cysteine mutant carboxysomes could not be purified from *E. coli* (Fig. S4), suggesting that cysteines play a non-essential structural role that strengthens the overall integrity of the complex, but may not be necessary for its assembly or function.

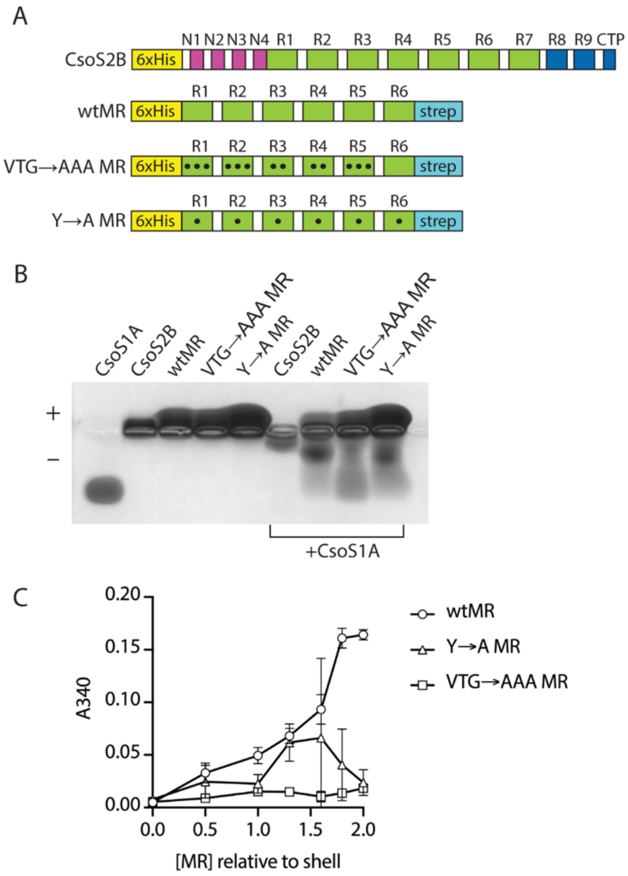


Figure 3. Shell protein CsoS1A binds to the MR of CsoS2, and mutations to VTG and Y residues perturb binding. (A) Purified constructs used in assays; all constructs have the ribosomal slip-site in R6 mutated to only produce the long form. Black circles indicate mutations within a repeat. Cartoons are not to scale with respect to domain sizes. (B) Native agarose protein gel of purified CsoS1A and CsoS2 variants. Quantification of this gel can be found in Fig. S5. (C) Turbidity at 10 minutes of the indicated constructs with defined molar ratios to shell (CsoS1A).

2.3.2 The CsoS2 Middle Region binds to shell proteins.

Dramatic loss of growth in VTG and Y mutant strains hinted that MR residues form interactions that are essential for carboxysome assembly. Previous studies showed full-length CsoS2 binds to shell protein CsoS1A (17), narrowing down candidate MR interaction partners to either shell proteins and/or CsoS2 itself. To biochemically assess the MR's binding interactions, we purified full-length CsoS2 and wild-type MR (wtMR) along with VTG and Y mutant variations of wtMR (Fig. 3A). In purifying wtMR, we wanted to identify MR interactions specifically since the NTD was already known to bind Rubisco (20) and CTD to shell (22, 23). Repeat 7 was left out of the wtMR construct because it occurs after the ribosomal slip site in Repeat 6, in an effort to eliminate potential confounding variables between the CsoS2A and CsoS2B isoforms.

A native agarose electrophoretic mobility shift assay revealed that both CsoS2 and wtMR bind to the hexameric shell protein CsoS1A, and that mutating VTG and Y perturbed this interaction (Fig. 3B and Fig. S5). CsoS1A showed a dramatic shift in mobility when mixed with full-length CsoS2B, and slightly less when mixed with wtMR. Mutating Y led to a subtle but even further decrease in shell mobility, while the mutant VTG construct showed the least binding to shell. We did not evaluate binding to Rubisco or CsoSCA, since it was previously shown that the MR does not interact with either of these proteins (20, 31).

Seeing the binding differences in this qualitative assay, we sought to further investigate the nature of the MR binding interaction. Interactions formed by disordered and/or repetitive proteins can often be monitored by a change in turbidity, which measures the transition from a soluble protein state to phase-separated condensate states. When more wtMR was added to CsoS1A, turbidity increased in a concentration-dependent manner, while no turbidity was observed for each protein alone (Fig. 3C and Fig. S6). The mutant VTG construct had little to no turbidity at any concentration, while the mutant Y construct showed intermediate behavior. Full-length CsoS2 displayed almost 5x more turbidity when mixed with CsoS1A compared to wtMR, reconfirming the robust contribution of the CTD to shell binding (Fig. S6) (21). Taken together with Fig. 3B, these results demonstrate that the VTG and Y residues participate in shell binding to the MR, yet they may contribute to the interaction in distinct ways.

2.3.3 Formation of phase-separated condensates is dependent on the CsoS2 Middle Region sequence.

Following the results of the turbidity assay, we confirmed via fluorescence microscopy that purified CsoS2 and wtMR indeed form phase-separated condensates when mixed with CsoS1A (Fig. 4A). However, when MR with mutated VTG or Y residues was mixed with CsoS1A, no condensates formed under these buffer conditions. All experiments were performed at 150 mM salt, mimicking typical intracellular conditions, and no condensates appeared when either MR or CsoS1A were observed on their own (Fig. S10). This suggests that these residues may be critical for forming low-affinity, highly multivalent interactions that drive CsoS2 and shell to phase separate during carboxysome biogenesis.

Interestingly, though both CsoS2 and wtMR formed phase separated condensates when mixed with shell, the condensates displayed distinct properties in their size and shape (Fig. 4B). Both condensates showed accumulated growth over 30 minutes (Figs. S7, S8, S9), but tended toward divergent shapes over the same growth period (Fig. S9). CsoS2 condensates were larger and more circular on average, while wtMR condensates were smaller and formed elongated structures (Fig. 4C, D, and E). A few additional protein and salt concentrations that were tested can be found in Fig. S18. Condensate shape can be an indicator of a liquid-to-solid phase transition, with liquid droplets often appearing more spherical and solid aggregates appearing more deformed or fibrillar (32, 33). The presence or absence of the CTD in the CsoS2 and wtMR constructs is a proxy for CsoS2A and CsoS2B, suggesting that these two proteins may contribute differently to the physical properties of the nascent carboxysome.

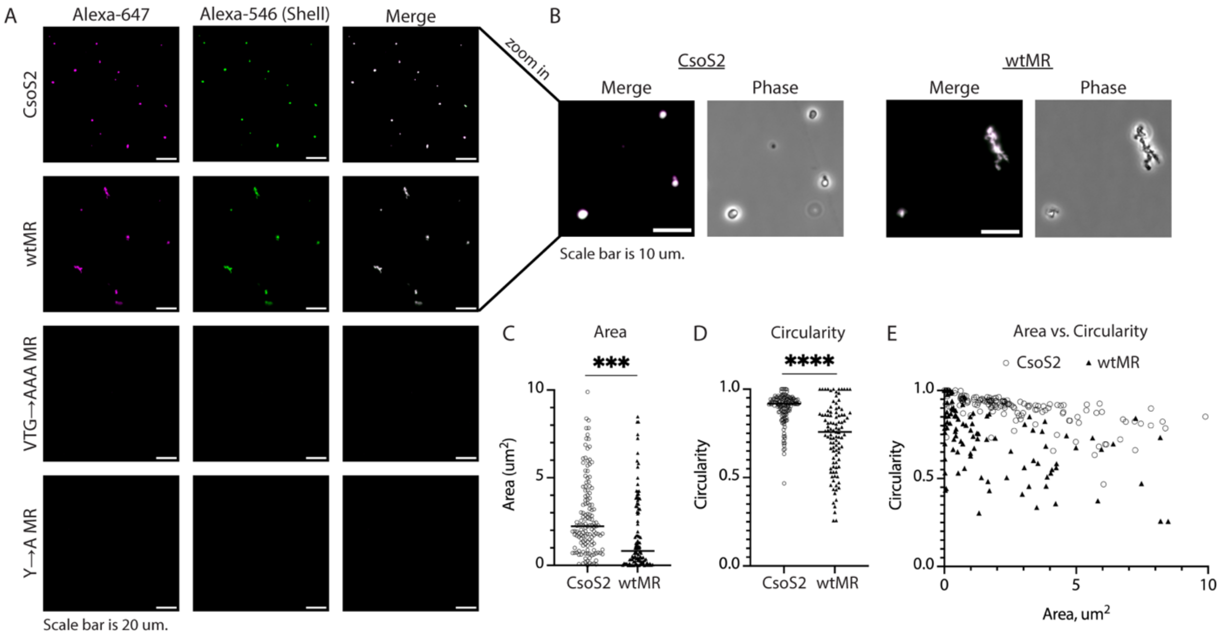


Figure 4. CsoS2 and wtMR form condensates when mixed with CsoS1A, but not when key residues are mutated. (A) Fluorescence microscopy of the indicated CsoS2 / MR protein variants with added CsoS1A, imaged at 30 minutes post mixing. All CsoS2 / MR variants are labeled in pink, CsoS1A is labeled in green, and the merge appears white at equally overlapping intensities. Scale bar is 20 µm. (B) Zoom-in of CsoS2 and wtMR droplets shown in (A). Scale bar is 10 µm. (C) Comparison of droplet area between CsoS2 and wtMR condensates, measured at 30 minutes. Significance of *** is $P \leq 0.001$ in an unpaired t-test. (D) Comparison of droplet circularity between CsoS2 and wtMR condensates, measured at 30 minutes. Circularity is calculated as 4π*area/perimeter^2, with 1.0 being a perfect circle and lower values indicating increasing shape elongation. Significance of **** is $P \leq 0.0001$ in an unpaired t-test. (E) Area vs. Circularity for all measured CsoS2 and wtMR condensates. For (C) and (D), the median is indicated by a black line.

2.3.4 In vitro carboxysome reconstitution and condensate properties.

Since CsoS2 and shell formed phase-separated condensates *in vitro*, we wanted to see if it was possible to fully reconstitute the carboxysome with its three major constituent components: CsoS2, shell, and Rubisco. The NTD of CsoS2 had been previously shown to form phase-separated condensates with Rubisco at low (20 mM) salt (20), but not yet demonstrated with full-length CsoS2 at physiological salt concentrations (150 mM). CsoS2 and Rubisco formed many small condensates at 5 minutes post mixing, but these condensates appeared to dissolve back into the soluble phase over 30 minutes (Fig. S11, S13). In contrast, when CsoS2, Rubisco, and shell (CsoS1A) were mixed, they formed robust condensates that grew significantly in size over 30 minutes (Fig. 5A and Fig. S12, S13). CsoSCA, the α-carboxysome carbonic anhydrase, forms phase separated condensates with Rubisco and the NTD of CsoS2 (Fig. S13.5), and would likely partition as a fourth component (34).

To better understand the 3-component condensate formation, CsoS2 and Rubisco were mixed and observed in a gasket on a microscope slide before adding shell at 10 minutes (Fig. S14). Condensates appeared within 30 seconds after the addition of shell. Larger droplets settled onto the focused plane of the slide over time, suggesting that droplet growth may occur via accretion of individual soluble components. There was also no significant difference in the average number of CsoS2-Rubisco-shell condensates per micrograph between 5 minutes and 30 minutes, further supporting a growth mechanism triggered and driven by the presence of shell (Fig. S13).

Interestingly, condensates were often observed to adhere next to one another without merging over time (Fig. 5B and D), a behavior that implied a more gel-like than liquid-like state (32). Protein liquidity in condensates can be sensitive to the solvent chemical environment, including to reducing / oxidizing (redox) conditions (14, 35, 36), which can alter local structure and dynamics via changes to reactive moieties like cysteine side chains.

We found that redox-dependent behavior was present in both intact carboxysomes and in reconstitutions. When carboxysomes, purified from either native or heterologous sources, were analyzed by SDS-PAGE under both reducing and

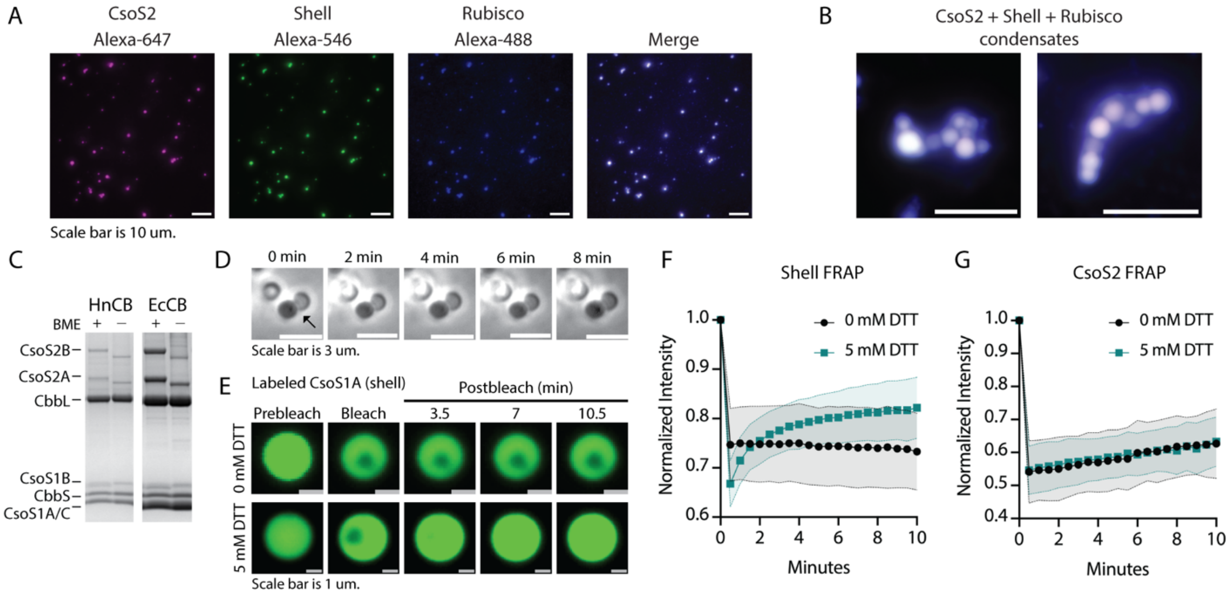


Figure 5. CsoS2, Shell, and Rubisco form condensates with liquid properties that differ in reducing vs. oxidizing conditions. (A) Individual channels of CsoS2 (pink) + shell (green) + Rubisco (blue) condensates, with the merge shown in white. Image was taken 30 minutes post mixing. (B) Zoomed-in examples of merged CsoS2 + shell + Rubisco condensates; scale bar is 10 µm. (C) PAGE gel of carboxysomes (CBs) purified from *H. neapolitanus* (HnCB) and *E. coli* (EcCB) with or without β-mercaptoethanol (BME). CsoS2A and B show distinct downshifts under oxidizing conditions. (D) Phase contrast of CsoS2 + shell + Rubisco condensates that do not merge over 8 minutes. (E) Example droplets from FRAP showing labeled CsoS1A with and without 5 mM DTT. (F) CsoS1A FRAP. (G) CsoS2 FRAP. (F, G) bleach occurs at ~30 seconds; black circles, 0 mM DTT; teal squares 5 mM DTT; normalized intensity, see Materials & Methods.

oxidizing conditions, CsoS2A and CsoS2B – and no other constituents – display a marked size shift, running as smaller under oxidizing conditions (Fig. 5C). Notably, this was a change in size from one homogenous species to another, suggesting that CsoS2 may undergo a specific redox-dependent structural change which could affect interactions between CsoS2 and its binding partners. To test this hypothesis, the liquidity of individual components was assessed using fluorescence recovery after photobleaching (FRAP). FRAP of CsoS2-Rubisco-shell condensates revealed that the shell experiences a dramatic difference in mobility depending on the redox conditions; it had no mobility under oxidizing conditions and recovered under reducing conditions (Fig. 5E and F). In contrast, CsoS2 showed nearly identical slow recovery under both oxidizing and reducing conditions (Fig. 5G). These results emphasize that the redox environment can independently modulate the mobility of proteins in carboxysome condensates, selectively tuning condensate properties and providing a window into how *in vivo* carboxysome assembly functions.

2.3.5 CsoS2 cysteines show biochemical effects in vitro

Although we were unable to detect physiological effects on growth in air in any of the C→S *in vivo* mutants, the cysteines still felt like tantalizing clues to a yet unobserved redox-dependent behavior of CsoS2. We purified a C→S MR variant with mutations in all repeats (except for repeat 3 which doesn't have cysteines) (Figure 6A). Running all the purified MR variants on a gel with and without reducing agent revealed three very unexpected results: 1) MR variants had a ladder banding pattern of higher molecular weight components, 2) these bands were unaffected by the presence of reducing agent, yet 3) these bands disappeared for the C→S MR variant (Figure 6B). Dimeric and trimeric bands were analyzed via mass spectrometry to identify the protein present, both in a VTG→AAA MR1 sample and a VTG→AAA MR1-6 sample. Unexpectedly, all four bands submitted were identified to be CsoS2, suggesting that the higher bands are oligomeric states of the MR (Table S3). Though oligomers did not collapse to monomers in the presence of reducing agent, both oligomeric and monomeric bands showed a small redox-dependent size shift, with a larger band in the reducing condition that likely indicates a more extended conformation.

Various conditions were tested to determine how the oligomers depended on the presence of cysteine, yet showed no response to reducing agents. A diverse panel of reducing agents was tested to see if oligomer reduction was chemical dependent. All had the same affect, showing the moderate shift to a higher size, the presence of oligomeric bands, and a marked absence of them for the C→S variant (Figure 6C). Increasing the concentration of reducing agent and including 8M urea also shockingly had no effect on oligomerization behavior, and adding urea on its own was comparable to the control sample (Figure 6D). Thinking that the MR oligomers might be kinetically trapped and inaccessible to reducing agent, the proteins were incubated for 21.5 hours prior to running the gel (Figure 6E). This similarly had no effect on oligomerization. Interestingly, the bands in Fig. 6D and E are all running higher than expected (~47-48 kDa vs. 39 kDa), which may be an effect of longer incubations of 1+ hours (1 hour for

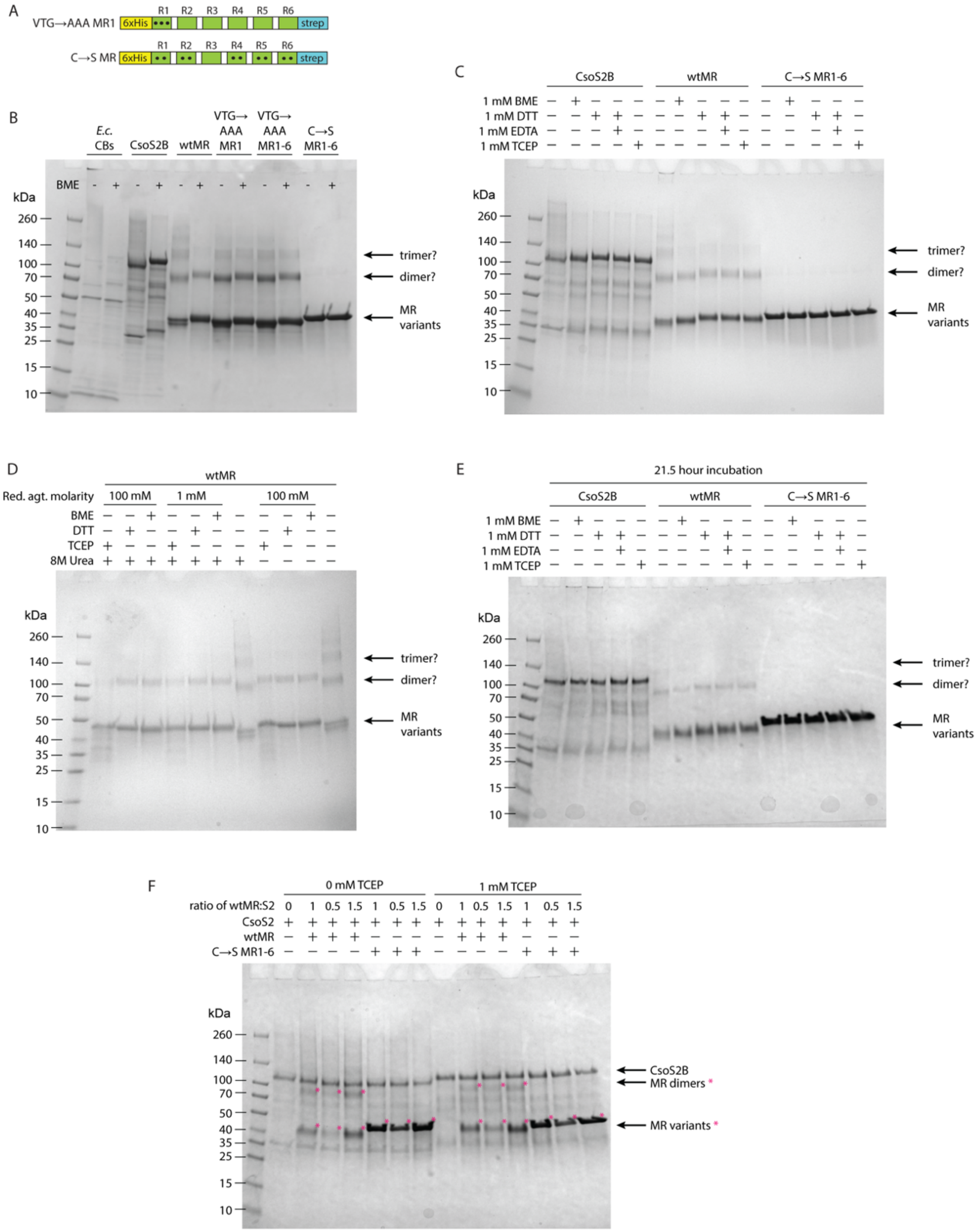


Figure 6. CsoS2 MR cysteines show biochemical phenotypes. (A) Additional MR constructs that were purified and used in the following assays, in addition to the constructs in Fig. 3. Note, there are no cysteines in R3. (B-F) In the following gels, all samples were boiled and run on denaturing SDS-PAGE gels. The expected molecular weight of wtMR is 38.6 kDa. The indicated proteins were run under the following conditions: (cont. on next page)

Fig. 6D and 21.5 hours for Fig. 6E) during which the protein may be unraveling to a maximally extended conformation. Oligomeric, SDS-resistant bands can be hallmarks of

Figure 6 caption, continued. (B) +/- 150 mM BME, 10 uM of each protein, 1 mg/ml CBs. (C) +/- 1 mM BME, DTT, EDTA, or TCEP, 5 uM of each protein. (D) +/- 1 or 100 mM BME, DTT, TCEP, and/or 8M urea with 1 hour pre-incubation, 5 uM of each protein. (E) +/- 1 mM BME, DTT, EDTA, or TCEP, with 21.5 hour incubation, 5 uM of each protein. (F) Mixed ratios of CsoS2, wtMR, and C→S MR +/- 1 mM TCEP, 5 uM of each protein. *E.c.* CBs, *E. coli* carboxysomes. BME, β-mercaptoethanol. DTT, dithiothreitol. EDTA, ethylenediamine tetraacetic acid. TCEP, tris(2-carboxyethyl)phosphine.

prions (37) (see discussion for more details). To test whether MR oligomers had prion-like behavior and the ability to recruit CsoS2 into an alternate conformation, CsoS2 and wtMR or C→S MR were mixed at various ratios with and without reducing agent. No new bands were observed, though perhaps a complementary experiment would be to repeat this but run the samples in a native agarose gel (commonly run for prions), where recruitment to an alternate conformation would look like a large smudge (37).

The oligomeric bands resisted denaturation when boiled, with added SDS, with added reducing agent, with added urea, and with all of these conditions combined. This led us to wonder whether irreversible covalent bonds may be forming. Higher-order oligomers in prion proteins, termed “subparticles”, have been observed previously (37, 38), as well as resistance of these oligomers to reducing agents (39). While reduction-resistant oligomers have been shown in prions with demonstrated cysteine disulfide bonding (39), they have also been observed in non-prion proteins with conserved cysteine residues, such as aquaporins (40, 41).

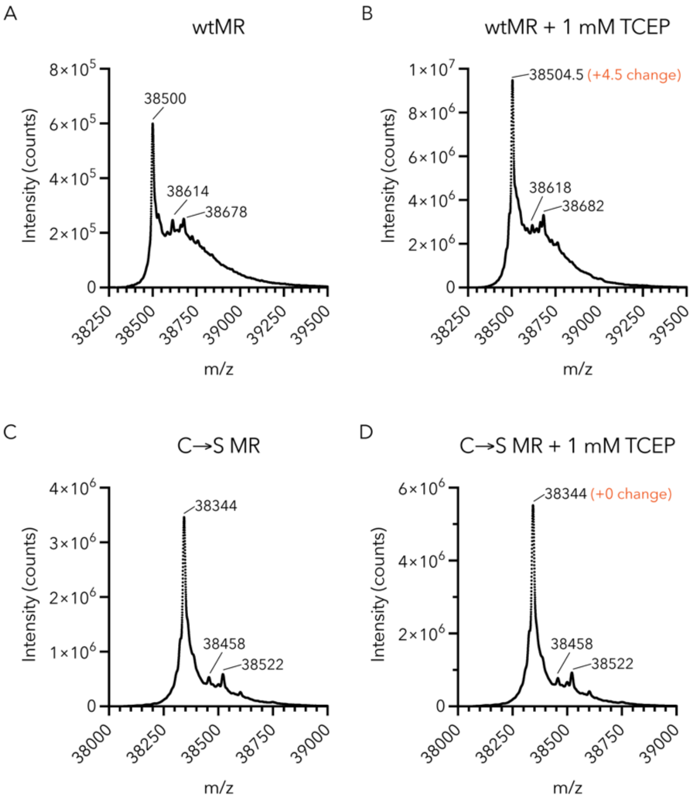


Figure 7. Intact mass spectrometry shows a highly specific m/z shift for reduced wtMR. (A) wtMR, (B) wtMR + 1 mM TCEP, (C) C→S MR, (D) C→S MR + 1 mM TCEP. Intact mass spec was performed by Tony Iavarone at QB3.

To test whether the oligomeric bands contained novel covalent bonds, we performed intact mass spectrometry on wtMR and C→S MR with and without reducing agent. No higher-order oligomers were observed in the results, but an unexpected mass shift did occur. The protein peak for wtMR showed a +4.5 change in m/z in the sample with added reducing agent compared to the control, while the C→S MR peak showed no change between the two samples (Figure 7). A +2 shift indicates a reduction of a disulfide bond, and one interpretation is that +4.5 is the reduction of two specific disulfide bonds. The other two peaks, +114 and +178, show no change between any of the samples, indicating that they are cysteine-independent modifications to the protein (+178 is likely a-N-gluconoylation, a common his-tag modification). If interpreted correctly, these results suggest that only four out of ten cysteines are forming highly specific disulfide bonds, and still leaves open the question of how the oligomers are forming so stably.

2.4 Discussion

Carboxysome assembly spans across length scales, from single amino acid interactions to thousands of proteins organizing themselves into a >300 megadalton compartment. In this work, we dissect each length scale to form a new model of how CsoS2 coordinates assembly of α-carboxysomes. With *in vivo* studies in the native host organism, we identified VTG and Y sequence motifs in the MR as essential for carboxysome assembly. These motifs appear to interact with the shell, an effect which is amplified by high valency across 7 MR repeats. Mutation of key residues *in vivo* and *in vitro* weakened this interaction. We further demonstrate that CsoS2, Rubisco, and shell can be reconstituted *in vitro* into spherical condensates, and that the liquidity of the shell can be tuned by the redox environment.

2.4.1 VTG and Y motifs and MR-shell binding

Overall, VTG and Y motifs contribute to many weak, transient interactions that en masse increase the binding affinity to shell proteins (Fig. 8A). It's not obvious at first glance how (V/I)(T/S)G motifs facilitate binding to the shell. A small step away to alanine abolished binding, suggesting important contributions from additional alkyl groups as well as the hydrogen bonding interactions from the hydroxyl group. In contrast, there is precedence for the importance of tyrosine residues in intrinsically disordered proteins (IDPs). Tyrosines often participate in pi-pi or cation-pi interactions with other aromatic or charged residues. IDPs with repetitive Y residues such as Fused in Sarcoma (FUS) showed reduced phase separation when greater numbers of Ys were mutated (42–44), similar to the MR.

The sticker and spacer model has emerged as a useful framework for understanding protein phase separation. In the model, proteins are divided into "sticker" regions responsible for intermolecular interactions and intervening disordered "spacers" (45). The "stickers" may range from single amino acid sidechains (such as the Ys in FUS) all the way to well-defined folded domains such as SSULs in the β-carboxysome scaffold protein CcmM (14). It remains to be seen where the MR lies along this spectrum, that is, whether an MR repeat acts as a single binding unit or a collection of short motifs, e.g. VTGs. A recently solved structure of the CTD bound to shell shows a well-conserved yet extended conformation (23). In contrast, modeling work from Oltrogge et al. (46) of MR repeat structure suggests a pseudo-threefold arrangement of the three VTG motifs that is mediated by a putative disulfide bond between the MR-specific cysteine residues. Moreover, this structure bears striking shape complementarity to the threefold axis found in shell-hexamer junctions and the VTGs could plausibly engage with CsoS1A-His79 in a manner similar to, but distinct from, the extended CTD conformation observed by Ni et al. (47). It is also possible that the highly conserved tyrosines of the MR repeats act as a stabilizing core for the VTG triads and make the MR repeats into a unitary interactor.

If MR binds to shell proteins, what is its role in the context of CsoS2B, which includes the additional shell binding CTD domain? Sequence – discussed above – along with valency and charge are three key differences. In *H. neapolitanus*, the MR has a valency of 7 repeats, while the CTD only has 2. Although these two repeats possess VTGs, they lack the conserved tyrosines, cysteines, and lysines (Fig. S15), and are also followed by the C-terminal peptide. The CTD has a pI of 9.4, making it positively charged at pH values close to 7 and promoting interaction with the negative shell luminal interface (21, 23). In contrast, the MR has a pI of 6.2. The roughly 2:1 ratio of MR to CTD (based on the 1:1 ratio of CsoS2A to CsoS2B) additionally amplifies the MR-shell interaction. However, too high of an A:B ratio may be detrimental to carboxysome formation - *in vitro*, MR and shell formed elongated condensates while CsoS2B and shell formed spherical condensates, and *in vivo*, CsoS2A alone is not sufficient to form carboxysomes (24). These differences likely act in concert to give the MR a mode of binding to the shell that is distinct from the CTD. Recent work from Oltrogge et al. proposes that the MR repeats bind areas of less shell curvature, i.e. the flat shell facets, while the CTD favors higher curvature associated with the vertices. In total, the balance of these interactions help set carboxysome size (46).

2.4.2 CsoS2 cysteines

A major unanswered question concerns the role of the CsoS2 cysteines in carboxysome formation. *In vivo* and *in vitro* data offer clashing viewpoints; there was no physiological phenotype when CsoS2 cysteines were mutated to serines, yet the C→S MR variant showed distinct, albeit confusing, biochemical phenotypes. It is unclear if the cysteine dependent multimeric assemblies of the MR have any biological function, or whether they are simply biochemical artifacts. If intermolecular disulfide bonding is not occurring, as suggested by the results of Figures 6 and 7, the only remaining distinction of a C→S mutation is the swapping of a sulfhydryl group for a hydroxyl group, which is more polar. If the cysteines have a purely polar role in the CsoS2-shell interaction, the change to serine would be quite subtle, and might explain the lack of an observed *in vivo* phenotype. Making C→A mutants would be a good future experiment to test this.

The oligomeric MR biochemical behavior also shows a similarity to the behavior of prions. Prions are proteins that have heritable phenotypes, allowing for “an unusual mechanism of information transfer that occurs via protein instead of nucleic acid” (48). Prions have a non-heritable conformation and a heritable conformation, the latter of which is capable of recruiting other copies of itself to assume the heritable conformation, usually leading to a phenotypic change. These heritable prion assemblies tend to form large, SDS-resistant bands on gels, which alerted me to the possibility of prion-like behavior in CsoS2. Prions are of course most famous for causing terrible brain diseases, so this discovery caused me to lose quite some sleep when I was doing these experiments. I reached out to Prof. Dan Jarosz, a prion specialist at Stanford, who assured me that transmission requires the same protein to be present in the host being infected, and that "because we don't have carboxysomes there is no risk to humans…BSL1 should be fine." Dave also confirmed that BSL1 was fine. And an abundance of studies on newly discovered prions is starting to re-define prions as essential organizing hubs for normal cellular processes (49). Still, it makes me queasy.

CsoS2 shows some behaviors that are consistent with the definition of a prion. Carboxysomes are heritable protein assemblies that define the physiology of the organism (50). One could argue that soluble protein components assume heritable conformations as they assemble into the carboxysome: shell proteins form sheets (30, 51, 52), Rubisco sometimes packs into fibrils (29, 30), and some evidence presented here suggests CsoS2 might assume an induced structure during the assembly process (Fig. 5C, Fig. 6, Fig. 7). Unbound copies of prion proteins are recruited into the bound conformation, leading to a phenotypic change. In fact, WT *H. neapolitanus* grown in high CO_2 expressed CsoS2 (Figure S3), yet it does not make carboxysomes (unpublished data from Luke Oltrogge), adding further intrigue to this theory. However, CsoS2 also lacks some classic prion identifiers: the heritable prion conformation is typically a self-templating amyloid (48), though not all prions show amyloid behavior (53). We looked at wtMR, VTG→AAA MR, and C→S MR variants on TEM, and none showed any distinct fibrillation (Figure S17). It is also debatable whether a large protein assembly such as the carboxysome should be defined as a prion, which are currently defined as single proteins, even though both represent protein-based and heritable modes of information transfer with downstream effects on cell physiology.

It should be noted that some non-prion proteins, such as certain aquaporins, display nearly identical biochemical behavior to the C→S MR variant as well (40, 41). While certain C→S mutations collapsed a dimer down to a monomer, others had no effect (41). The purified CsoS2 and MR proteins used in this study have all cysteines mutated to serines; it's possible that only some cysteines (perhaps four of them? See Fig. 7) are responsible for oligomer formation, and that they mask non-participating cysteines. Future experiments may want to probe exactly which cysteines are responsible for oligomer formation, though the combinatorics (10!/((10-4)!4!) = 210) make this experiment tricky.

2.4.3 Carboxysome assembly

The ability to study carboxysome assembly both *in vivo* and in *vitro* has many benefits, though it must be acknowledged that carboxysome condensates are not true carboxysomes. They are thousands of times larger in volume, do not contain all carboxysome components at exact ratios found *in vivo*, and lack the architectural organization of an outer shell layer and icosahedral shape. However, they are extremely useful as a proxy tool to biochemically interrogate protein interactions that are challenging to study *in vivo* (Fig. 8B). From condensates, we learn that CsoS2 and Rubisco are at the edge of solubility and weakly interact at physiological salt concentrations. Addition of shell leads to robust condensate formation, identifying the shell-CsoS2 interaction as the main driver of local phase separation. Notably, there is no evidence of organization in these condensate assays; we assume CsoS2, Rubisco, and shell are homogeneously mixed.

Extrapolating what this tells us about *in vivo* α-carboxysome formation (Fig. 8C), it is known that CsoS2, Rubisco, and shell are transcribed and translated from the same operon in distinct ratios (54, 55). It is thus likely that carboxysome proteins interact immediately during coincident expression. The exact details of carboxysome biogenesis remain uncertain. One pathway posits that these initial interactions drive local phase separation on a nanoscopic scale. Cryoelectron tomography evidence for this is mixed - Rubisco clusters have been observed in *Synechococcus* sp. WH8109 (56) but thus far not in *H. neapolitanus* (57). These same studies both only observe a maximum of one partially assembled carboxysome per cell, which suggests that nucleation of on-pathway assembly is kinetically limited, even if there are multiple phase separation events creating small Rubisco aggregates in the cytosol (Fig. 8C, local phase separation panel). Tomographic images of these partially assembled carboxysomes show one to four shell faces with observable packed Rubisco cargo, implying that shell crystallization - i.e. the formation of a 2D lattice - nucleates assembly from that site outwards. This further supports a model in which the shell-CsoS2 interaction is the main driver of assembly, rather than a Rubisco-centric model.

In the reduced cytosol, the shell is more mobile and forms interactions with both CsoS2 and itself from within the condensate and via outside accretion to organize an outer layer. At a certain volume it becomes thermodynamically favorable for shell proteins to fully encapsulate the carboxysome condensate, blocking additional growth and sealing off a functional carboxysome (57, 58). To kinetically trap carboxysome growth or dissolution at a precise size is perhaps even a role of the shell *in vivo*, in addition to concentrating CO_2. This is the key step where condensates and carboxysomes differ - it is still unknown what branches the completion of a 150 nm compartment from continued growth to a micron-sized particle. It might simply require tweaking of *in vitro* reaction conditions - salt concentration, protein concentration, molecular crowding, etc. - to tilt the preference towards smaller compartments. Future work will aim to not only establish the precise conditions to form nm-sized carboxysomes *in vitro*, but also to confirm that they can carboxylate CO_2.

2.5 Conclusion

Carboxysomes are a fascinating model system to understand how the coordinated actions of thousands of proteins build an essential cellular structure. Remarkably, the instructions for compartment assembly are encoded solely in the sequences of its constituent proteins. Here we establish that there is a molecular grammar to the CsoS2 MR sequence and how disruption of even a small number of residues diminishes binding to shell and prevents carboxysome formation. This work contributes new motifs to the growing dictionary of known sequence determinants of phase separation and microcompartment formation. Predicting whether a protein will phase separate and form a compartment, along with the conditions that affect this interaction, will continue to be important for informing broader efforts to engineer carboxysomes and other diverse microcompartments in biological systems.

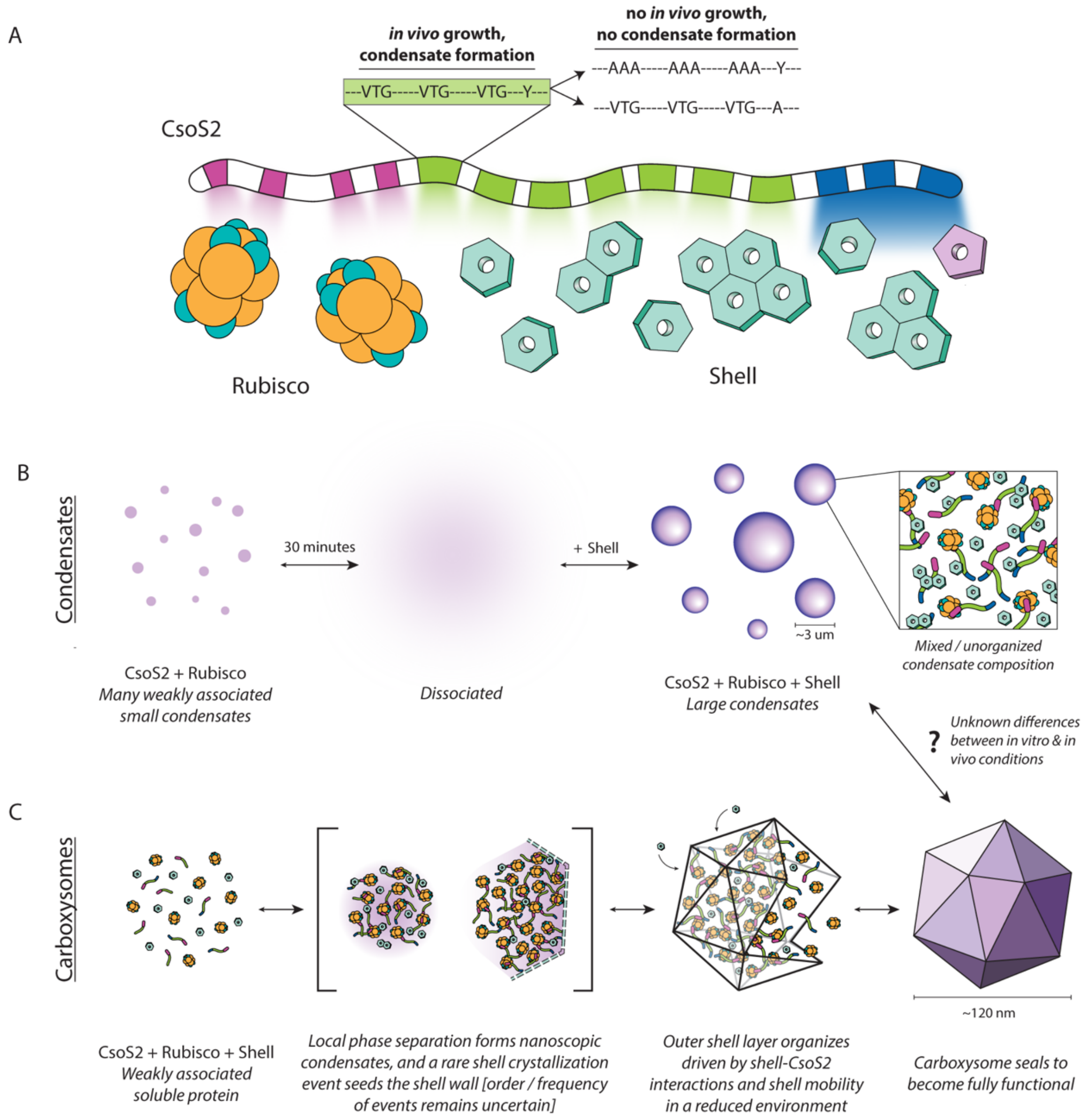


Figure 8. Model of CsoS2 interactions driving condensate and carboxysome assembly. (A) Cartoon model of known interactions between CsoS2 motifs and binding partners, with colored spotlights highlighting the relative strength of each interaction. Pink blocks are NTD repeats, green blocks are MR repeats, and blue blocks are CTD repeats and the CTP. The pullout box shows the sequence residues that are important for *in vivo* growth and condensate formation, and the mutated variants that do not grow or form condensates. (B) Model of condensate formation. Weak associations between CsoS2 and Rubisco tend towards dissolution at equilibrium but addition of shell precipitates large condensates. Condensates are assumed to be a mixture of all three proteins with no clear shell layer. (C) Model of *in vivo* carboxysome assembly, informed by condensate biochemistry and previous studies (52, 53). All carboxysome components phase separate locally on a nanoscopic scale, initiating a rare shell crystallization event under as of yet unknown conditions (see text for details). Shell-CsoS2 interactions drive assembly and organization, leading to the formation of sealed and functional compartments.

2.6 Materials & Methods

2.6.1 CsoS2 MR and CTD consensus sequences

All sequences in IMG matching the CsoS2 Pfam (PF12288) were downloaded in May 2020 for a total of 770 sequences. Partial sequences and those with ambiguous residue assignments were discarded, and the set was dereplicated to 95% protein sequence identity using usearch (59). These 272 remaining sequences were analyzed for peptide motifs using the MEME suite (60). The MR and CTD repeat motif positions were identified using MAST for a total of 2190. These repeat sequences were extracted with 15 aa of buffer on either side and then all aligned against each other, including both MR and CTD types, using mafft (61). FastTree was used to build a phylogenetic tree of all the repeats which clearly separated into two major clades: one with MR repeats and one with CTD repeats (62). A number of repeats had been misidentified by MAST as evidenced by their membership in the opposing clade. Notable among these is R7 from *H. neapolitanus* which the phylogeny strongly suggests is actually an MR repeat. The sequences belonging to the MR repeat clade (1662) and CTD repeat clade (528) were aligned again with mafft but this time only against members of their respective clades. Weblogo3 was used to create sequence logos for the two repeat classes from these alignments (63).

2.6.2 H. neapolitanus strain generation

Wild type (WT) *Halothiobacillus neapolitanus* is strain c2, ATCC 23641. The *ΔcsoS2* strain was made by homologous recombination of a spectinomycin resistance cassette into the native CsoS2 locus. Complement and mutant strains were generated by homologous recombination of the new CsoS2 sequence into a neutral site on the genome in the *ΔcsoS2* background strain. The insertion region corresponds to bases 2428660 - 2429201 on the genome. Plasmids were made by Golden Gate cloning into a neutral site destination vector. The neutral site vector contained the following features from 5'-3': *H. neapolitanus* upstream homology arm (bases 2428121 - 2428660), KanR, LacIQ, pTRC promoter, gene of interest (CsoS2), rrnB terminator, *H. neapolitanus* downstream homology arm (bases 2429201 - 2429703). All sequences contained an intact frame shifting site in CsoS2.

To transform *H. neapolitanus*, 10 ml of DSMZ-68 medium was inoculated per transformation and grown at 30°C and 5% CO_2. Cells were collected when the pH indicator had turned gray or light yellow indicating a pH of ~ 5.5-6.5 (1-2 days of growth). Cells were pelleted at 4000 xg for 10 minutes at 4°C and washed with cold milliQ water twice. After the third spin, cells were resuspended in 50-100 µl cold milliQ water. Cells were mixed with 500 ng of linearized plasmid and placed in cold electroporation cuvettes, then electroporated at 19 kV/cm, 200 mA, and 25 µF before immediate resuspension in 1 ml cold DSMZ-68 without antibiotic. Recovery occurred during overnight incubation at 30°C and 5% CO_2 before plating on DSMZ-68 agar plates containing the selection antibiotic. Colonies usually appeared after 3-4 days of growth at 30°C.

The genotype was confirmed via colony PCR and sequencing (Figure S16). For colony PCR, 1 colony (or part of one from a streaked plate) was resuspended in 10 ul of DSMZ. 1 ul from the 10 was used as genomic material for a colony PCR. The other 9 ul were inoculated into 6 ml of DSMZ with antibiotic for later glycerol storage if the colony showed a positive CsoS2 band. PCR primer sequences were gagtagggtacaaagtgtccacct and cgataaaaccgggcagcaaat. PCRs were performed with Q5 polymerase in 50 ul with a 3:55 extension time and 67° Tm. CsoS2 insertion bands appeared at 7820 base pairs and WT bands appeared at 2553 base pairs.

A list of *H. neapolitanus* strains with genotype, resistance, and induction can be found in Table S1.

2.6.3 H. neapolitanus selection assays

H. neapolitanus was inoculated from a colony on a plate into DSMZ-68 medium with the appropriate antibiotic (none for WT, 10 µg/ml spectinomycin for *ΔcsoS2*, and 10 µg/ml spectinomycin + 2 µg/ml kanamycin for complement / mutant strains) and 100 µM IPTG. Colonies were grown 1-2 days in 5% CO_2 until the medium had turned gray, which corresponded to OD600 ~0.1-3. Cells were washed with DSMZ-68 without pH indicator added to collect a more accurate OD600. All strains were normalized to OD600 of 0.1 and a 10x dilution series was generated. Strains were plated onto dry DSMZ-68 plates with the appropriate antibiotic and 100 µM IPTG and grown in either 5% CO_2 or air at 30°C.

2.6.4 Protein expression and purification

All proteins (6xHis-CsoS2B, 6xHis-wtMR-strep, 6xHis-(VTG→AAA MR)-strep, 6xHis-(VTG→AAA MR1)-strep, 6xHis-(C→S MR)-strep, 6xHis-(Y→A MR)-strep, 6xHis-Rubisco, and strep-CsoS1A) were individually cloned into pET-14-based destination vectors with ColE1 origin, T7 promoter, and carbenicillin resistance (See Table S2 for all protein sequences used in this study). All sequences are from *H. neapolitanus*. For the VTG→AAA MR construct, all VTG and VSG sites were mutated to AAA in repeats 1-5. For the VTG→AAA MR1 construct, all VTG and VSG sites were mutated to AAA in repeat 1 only. For the Y→A MR construct, all Y sites were mutated to A in repeats 1-6. For the C→S MR construct, all C sites were mutated to S in repeats 1-6. Plasmids were transformed into *E. coli* BL21-AI cells. Cells were grown in LB medium at 37°C with appropriate antibiotic until mid-log phase (OD_{600} of 0.3-0.5), at which point 0.2% L-arabinose was added to induce protein expression and the temperature lowered to 18°C. Cells were grown overnight before pelleting at 5000 xg the next day and freezing at -20°C.

Frozen cell pellets were thawed and resuspended in lysis buffer (50 mM Tris, 300 mM NaCl, 20 mM imidazole, pH 7.5) with the addition of 1 mM phenylmethanesulfonyl fluoride (PMSF), 0.1 µl/ml benzonase, and 0.1 mg/ml lysozyme. Cells were lysed on an

Avestin EmulsiFlex-C3 homogenizer and clarified at 27,000 xg for 45-60 minutes. All subsequent purification steps were performed at room temperature. Supernatant was added to a Ni-Sepharose resin in a gravity column, washed with wash buffer (50 mM HEPES, 300 mM NaCl, 60 mM imidazole, pH 7.5) and eluted with elution buffer (50 mM HEPES, 300 mM NaCl, 300 mM imidazole, pH 7.5). Proteins with a strep tag (His-wtMR-strep, His-(VTG→AAA MR)-strep, His-(Y→A MR)-strep) were further cleaned up on a Strep-Tactin resin on a gravity column. The entire elution was loaded onto the column, washed with wash buffer (50 mM HEPES, 300 mM NaCl) and eluted with elution buffer (50 mM HEPES, 300 mM NaCl, 2.5 mM d-desthiobiotin) before adding 10% glycerol, flash freezing in liquid N_2, and storing at -80°C.

CsoS2B was purified the same way through the His elution step, then further cleaned up using size exclusion chromatography. Eluted protein was loaded onto a HiPrep 16/60 Sephacryl S-200 HR column equilibrated in 50 mM HEPES, 300 mM NaCl buffer on an Akta Pure chromatography system. Fractions with full-length protein were concentrated on Amicon Ultra 15 Ultracel 30K filters before adding 10% glycerol, flash freezing in liquid N_2, and storing at -80°C.

Strep-CsoS1A was lysed and clarified the same way as above in 50 mM HEPES, 150 mM NaCl lysis buffer, then purified on a Strep-Tactin resin on a gravity column. Clarified lysate was loaded onto the column and allowed to flow through, followed by a wash step and elution with elution buffer (50 mM HEPES, 150 mM NaCl, 2.5 mM D-Desthiobiotin) before adding 10% glycerol, flash freezing in liquid N_2, and storing at -80°C.

6xHis-Rubisco was lysed and clarified the same way as above yet with a different lysis buffer (50 mM Tris, 150 mM NaCl, 20 mM imidazole, pH 7.5) and purified on a Ni-Sepharose resin on a gravity column the same way as above. Wash buffer was 50 mM Tris, 150 mM NaCl, 60 mM imidazole, pH 7.5. Elution buffer was 50 mM Tris, 150 mM NaCl, 300 mM imidazole, pH 7.5. Eluted protein was buffer exchanged on a 2 mL Zeba desalting column before adding 10% glycerol, flash freezing in liquid N_2, and storing at -80°C.

2.6.5 Carboxysome expression and purification

Carboxysomes were expressed in *E. coli* BW25113 off of the pHnCB10 plasmid (as described in Bonacci et al.(10)) with 500 µM IPTG induction at mid-log phase. Cells were grown overnight at 18°C and pelleted the next day. Carboxysomes were purified as described previously (20). Briefly, cell pellets were lysed using B-PER reagent with the addition of 1 mM PMSF, 0.1 µl/ml benzonase, and 0.1 mg/ml lysozyme. Lysis took place for 45 minutes at room temperature while shaking. Lysate was spun for 20 minutes at 12,000 xg, the supernatant collected, and then spun again for 30 minutes at 40,000 xg and the supernatant discarded. The pellet was resuspended in 200 µl TEMB (10 mM Tris, 10 mM MgCl2, 1 mM EDTA, pH 8.0) on ice with gentle rocking for 1 hour to overnight, with additional resuspension via pipette if needed. The resuspended pellet was clarified for 3 minutes at 1000 xg before loading onto a 5-step sucrose gradient (10, 20, 30, 40, and 50% w/v sucrose in TEMB). Gradients were spun for 15 minutes at

105,000 xg or longer depending on the size of the prep. The gradient was fractionated and analyzed with SDS-PAGE. Fractions containing carboxysomes were pooled and centrifuged for 30-90 minutes at 105,000 xg, resuspended in TEMB, and stored at 4°C. The SDS-PAGE gel in Figure 5C was run with 1% (v/v) β-mercaptoethanol (BME).

2.6.6 Native agarose protein gel

The gel was made from Tris-acetate-EDTA (TAE) buffer with 1% agarose. Samples were mixed (buffer: 50 mM HEPES, 150 mM NaCl) and cooled to room temperature over 45 minutes before adding native loading dye. Samples were not boiled. 5 µg of protein was loaded into each well for the controls; for mixed samples 5 µg of CsoS1A was added in addition. The gel was run for 60 minutes at 60 volts in native buffer (25 mM Tris, 200 mM Glycine). The gel was stained for 1 hour with Gel Code Blue, then destained with water until most of the stain had dissipated from the background. Gel quantification was done in FIJI.

2.6.7 Turbidity assays

Protein was thawed to room temperature before mixing. All samples were prepared to a final buffer composition of 50 mM HEPES, 150 mM NaCl. All samples contained 9 µM CsoS1A (except for the 0 µM shell sample). Concentrations of MR variants were: 0, 4.5, 9, 12, 14, 16, and 18 µM. The 0 µM shell control had 18 µM of MR. For CsoS2, concentrations tested were 4.5 and 14 µM. The 0 µM shell control had 14 µM of CsoS2. 40 µl were pipetted into a Nunc 384 well transparent plate and data collected on a Tecan Spark plate reader.

2.6.8 Condensate microscopy and quantification

Strep-CsoS1A (shell) was labeled with Alexa546 NHS Ester, at a ratio of 2x dye to hexamer. All CsoS2 and MR variants were labeled with Alexa647 NHS Ester, at a ratio of ⅙ dye to monomer. Rubisco was labeled with Alexa488 TFP Ester at a ratio of ⅓ dye to L8S8 hexadecamer. Prior to dyeing, Rubisco was buffer exchanged into 50 mM HEPES, 150 mM NaCl buffer on a Spin-X UF Corning 100K 0.5 ml filter tube. Labeling occurred for 1 hour in the dark at room temperature. Thermo Fluorescent Dye Removal Columns (#22858) were used to wash away the unconjugated dye, using an equal amount of resin to the volume of the sample. Proteins were thawed to room temperature before mixing in a PCR tube in a final buffer concentration of 50 mM HEPES, 150 mM NaCl. All proteins were at a concentration of 10 µM, except for the Rubisco+CsoS2+shell sample, which had 7.9 µM Rubisco, 6.1 µM CsoS2, and 17.5 µM shell. For the gasket experiment (Fig. S14), 2 µl of shell at 105 µM was added at the 10 minute mark to a 10 µl mix of 10 µM Rubisco and 10 µM CsoS2.

At 5 minutes and 30 minutes, a 1 µl sample was taken from the tube and pipetted onto a microscope slide (VWR micro cover glass 24x60mm No.1) and a coverslip added (VWR micro cover glass 24x30mm No.1). Samples were imaged on a Zeiss Axio

Observer Z1 inverted fluorescence microscope at 100x magnification with an oil immersion objective. The gasket in Fig. S14 is a Coverwell Perfusion Chamber 8x9mm diameter by 0.9mm depth (#622105). For gasket experiments, the final reaction volume was 12 ul. The Alexa546 channel appears as green, the Alexa647 channel appears as magenta, and the Alexa488 channel appears as blue.

All images were analyzed in FIJI. The intermodes thresholding algorithm was used to define droplets and make a mask before taking measurements. Condensates under 0.002 um^2 and over 10 um^2 were discarded due to false positives of misclassified droplets during the thresholding process.

2.6.9 FRAP measurements

FRAP experiments were done on a Leica STELLARIS 5 microscope with a white light laser. Each image was taken as a z-stack. A pre-bleach image was taken, and then droplets were bleached at 499, 557, and 653 nm at 40% laser intensity. A post-bleach timelapse took an image every 30 seconds for 10 minutes. For image analysis, images were first converted into average projections using the LASX microscope software, then further analyzed on FIJI. Drift correction was applied to each channel using StackReg (translation), then the background subtracted using a 50 pixel radius. FRAP measurements were analyzed using the method described in Guillén-Boixet et al. (64).

2.6.10 Western blots

5 ml of each strain were grown at 30°C and 5% CO_2 (except for WT which was grown in air) in DSMZ-68 medium with appropriate antibiotic and with or without 100 µM IPTG. At early log phase (indicated by grey or light-yellow pH indicator in the medium), cells were pelleted at 4000 xg for 10 minutes, the supernatant discarded, and frozen at -20°C for later analysis. For analysis, cells were thawed with 200 ml B-PER reagent, plus 1 mM PMSF, 0.1 µl/ml benzonase, and 0.1 mg/ml lysozyme (final concentrations). Lysis occurred over 45 minutes at room temperature while shaking. Samples were mixed with loading dye containing BME and boiled for 6 minutes. For the blot in (a), ~25 µg of protein was loaded per well (+/- 3 µg), for (b) 25 µg, and for (c) 50 µg (12.5 µg for WT). Samples were run on a Biorad TGX 4-20% gel for 40 minutes at 180 V. The gel was rinsed in water before transferring onto a PVDF membrane using a Biorad TransBlot Turbo for 10 minutes at 2.5A and 25 V. The membrane was blocked in TBST (50 mM Tris pH 7.5, 150 mM NaCl, 0.1% Tween-20, 5% rehydrated milk) overnight at 4°C while shaking. The next morning the buffer was replaced with 10 ml new TBST (2.5% milk) and primary antibody added (polyclonal rabbit antibodies ordered from GenScript). The antibody used for each blot is indicated in the figure. Both antibodies were added at a 1:2000 dilution. Blots were incubated with primary for 1 hour at room temperature, rinsed 3x with TBST, then incubated with secondary (Goat-HRP anti-Rabbit IgG) at 1:10000 dilution in TBST (1% milk) for 1 hour. Blots were rinsed 3x with TBST for 15 minutes before adding 12 ml of BioRad Clarity Western ECL Substrate, incubating for 5 minutes, and imaging.

2.6.11 SDS-PAGE gels

SDS-PAGE gels were BioRad 4-20% Criterion TGX Precast Midi Protein Gels, 18 well. Samples were mixed with the indicated reducing agents, allowed to incubate depending on the experiment, and then boiled for 7 minutes. Gels were run for 40 minutes at 180V. Gels were stained with GelCode Blue.

2.7 References

1. Turnšek, J. B., Oltrogge, L. M., and Savage, D. F. (2023) Conserved and repetitive motifs in an intrinsically disordered protein drive α-carboxysome assembly. *bioRxiv*. 10.1101/2023.07.08.548221
2. Marcus, Y., Berry, J. A., and Pierce, J. (1992) Photosynthesis and photorespiration in a mutant of the cyanobacterium Synechocystis PCC 6803 lacking carboxysomes. *Planta*. **187**, 511–516
3. Badger, M. R., and Price, G. D. (1992) The CO2concentrating mechanism in cyanobactiria and microalgae. *Physiol. Plant.* **84**, 606–615
4. Rae, B. D., Long, B. M., Badger, M. R., and Price, G. D. (2013) Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO2 fixation in cyanobacteria and some proteobacteria. *Microbiol. Mol. Biol. Rev.* **77**, 357–379
5. Mangan, N. M., Flamholz, A., Hood, R. D., Milo, R., and Savage, D. F. (2016) pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5354-62
6. Kerfeld, C. A., and Melnicki, M. R. (2016) Assembly, function and evolution of cyanobacterial carboxysomes. *Curr. Opin. Plant Biol.* **31**, 66–75
7. Shively, J. M., Ball, F., Brown, D. H., and Saunders, R. E. (1973) Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of Thiobacillus neapolitanus. *Science*. **182**, 584–586
8. Cannon, G. C., and Shively, J. M. (1983) Characterization of a homogenous preparation of carboxysomes from Thiobacillus neapolitanus. *Arch. Microbiol.* **134**, 52–59
9. English, R. S., Lorbach, S. C., Qin, X., and Shively, J. M. (1994) Isolation and characterization of a carboxysome shell gene from Thiobacillus neapolitanus. *Mol. Microbiol.* **12**, 647–654
10. Bonacci, W., Teng, P. K., Afonso, B., Niederholtmeyer, H., Grob, P., Silver, P. A., and Savage, D. F. (2012) Modularity of a carbon-fixing protein organelle. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 478–483
11. Kinney, J. N., Axen, S. D., and Kerfeld, C. A. (2011) Comparative analysis of carboxysome shell proteins. *Photosynth. Res.* **109**, 21–32
12. Liu, L.-N. (2021) Advances in the bacterial organelles for CO2 fixation. *Trends Microbiol.* 10.1016/j.tim.2021.10.004
13. Long, B. M., Tucker, L., Badger, M. R., and Price, G. D. (2010) Functional

cyanobacterial beta-carboxysomes have an absolute requirement for both long and short forms of the CcmM protein. *Plant Physiol.* **153**, 285–293

14. Wang, H., Yan, X., Aigner, H., Bracher, A., Nguyen, N. D., Hee, W. Y., Long, B. M., Price, G. D., Hartl, F. U., and Hayer-Hartl, M. (2019) Rubisco condensate formation by CcmM in β-carboxysome biogenesis. *Nature*. **566**, 131–135
15. Ryan, P., Forrester, T. J. B., Wroblewski, C., Kenney, T. M. G., Kitova, E. N., Klassen, J. S., and Kimber, M. S. (2019) The small RbcS-like domains of the β-carboxysome structural protein CcmM bind RubisCO at a site distinct from that binding the RbcS subunit. *J. Biol. Chem.* **294**, 2593–2603
16. Zang, K., Wang, H., Hartl, F. U., and Hayer-Hartl, M. (2021) Scaffolding protein CcmM directs multiprotein phase separation in β-carboxysome biogenesis. *Nat. Struct. Mol. Biol.* **28**, 909–922
17. Cai, F., Dou, Z., Bernstein, S. L., Leverenz, R., Williams, E. B., Heinhorst, S., Shively, J., Cannon, G. C., and Kerfeld, C. A. (2015) Advances in understanding carboxysome assembly in prochlorococcus and synechococcus implicate csos2 as a critical component. *Life*. **5**, 1141–1171
18. Desmarais, J. J., Flamholz, A. I., Blikstad, C., Dugan, E. J., Laughlin, T. G., Oltrogge, L. M., Chen, A. W., Wetmore, K., Diamond, S., Wang, J. Y., and Savage, D. F. (2019) DABs are inorganic carbon pumps found throughout prokaryotic phyla. *Nature Microbiology*. **4**, 2204–2215
19. Borden, J. S., and Savage, D. F. (2021) New discoveries expand possibilities for carboxysome engineering. *Curr. Opin. Microbiol.* **61**, 58–66
20. Oltrogge, L. M., Chaijarasphong, T., Chen, A. W., Bolin, E. R., Marqusee, S., and Savage, D. F. (2020) Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation. *Nat. Struct. Mol. Biol.* **27**, 281–287
21. Tan, Y. Q., Ali, S., Xue, B., Teo, W. Z., Ling, L. H., Go, M. K., Lv, H., Robinson, R. C., Narita, A., and Yew, W. S. (2021) Structure of a Minimal α-Carboxysome-Derived Shell and Its Utility in Enzyme Stabilization. *Biomacromolecules*. **22**, 4095–4109
22. Li, T., Jiang, Q., Huang, J., Aitchison, C. M., Huang, F., Yang, M., Dykes, G. F., He, H.-L., Wang, Q., Sprick, R. S., Cooper, A. I., and Liu, L.-N. (2020) Reprogramming bacterial protein organelles as a nanoreactor for hydrogen production. *Nat. Commun.* **11**, 5448
23. Ni, T., Jiang, Q., Ng, P. C., Shen, J., Dou, H., Zhu, Y., Radecke, J., Dykes, G. F., Huang, F., Liu, L.-N., and Zhang, P. (2023) Intrinsically disordered CsoS2 acts as a general molecular thread for α-carboxysome shell assembly. *bioRxiv*. 10.1101/2023.06.24.546370
24. Chaijarasphong, T., Nichols, R. J., Kortright, K. E., Nixon, C. F., Teng, P. K., Oltrogge, L. M., and Savage, D. F. (2016) Programmed Ribosomal Frameshifting Mediates Expression of the α-Carboxysome. *J. Mol. Biol.* **428**, 153–164
25. Chaijarasphong, T. (2016) *Towards an In Vitro Reconstitution of the Alpha-Carboxysome*. Ph.D. thesis, UC Berkeley, [online] https://escholarship.org/uc/item/6f39j9cf (Accessed July 6, 2023)
26. Baker, S. H., Lorbach, S. C., Rodriguez-Buey, M., Williams, D. S., Aldrich, H. C.,

and Shively, J. M. (1999) The correlation of the gene csoS2 of the carboxysome operon with two polypeptides of the carboxysome in thiobacillus neapolitanus. *Arch. Microbiol.* **172**, 233–239

27. Cannon, G. C., Baker, S. H., Soyer, F., Johnson, D. R., Bradburne, C. E., Mehlman, J. L., Davies, P. S., Jiang, Q. L., Heinhorst, S., and Shively, J. M. (2003) Organization of carboxysome genes in the thiobacilli. *Curr. Microbiol.* **46**, 115–119
28. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Žídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D., and Velankar, S. (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444
29. Metskas, L. A., Ortega, D., Oltrogge, L. M., Blikstad, C., Lovejoy, D. R., Laughlin, T. G., Savage, D. F., and Jensen, G. J. (2022) Rubisco forms a lattice inside alpha-carboxysomes. *Nat. Commun.* 10.1038/s41467-022-32584-7
30. Evans, S. L., Al-Hazeem, M. M. J., Mann, D., Smetacek, N., Beavil, A. J., Sun, Y., Chen, T., Dykes, G. F., Liu, L.-N., and Bergeron, J. R. C. (2023) Single-particle cryo-EM analysis of the shell architecture and internal organization of an intact α-carboxysome. *Structure*. **31**, 677-688.e4
31. Blikstad, C., Dugan, E. J., Laughlin, T. G., Liu, M. D., Shoemaker, S. R., Remis, J. P., and Savage, D. F. (2021) Discovery of a carbonic anhydrase-Rubisco supercomplex within the alpha-carboxysome. *bioRxiv*. 10.1101/2021.11.05.467472
32. Alberti, S., Gladfelter, A., and Mittag, T. (2019) Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell*. **176**, 419–434
33. Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A. P., Kim, H. J., Mittag, T., and Taylor, J. P. (2015) Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell*. **163**, 123–133
34. Blikstad, C., Dugan, E. J., Laughlin, T. G., Turnšek, J. B., Liu, M. D., Shoemaker, S. R., Vogiatzi, N., Remis, J. P., and Savage, D. F. (2023) Identification of a carbonic anhydrase–Rubisco complex within the alpha-carboxysome. *Proceedings of the National Academy of Sciences*. **120**, e2308600120
35. Kato, M., Yang, Y.-S., Sutter, B. M., Wang, Y., McKnight, S. L., and Tu, B. P. (2019) Redox State Controls Phase Separation of the Yeast Ataxin-2 Protein via Reversible Oxidation of Its Methionine-Rich Low-Complexity Domain. *Cell*. **177**, 711-721.e8
36. Huang, X., Chen, S., Li, W., Tang, L., Zhang, Y., Yang, N., Zou, Y., Zhai, X., Xiao, N., Liu, W., Li, P., and Xu, C. (2021) ROS regulated reversible protein phase separation synchronizes plant flowering. *Nat. Chem. Biol.* **17**, 549–557
37. Bagriantsev, S. N., Kushnirov, V. V., and Liebman, S. W. (2006) Analysis of Amyloid Aggregates Using Agarose Gel Electrophoresis. in *Methods in Enzymology*, pp. 33–48, Academic Press, **412**, 33–48
38. Bagriantsev, S., and Liebman, S. W. (2004) Specificity of prion assembly in vivo.

[PSI+] and [PIN+] form separate structures in yeast. *J. Biol. Chem.* **279**, 51042–51048
39. Welker, E., Raymond, L. D., Scheraga, H. A., and Caughey, B. (2002) Intramolecular VersusIntermolecular Disulfide Bonds in Prion Proteins*. *J. Biol. Chem.* **277**, 33477–33481
40. Kirscht, A., Survery, S., Kjellbom, P., and Johanson, U. (2016) Increased Permeability of the Aquaporin SoPIP2;1 by Mercury and Mutations in Loop A. *Front. Plant Sci.* **7**, 1249
41. Bienert, G. P., Cavez, D., Besserer, A., Berny, M. C., Gilis, D., Rooman, M., and Chaumont, F. (2012) A conserved cysteine residue is involved in disulfide bond formation between plant plasma membrane aquaporin monomers. *Biochem. J.* **445**, 101–111
42. Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., Mirzaei, H., Goldsmith, E. J., Longgood, J., Pei, J., Grishin, N. V., Frantz, D. E., Schneider, J. W., Chen, S., Li, L., Sawaya, M. R., Eisenberg, D., Tycko, R., and McKnight, S. L. (2012) Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell*. **149**, 753–767
43. Lin, Y., Currie, S. L., and Rosen, M. K. (2017) Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. *J. Biol. Chem.* **292**, 19110–19120
44. Wang, J., Choi, J.-M., Holehouse, A. S., Lee, H. O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A., Drechsel, D., Poser, I., Pappu, R. V., Alberti, S., and Hyman, A. A. (2018) A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell*. **174**, 688-699.e16
45. Ginell, G. M., and Holehouse, A. S. (2023) An Introduction to the Stickers-and-Spacers Framework as Applied to Biomolecular Condensates. in *Phase-Separated Biomolecular Condensates: Methods and Protocols* (Zhou, H.-X., Spille, J.-H., and Banerjee, P. R. eds), pp. 95–116, Springer US, New York, NY
46. Oltrogge, L. M., Chen, A. W., Chaijarasphong, T., Turnšek, J. B., and Savage, D. F. (2023) α-carboxysome size is controlled by the disordered scaffold protein CsoS2. *bioRxiv*. 10.1101/2023.07.07.548173
47. Ni, T., Jiang, Q., Ng, P. C., Shen, J., Dou, H., Zhu, Y., Radecke, J., Dykes, G. F., Huang, F., Liu, L.-N., and Zhang, P. (2023) Intrinsically disordered CsoS2 acts as a general molecular thread for α-carboxysome shell assembly. *Nat. Commun.* **14**, 5512
48. Brown, J. C. S., and Lindquist, S. (2009) A heritable switch in carbon source utilization driven by an unusual yeast prion. *Genes Dev.* **23**, 2320–2332
49. Jakobson, C. M., and Jarosz, D. F. (2018) Organizing biochemistry in space and time using prion-like self-assembly. *Current Opinion in Systems Biology*. **8**, 16–24
50. Savage, D. F., Afonso, B., Chen, A. H., and Silver, P. A. (2010) Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science*. **327**, 1258–1261
51. Faulkner, M., Rodriguez-Ramos, J., Dykes, G. F., Owen, S. V., Casella, S., Simpson, D. M., Beynon, R. J., and Liu, L.-N. (2017) Direct characterization of the

native structure and mechanics of cyanobacterial carboxysomes. *Nanoscale*. **9**, 10662–10673

52. Tsai, Y., Sawaya, M. R., Cannon, G. C., Cai, F., Williams, E. B., Heinhorst, S., Kerfeld, C. A., and Yeates, T. O. (2007) Structural analysis of CsoS1A and the protein shell of the Halothiobacillus neapolitanus carboxysome. *PLoS Biol.* **5**, e144
53. Chakravarty, A. K., Smejkal, T., Itakura, A. K., Garcia, D. M., and Jarosz, D. F. (2020) A Non-amyloid Prion Particle that Activates a Heritable Gene Expression Program. *Mol. Cell.* **77**, 251-265.e9
54. Cai, F., Heinhorst, S., Shively, J. M., and Cannon, G. C. (2008) Transcript analysis of the Halothiobacillus neapolitanus cso operon. *Arch. Microbiol.* **189**, 141–150
55. Sun, Y., Harman, V. M., Johnson, J. R., Brownridge, P. J., Chen, T., Dykes, G. F., Lin, Y., Beynon, R. J., and Liu, L.-N. (2022) Decoding the Absolute Stoichiometric Composition and Structural Plasticity of α-Carboxysomes. *MBio*. **13**, e0362921
56. Dai, W., Chen, M., Myers, C., Ludtke, S. J., Pettitt, B. M., King, J. A., Schmid, M. F., and Chiu, W. (2018) Visualizing Individual RuBisCO and Its Assembly into Carboxysomes in Marine Cyanobacteria by Cryo-Electron Tomography. *J. Mol. Biol.* **430**, 4156–4167
57. Iancu, C. V., Morris, D. M., Dou, Z., Heinhorst, S., Cannon, G. C., and Jensen, G. J. (2010) Organization, structure, and assembly of alpha-carboxysomes determined by electron cryotomography of intact cells. *J. Mol. Biol.* **396**, 105–117
58. Rotskoff, G. M., and Geissler, P. L. (2018) Robust nonequilibrium pathways to microcompartment assembly. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 6341–6346
59. Edgar, R. C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. **26**, 2460–2461
60. Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015) The MEME Suite. *Nucleic Acids Res.* **43**, W39-49
61. Katoh, K., and Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780
62. Price, M. N., Dehal, P. S., and Arkin, A. P. (2010) FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One*. **5**, e9490
63. Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190
64. Guillén-Boixet, J., Kopach, A., Holehouse, A. S., Wittmann, S., Jahnel, M., Schlüßler, R., Kim, K., Trussina, I. R. E. A., Wang, J., Mateju, D., Poser, I., Maharana, S., Ruer-Gruß, M., Richter, D., Zhang, X., Chang, Y.-T., Guck, J., Honigmann, A., Mahamid, J., Hyman, A. A., Pappu, R. V., Alberti, S., and Franzmann, T. M. (2020) RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell*. **181**, 346-361.e17

Chapter 3:
Measuring permeability and redox in the carboxysome

3.1 Chapter Summary

Auxotrophic growth of bacteria harboring carboxysomes relies on the flow of essential metabolites across the carboxysome shell layer. Bicarbonate and ribulose bisphosphate must enter, 3-phosphoglycerate must exit, and crucially, CO_2 must be retained. Though this flux is well understood, little is known about exactly how the carboxysome establishes a selective permeability barrier, and how selective this barrier actually is. Carboxysome permeability properties also have implications for redox reactions inside the carboxysome, since reactive oxygen species or reductants face the same shell barrier as all other cellular compounds. This chapter explores both carboxysome permeability and internal redox environment using an encapsulated redox sensor, roGFP. Two parallel experimental approaches were taken: a bulk assay and single-particle analysis. Both approaches found that purified carboxysomes were permeable to the reducing agent TCEP, displaying a reduced internal environment over time. Measuring chemical kinetics remains an important, yet tricky, assay for carboxysomes, and this work details methods and advances towards quantifying permeability.

3.2 Introduction

The work in this chapter emerged from a series of hypotheses about carboxysomes held by the lab since before I joined. Mainly, the hypothesis was that carboxysomes are oxidizing on the interior, creating a strategic chemical environment that exists as a protected haven from the reducing environment of the bacterial cytosol. By this theory, the shell protects carboxysomal proteins from the barrage of cellular reducing agents (glutathione, thioredoxin), promoting the formation of oxidized disulfide bonds in carboxysomal proteins, thereby allowing for strategic localized regulation of carboxysomal protein activity.

This theory has a good amount of experimental backing. In a fantastic paper by Price & Badger in 1989, they showed that “expression of human carbonic anhydrase in the cyanobacterium *Synechococcus PCC7942* creates a high CO_2-requiring phenotype” (1). The quoted part is the title of their paper. The paper demonstrates how carbonic anhydrase (CA) activity outside of the carboxysome short-circuits the CCM, leading to a loss of cellular C_i accumulation and a high-CO_2 requiring phenotype. They conclude that the CA must be inactive in the cytosol and active specifically in the carboxysome for the CCM to function. What they didn’t show is how the cell knows whether or not the CA is inside the carboxysome. A mechanism in which the CA is only active when it is encapsulated would make the most sense.

Turns out that carboxysomes use redox as a CA encapsulation sensor, at least for β-carboxysomes. In *Thermosynechococcus elongatus BP-1*, a β-cyanobacterium, the CcmM C-terminal γ-CA is oxidatively activated (2). β-carboxysomes were also observed *in vivo* via fluorescence microscopy with encapsulated roGFP to oxidize as they matured (3). A similar mechanism has not yet been demonstrated for CsoSCA, the α-carboxysome CA, though it was shown that DTT and βME can somewhat inhibit activity of the *H. neapolitanus* CsoSCA (4).

It's also impossible to ignore the abundance of cysteines in carboxysomal scaffold proteins, as I pointed out in Chapters 1 and 2. Cysteine doublets often hint at a role for oxidatively induced disulfide bonding. In addition to the above mentioned γ-CA, cysteine disulfide bonds are undeniably important in CcmM. The reduced form of CcmM was shown to bind Rubisco with higher affinity but had less mobility in phase separated droplets (5). This supports a carboxysome biogenesis pathway where the reducing cytosol makes it easier for CcmM and Rubisco to associate, the carboxysome forms, and then the CcmM structure relaxes and becomes more mobile in the oxidized carboxysome lumen, which may be beneficial for carboxysome function. It's like protein dating: they find each other under favorable circumstances, a bond forms, and then they relax into a steady relationship (6). The study authors mutated the CcmM cysteines and saw a growth phenotype, at least. Can't say as much for the cysteines in CsoS2 (see Chapter 2).

The other reason carboxysomes likely have an oxidizing interior is that O_2 can (probably) simply slip through the shell pores. It is significantly smaller than other molecules that have to enter, such as HCO_3^- and 3-PG. This is hotly debated since a "selective barrier" mechanism makes logical sense to concentrate CO_2 near Rubisco and eliminate the competitive O_2 reaction. And some studies seem to prove this: oxygen-sensitive hydrogenases were shown to function in carboxysomes, though molecular crowding may have also been playing a role (7). Molecular dynamics simulations have found that the selective barrier exists (8), but others make mathematical arguments against it (9).

An experimental measurement of the internal carboxysome redox environment would help settle the debate on these longstanding questions. That's where I came in during the fine spring of 2018. I was to design an assay to test this hypothesis that would be "so easy"; a paper within a year type of thing. Well, you can already guess where this is going.

The idea was to kill two birds with one stone: determine the carboxysome redox environment while simultaneously measuring carboxysome permeability. I would target a redox-sensitive GFP (roGFP) to the interior of the carboxysome, which would allow me to measure an internal vs. external redox state. I would then challenge purified roGFP carboxysomes with a panel of redox agents of different sizes: βME, DTT, TCEP, Glutathione, and finally, the protein thioredoxin (Figure 1). This would allow me to see

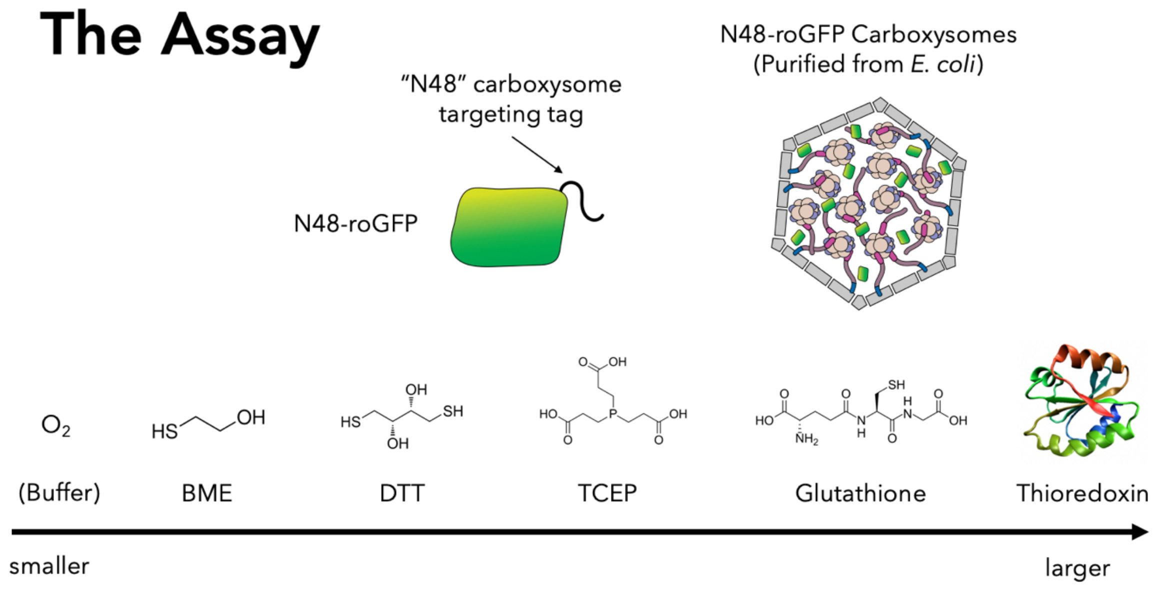
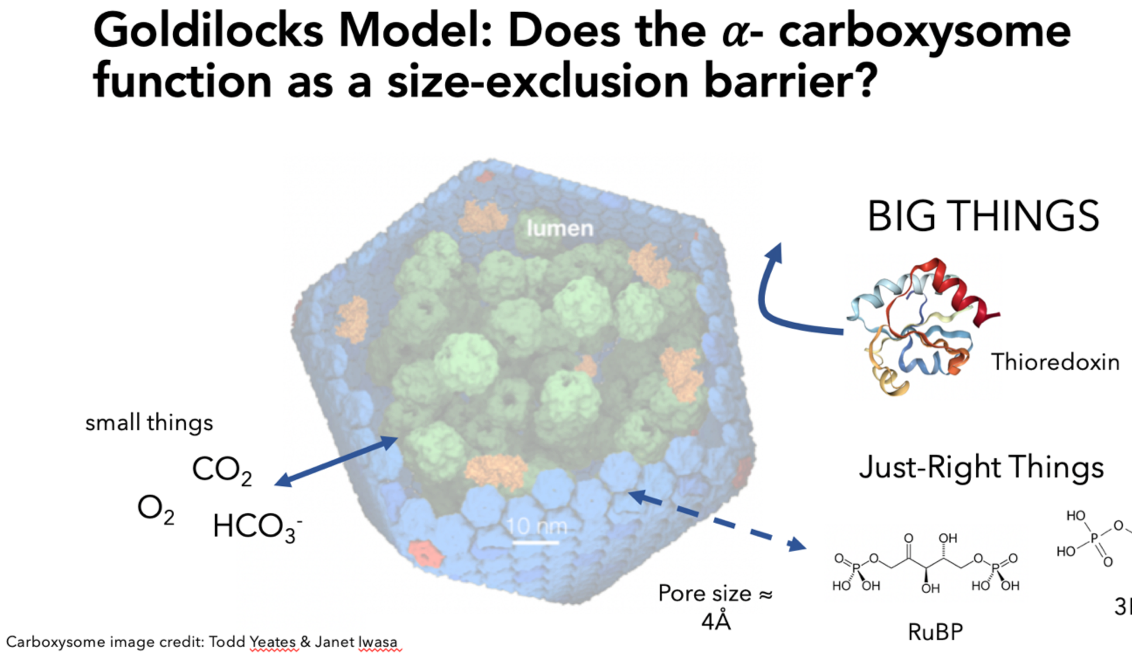


Figure 1. The roGFP Assay. roGFP was tagged with an N48 carboxysome targeting tag pulled from *H. neapolitanus* CsoSCA (18). Note, N48 is a misnomer as it is actually 53 peptides. Cissi never changed the name, ah well. This targeted roGFP to the carboxysome interior. The bottom panel shows the different redox agents sampled in this assay, from smallest to largest.

which redox agents enter through the carboxysome pores and reduce the roGFP, and which face a size restriction barrier, thereby serving as a measurement of carboxysome permeability.

My hypothesis was that some or all chemical reducing agents would get in, perhaps drawing a line at the larger ones such as TCEP and Glutathione, but that the protein thioredoxin (Trx) would encounter a barrier. I called this the Goldilocks Model (Figure 2); my experiment would figure out what size was "just right". I was pretty stoked about the permeability assay – no one (back then, and even to this date) has been able to measure carboxysome permeability. Plenty of papers compare pore sizes from crystal structures of shell proteins (10–12) and mathematically model permeability (8, 9, 13, 14), but none has ever visualized a metabolic substrate (CO_2, HCO_3^-, RuBP, 3PG) transiting the pores or measured it experimentally. How does the carboxysome restrict leakage of CO_2 while freely allowing the passage of significantly larger RuBP and 3PG?

So I ran the assay, and the data perfectly supported our hypothesis! Just kidding; if it were so easy you'd probably be reading a concisely written copy-pasted paper instead of these ramblings buried in a thesis. Firstly, there was a fundamental problem with the assay, which is that you could never be sure whether the carboxysomes you purified were whole or broken. This simple fact made most of the data uninterpretable, especially when the data didn't show that the carboxysomes were definitely *not* broken (i.e. adding thioredoxin, a large protein, and not seeing reduction would have shown this). Secondly, the assay required A LOT of carboxysomes, and they are not easy to purify in large quantities! (As a side note, I spent a summer learning to use the



Goldilocks Model: Does the α- carboxysome function as a size-exclusion barrier?

Carboxysome image credit: Todd Yeates & Janet Iwasa

Figure 2. The Goldilocks Model. The model states that small molecules (like CO_2, O_2, and HCO_3^-) enter freely through the pores, big molecules (like thioredoxin) are excluded, and Just Right molecules (like RuBP and 3PG) also enter, but may exist at the upper limit of pore permeability.

basement bioreactor and grew ~20L of *Halo* (Figure S14) in preparation for purifying native carboxysomes for these experiments, after first doing all the troubleshooting with less precious *E. coli* carboxysomes. Those pellets are still in the freezer, preserved for their memory more than anything else.)

I was pushing on this, and then the pandemic hit. We sheltered in place, and I did a lot of reading on carboxysomes. I got some fun new project ideas from that period, which you can read about in the Appendix (A.3). And I considered the fundamental issues of this roGFP assay, now coupled with the fact that we were working shifts in lab and it was hard to do carboxysome purifications with half the lab time.

This is when I learned the economic lesson of the Sunk Cost Fallacy, which I now understand on a very personal level. Sunk Costs are costs that have already been spent and cannot be recovered. The Sunk Cost Fallacy describes the irrational human behavior of deciding to stick with something simply because you've invested time/money/energy/whatever in it, even though it's the wrong decision to keep pursuing it. With frustrated disappointment I decided to cut my losses and work on another project. The new project turned into my first author paper (see Chapter 2).

Some time after abandoning this roGFP project and moving on to greener pastures, a cool collaboration opportunity popped up surrounding the very questions at the core of this project. The Moerner Lab at Stanford are experts in optical microscopy and single molecule studies. That's an understatement. W.E. Moerner won the Nobel Prize in 2014 for the development of single-molecule microscopy. Spoiler alert, I'm now on multiple papers with a Nobel Prize winner, a scientist's dream! Among other things, the Moerner Lab invented the Anti-Brownian ELectrokinetic (ABEL) trap, which counteracts the Brownian motion of objects in solution by applying electric fields, allowing them to effectively trap single fluorescently labeled proteins in solution. In 2019, they created a new ABEL trap that could trap particles based on their scattering profile rather than fluorescence, called the Interferometric Scattering Anti-Brownian ELectrokinetic (ISABEL) trap. The carboxysome was the perfect subject – large enough to scatter light, small enough to be trapped, and ripe with interesting questions that could be addressed via single molecule studies. Questions exactly like the ones I was trying to answer in my bulk assay. This innovative technology is known as Carboxysome Redox Insights from Single Particle Results (CRISPR). So let it be recorded in history that I worked on CRISPR with a Nobel Prize winning scientist!

We (Luke & I) started collaborating with the Moerner group, in particular William (Memo) Carpenter and Abhijit Lavania, a postdoc and graduate student in the lab. My work was of particular interest to Memo, who was excited about using the ISABEL trap to probe questions of carboxysome redox and permeability. We supplied the carboxysomes and biological intel, and they did the optical trapping and expert analyses of physical properties. With Abhijit's work we made precise measurements of carboxysome size and mass (15), and with Memo's work we were able to measure carboxysome redox properties (16). The kicker is, my bulk data showed the same result as the single particle data: carboxysomes were permeable to TCEP, and showed internal reduction over time. Seeing that made me feel like the few years I spent on this weren't a total waste of time. So now, dear reader, you can view the bulk data accompaniment to Carpenter et al., exclusively presented in this thesis!

3.3 roGFP as a carboxysome redox sensor and bulk assay optimization

Redox-sensitive GFP (roGFP) was invented in 2004 as a simple and non-invasive way to measure the redox potential inside cells (17). Hanson *et al.* engineered GFP to have two closely interacting surface-exposed cysteines that can rapidly and reversibly form a disulfide bond under oxidizing conditions. roGFP has two fluorescence excitation maxima at 400 and 490 nm. Under reducing conditions, the 490 peak raises while the 400 peak drops, and the reverse happens in oxidizing conditions. The resulting ratiometric 400/490 nm measurement acts as a redox reporter.

Purified roGFP worked in our hands as expected with all of the redox agents I intended to use in the assay (Figure 3A). 400/490 ratios are shown in Table 1. To establish control concentrations and time settings, I ran a concentration sweep of each reducing agent with roGFP. Figure 3B shows the controls for DTT with 200 nM roGFP over 10 hours (Figure 3B). This established that 1000x DTT (200 uM) to roGFP is an adequate concentration to reduce roGFP in under an hour. Reaction conditions for Thioredoxin (Trx) are shown in Figure 3C and D. Trx + NADPH + roGFP showed consistent, yet slow, reduction over 3 hours (orange trace) while the addition of Thioredoxin reductase (TrxR) led to re-oxidation over ~75 minutes (pink trace) (Figure

3D). NADPH oxidation due to NADPH reduction of Trx could be traced simultaneously, as measured by the decrease in 340nm signal (Figure 3C). The addition of TrxR showed an almost instantaneous drop in A340 at time 0, indicating a large pool of oxidized NADPH, yet why this did not carry over to faster roGFP reduction remains unclear. These control experiments established Trx + NADPH as the appropriate condition to reduce roGFP, without the addition of TrxR.

After establishing basic roGFP controls, I purified carboxysomes with internally targeted roGFP. The 'N48' carboxysome localization tag from CsoSCA (18) was fused to the N-terminus of roGFP to facilitate loading into the carboxysome. Purified

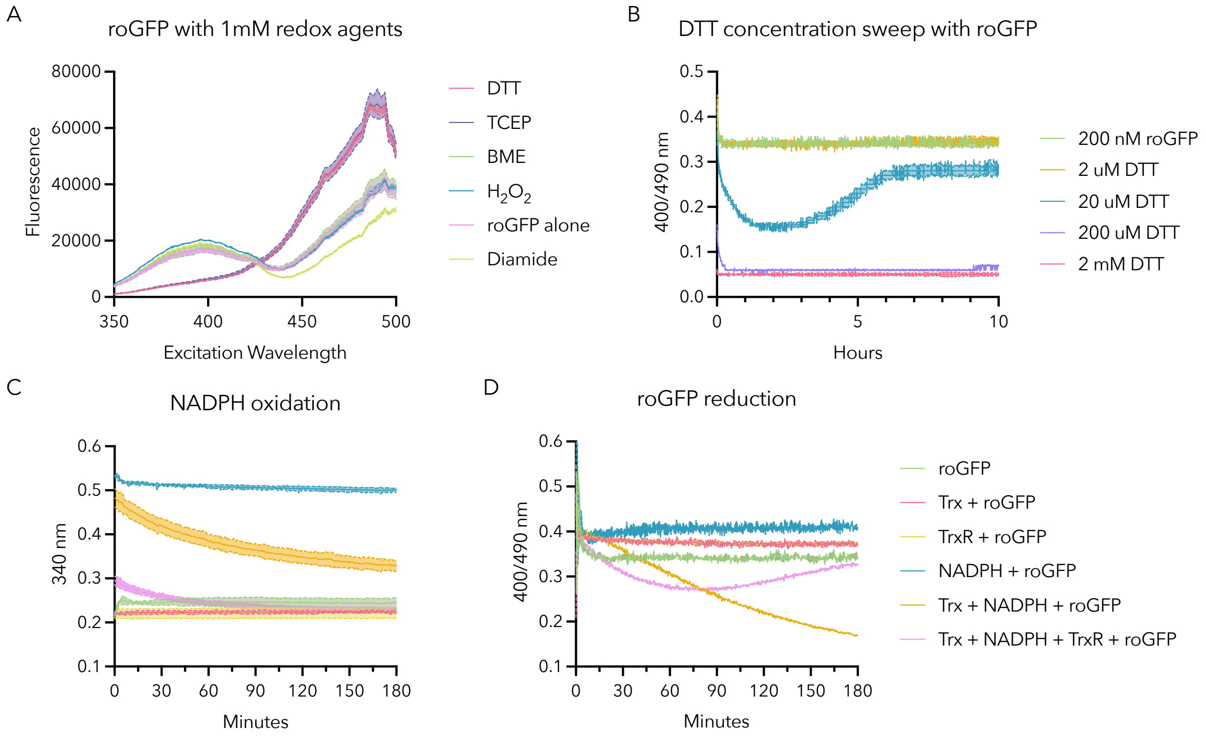


Figure 3. roGFP works *in vitro* as expected. (A) Controls showing the reversible shift in fluorescence output at 400 and 490 nm as demonstrated by Hanson *et al.* (17). 2 uM roGFP was used, and 1 mM reducing agents. (B) DTT concentration sweep with 200 nM roGFP over 10 hours. (C) NADPH oxidation as measured by the decrease in signal at 340 nm. (D) roGFP oxidation/reduction with Trx + NADPH and other controls over 3 hours.

Table 1. 400/490 ratios for the panel of redox agents in Figure 3A.

Redox agent	400/490
TCEP	0.09
DTT	0.09
BME	0.44
roGFP	0.44
H_2O_2	0.52
Diamide	0.67

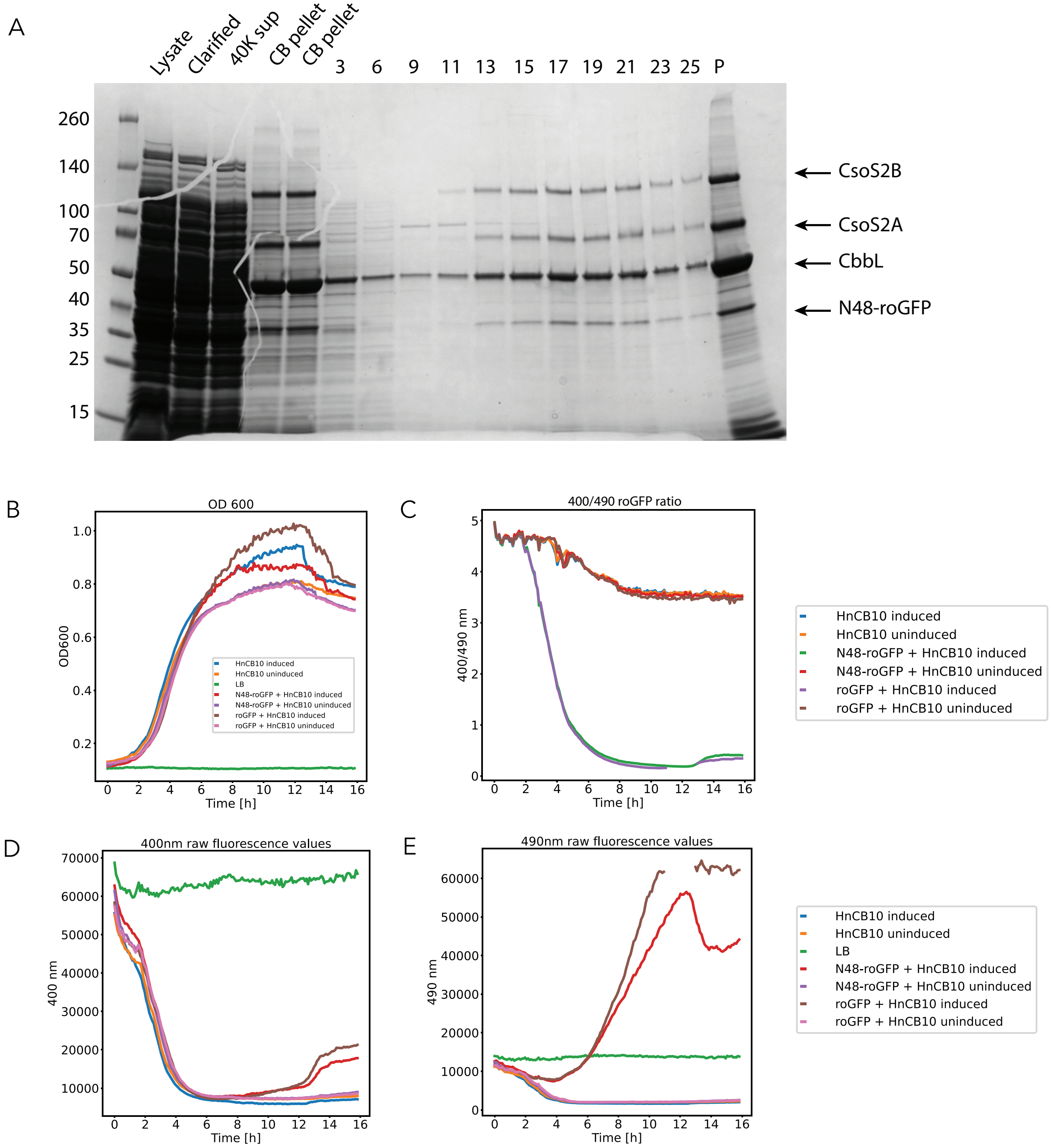


Figure 4. N48-roGFP carboxysome purification and *in vivo* studies of N48-roGFP expression. (A) *E. coli* expressing carboxysomes off of the pHnCB10 plasmid and N48-roGFP off of a separate plasmid could be readily purified using standard methods. The N48-roGFP band (~34 kDa) appears in the expected fractions of the sucrose gradient alongside other carboxysomal proteins. Lower bands (<15 kDa) cannot be seen because they ran off of the gel. (B) Growth of induced and uninduced cells harboring carboxysomes and roGFP or N48-roGFP. (C) 400/490 ratio over the growth period. (D) 400 nm raw fluorescence values. (E) 490 nm raw fluorescence values.

carboxysomes showed a robust N48-roGFP band (~34 kDa) on an SDS-PAGE gel, confirming successful loading (Figure 4A). I additionally tested tagging the Rubisco large subunit with roGFP, but I was unable to purify carboxysomes with this construct (Figure S15A, B). As an aside, *H. neapolitanus* expressing CbbL-roGFP displayed some interesting qualitative phenotypic variances compared to the strain expressing N48-roGFP (Figure S15C, D).

I did some initial *in vivo* experiments with *E. coli* expressing carboxysomes (via the pHnCB10 plasmid) and roGFP, to see if I could observe changes in carboxysomally encapsulated N48-roGFP redox state without lysing cells and going through purification. Cells grew similarly under all conditions (Figure 4B). Induced cells showed a significant drop in 400/490, indicating roGFP reduction in the reducing cytosol (Figure 4C). However, both N48-roGFP and roGFP were reduced, likely because a significant portion of N48-roGFP was not encapsulated in carboxysomes (this was separately confirmed by the large fraction of soluble N48-roGFP present at the top of the sucrose gradient during carboxysome purification). Given these results I did not proceed further with *in vivo* experiments to evaluate the encapsulated N48-roGFP redox state.

Under current assay conditions, a consistent and unexplainable initial downward slope would appear for the oxidized roGFP 400/490 ratio, whereas no decrease appeared for the N48-roGFP carboxysome samples (Figure 5A). A decrease in 400/490 typically indicates reduction, yet no reducing agent had been added, or any other assay conditions changed. The drop was most obvious for the oxidized roGFP sample, but may also have been occurring in the reduced sample and affecting the measured rate. Further investigating the signal by looking at the raw 400 nm and 490 nm traces, roGFP showed an uncharacteristic loss of fluorescence over time. The 400 nm oxidized roGFP trace showed a decrease, when it would be expected to remain flat (Figure 5B). The 490 nm trace showed a leveling off of the reduced roGFP sample when it should increase in the presence of reducing agent, and a decreased fluorescence signal for the oxidized sample when it should remain flat (Figure 5C). In contrast, the N48-roGFP carboxysome traces appeared normal, with the 490 nm DTT trace tracking upward as expected (Figure 5C).

After talking with some labmates (a day in the library saves a month in the lab; an hour's discussion saves a day in the library), I investigated adsorption as a possible cause. Adsorption is the adhesion of molecules to a solid surface, in this case adhesion

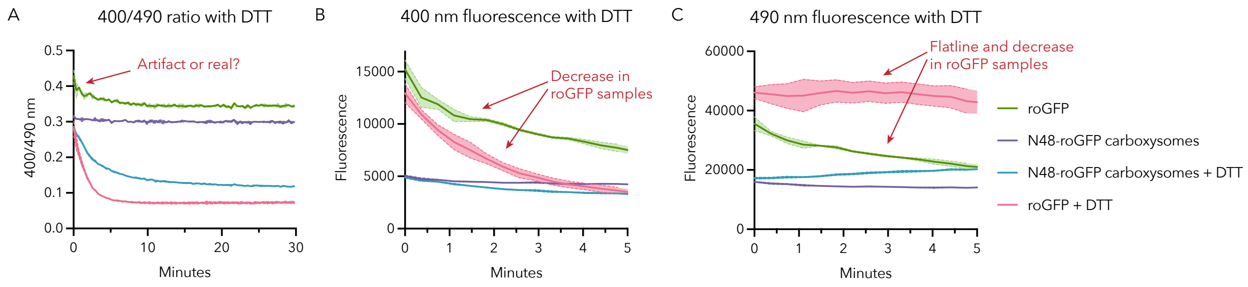


Figure 5. Unexplained initial drop in the 400/490 ratio for roGFP samples. (A) A drop in the 400/490 ratio for roGFP can be seen within the first 5 minutes of the assay before reaching equilibrium. (B) Raw 400 nm values show sharp decreases in fluorescence for roGFP. (C) Raw 490 nm values show sharp decreases or plateaus in fluorescence for roGFP.

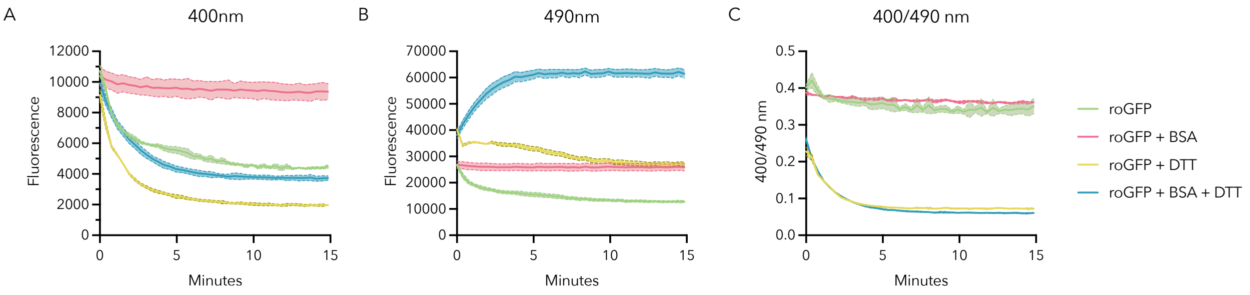


Figure 6. roGFP adsorption controls. All plots include the samples roGFP (green), roGFP + BSA (pink), roGFP + DTT (yellow), roGFP + DTT + BSA (blue). (A) Fluorescence at 400 nm. (B) Fluorescence at 490 nm. (C) 400/490 nm ratio.

of roGFP to the 96-well plate, which could alter the optical properties of roGFP. I added 1 mg/ml of BSA to the roGFP samples to passivate the sticky surface and repeated the assay. This remedied the signal and proved adsorption to be the cause of the issue. The raw 400 nm fluorescence trace for oxidized roGFP + BSA stayed flat instead of decreasing, and for reduced roGFP both traces decreased as expected (Figure 6A). The largest difference could be seen in the 490 nm fluorescence trace, with reduced roGFP + BSA increasing dramatically over 5 minutes and remaining high as expected, while the sample without BSA decreased over the same amount of time (Figure 6B). The 400/490 plot shows the cumulative effect, most visible in the oxidized sample traces, with roGFP displaying a decrease over time and roGFP + BSA remaining flat (Figure 6C). This also explains why the carboxysome samples did not show this artifact in Figure 5 – N48-roGFP was encapsulated, and therefore did not adsorb onto the plate and display altered optical properties. Following this optimization step, I included BSA in all of my roGFP samples in future assays.

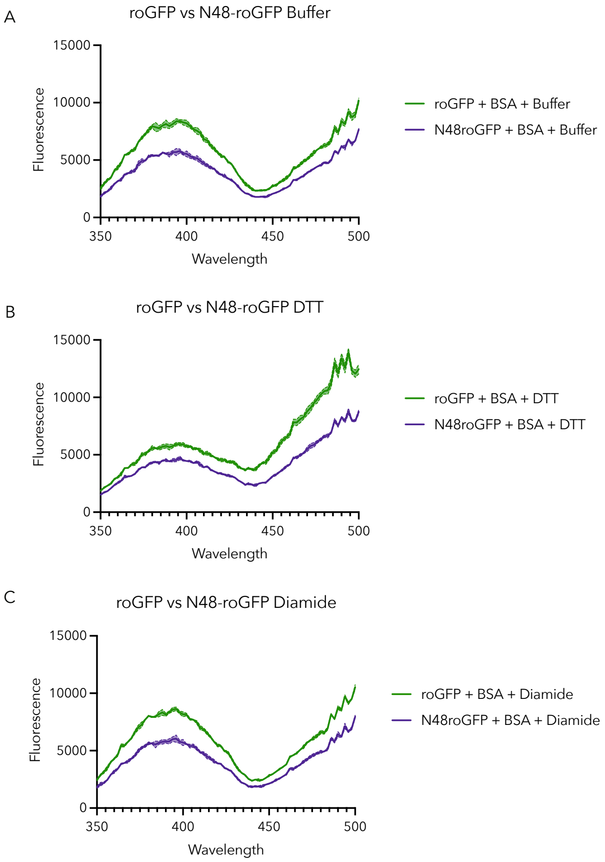


Figure 7. Comparing fluorescence properties of roGFP and N48-roGFP in different redox environments. roGFP (green), N48-roGFP (purple). A) Buffer (oxidizing). B) DTT (reducing). C) Diamide (oxidizing).

While the experiments in Figures 3-6 helped to establish assay conditions, I decided to switch to N48-roGFP rather than roGFP as a more appropriate control. All the carboxysomal roGFPs have the N48 tag, which is a carboxysome localization tag taken from the N terminal 53 amino acids of CsoSCA (it really should be called N53, but Cissi never changed the name). Comparing roGFP to N48-roGFP, N48-roGFP displayed lower fluorescence in buffer, DTT, and diamide conditions (Figure 7A, B, and C). However, its redox-responsive properties of lower 400/490 in reducing conditions and higher 400/490 in oxidizing conditions remained unchanged, confirming its use as an appropriate control (Table 2).

Table 2. 400/490 values for roGFP and N48roGFP in reducing and oxidizing conditions.

	400/490 values		
	Buffer	DTT	Diamide
roGFP	0.98	0.43	0.93
N48roGFP	0.89	0.54	0.89

3.4 Bulk assay results

3.4.1 All tested redox agents access and reduce N48-roGFP inside carboxysomes

Five redox agents of varying sizes were tested to compare rates of reduction between soluble N48-roGFP and N48-roGFP encapsulated in carboxysomes: β-mercaptoethanol (BME), dithiothreitol (DTT), tris(2-carboxyethyl)phosphine (TCEP), thioredoxin (Trx), and glutathione. The chemical compound or structure is shown to the left of each plot involving that chemical (Figure 8). To reiterate, the hypothesis was that smaller chemicals would be able to access and reduce N48-roGFP through the carboxysome pores, and that the bigger the molecule got, the slower the rate of reduction would be until ceasing altogether for impermeable molecules.

All redox agents showed robust and measurable reduction of the control, soluble N48-roGFP (Figure 8 A, C, E, G, and I). The rate of reduction for each redox agent increased in a concentration-dependent manner. Oxidized N48-roGFP 400/490 values (those at 0 uM reducing agent) hovered around 0.4. “Fully reduced” values ranged from 0.1 – 0.2, though some redox agents may have required higher concentrations than those tested to achieve complete reduction.

Similarly, and perhaps surprisingly, all redox agents were able to reduce carboxysomal N48-roGFP (Figure 8 B, D, F, and H). This was expected for BME, the smallest molecule, and DTT and TCEP, which approximate RuBP in size and structure. However, to our dismay, the protein thioredoxin was also able to enter carboxysomes and reduce N48-roGFP (Figure 8H). For whatever reason I did not test glutathione, probably because I was so frustrated with the assay results not turning out as hoped for. Like the controls, the rate of reduction increased in a concentration-dependent manner, though fewer concentrations were tested due to a limited supply of carboxysomes. Interestingly, the range of oxidized-to-reduced 400/490 values was much narrower for carboxysome samples. Oxidized values were about 0.3, while “fully reduced” values

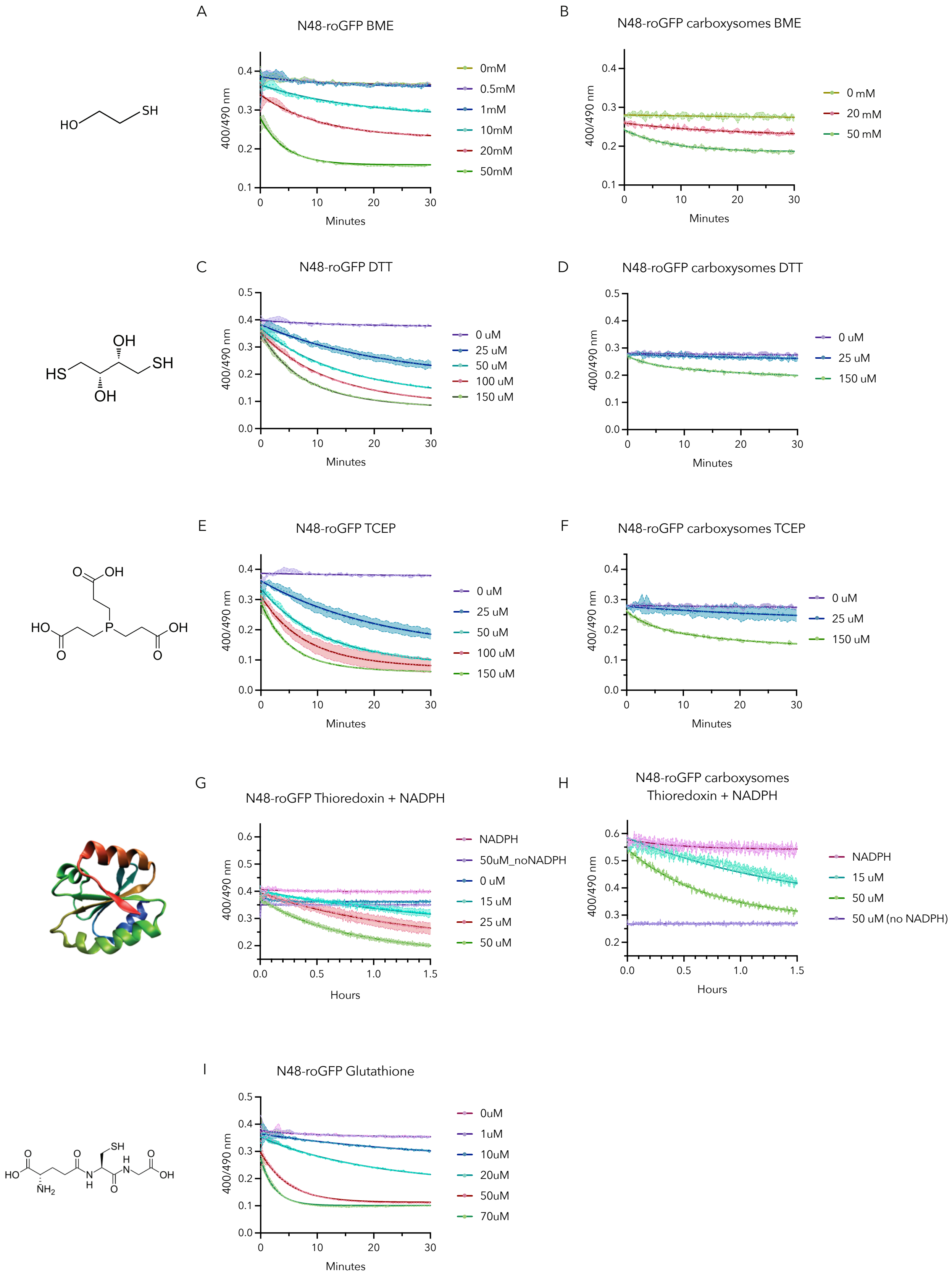

A
N48-roGFP BME
400/490 nm
Minutes
0mM
0.5mM
1mM
10mM
20mM
50mM
HO
SH
B
N48-roGFP carboxysomes BME
0 mM
20 mM
50 mM
C
N48-roGFP DTT
0 uM
25 uM
50 uM
100 uM
150 uM
HS
SH
OH
OH
D
N48-roGFP carboxysomes DTT
0 uM
25 uM
150 uM
E
N48-roGFP TCEP
0 uM
25 uM
50 uM
100 uM
150 uM
HO
OH
O
P
F
N48-roGFP carboxysomes TCEP
0 uM
25 uM
150 uM
G
N48-roGFP Thioredoxin + NADPH
Hours
NADPH
50uM_noNADPH
0 uM
15 uM
25 uM
50 uM
H
N48-roGFP carboxysomes Thioredoxin + NADPH
NADPH
15 uM
50 uM
50 uM (no NADPH)
I
N48-roGFP Glutathione
0uM
1uM
10uM
20uM
50uM
70uM
HO
NH$_2$
SH
N
H
OH

Figure 8 (previous page). Redox agents reduce both soluble N48-roGFP and carboxysomal N48-roGFP. Results for soluble N48-roGFP are shown in the left column, and results for carboxysomal N48-roGFP are shown in the right column. Each row corresponds to a different redox agent. Structures of all applied redox agents are shown to the left of each plot. A) N48-roGFP with BME. B) N48-roGFP carboxysomes with BME. C) N48-roGFP with DTT. D) N48-roGFP carboxysomes with DTT. E) N48-roGFP with TCEP. F) N48-roGFP carboxysomes with TCEP. G) N48-roGFP with thioredoxin + NADPH. H) N48-roGFP carboxysomes with thioredoxin + NADPH. I) N48-roGFP with glutathione. J) Yeah, use your imagination.

were around 0.15 – 0.2. This may be due to the overall lower signal from carboxysomal N48-roGFP and potentially altered spectral properties as a result of encapsulation. The notable exceptions are 400/490 values from the thioredoxin sample (Figure 8H), which range from 0.25 – 0.55. The jump in signal is likely due to NADPH contribution to the 400 nm fluorescence reading, as the NADPH alone sample trace is at the maximum 0.55 value. As NADPH is oxidized and loses spectral signal, roGFP is simultaneously reduced, contributing to the artifactually larger range in 400/490 values. However, it is unclear why the difference between controls (NADPH + N48-roGFP vs. Trx + N48-roGFP) is so much larger for carboxysomes than for soluble N48-roGFP.

While this discussion remains qualitative for now, a quantitative comparison is coming up in Figure 11.

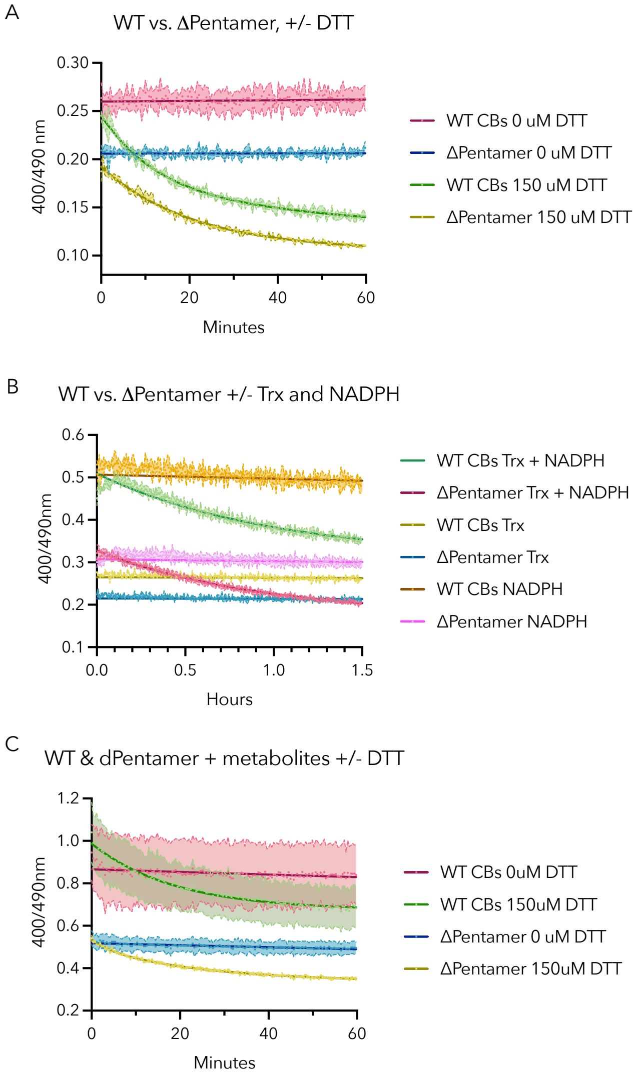


Figure 9. ΔPentamer carboxysomes show similar N48-rGFP reduction rate to WT carboxysomes. A) DTT, B) Trx + NADPH, C) DTT with added metabolites.

3.4.2 Altering carboxysome permeability has little effect on reduction rate

Following the results of Figure 8, we wondered if changing the permeability of carboxysomes would yield any differences in the N48-roGFP reduction rate. To permeabilize carboxysomes without affecting their structure, I deleted the pentamer proteins, effectively introducing 12 large holes into the carboxysome (19). Our hypothesis was that reduction rates would increase due to easier access to the carboxysome interior. However, we saw that reduction rates remained roughly equal between Δpentamer and WT carboxysomes (Figure 9) (quantitative comparison in Figure 11). The major difference between Δpentamer and WT carboxysome samples was the downshift in 400/490 values, which may be due to altered N48-roGFP spectral properties in the two different types of carboxysomes. I additionally tested a new condition where I added metabolites that may be involved in the function of carboxysomal proteins: ATP, α-ketoglutarate (αKG), ribulose bisphosphate (RuBP), GTP, ADP, and NADPH. This similarly had little effect on the rate, but contributed to upshifted 400/490 values due to spectral contributions from the added compounds.

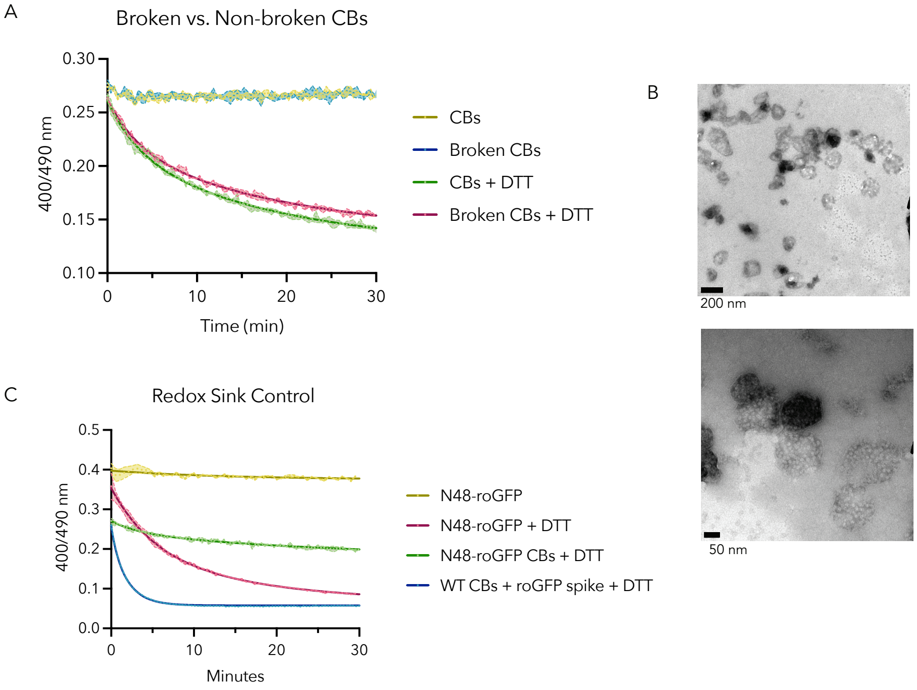


Figure 10. Broken carboxysome and redox sink controls. A) Broken vs. unbroken carboxysomes (CBs), +/- DTT. B) TEM of unbroken N48-roGFP carboxysomes used in the assay, two images of the same sample. C) Redox sink control testing spiked in roGFP with WT carboxysomes (blue trace).

In a last effort to determine if increasing carboxysome permeability would lead to faster reduction rates, establishing that there is indeed a permeability barrier imposed by the carboxysome, I tried breaking carboxysomes using a freeze-thaw method. To break carboxysomes, I put them at -20°C for 30 minutes and then let them thaw at room temperature. When tested in the roGFP assay, broken carboxysomes had a slower rate of reduction than unbroken carboxysomes (Figure 10A). There are three possible explanations for this result: 1) my unbroken carboxysomes were actually broken, 2) my broken carboxysomes were actually not broken, 3) aggregates from broken carboxysomes somehow made the N48-roGFP harder to access and reduce (unlikely). Whatever the status of my poor carboxysomes, I felt pretty broken after seeing these results. TEM of purified WT carboxysomes showed them to be somewhat broken and cracked, with Rubisco spewing out like the contents of an upset stomach (Figure 10B). But then again, this could just be an artifact of the staining process and not a reflection of how the carboxysomes were when used in the assay. This gets back to one of the fundamental issues with the assay: you could never know if purified carboxysomes were whole or broken, complicating interpretation of assay results.

3.4.3 Carboxysomes do not act as a "redox sink"

There was one last control to test: determining whether or not carboxysomes were acting as "redox sinks" due to the abundant number of reducible residues on each carboxysome. There are far more reducible residues per carboxysome than there are roGFPs; other carboxysomal proteins could potentially lower the overall concentration of reducing agent, therefore slowing the rate of reduction of N48-roGFP. It was unlikely this was happening because I added 1000x reducing agent to N48-roGFP in my assay, but it was worth testing. I took WT carboxysomes without encapsulated N48-roGFP and spiked in soluble roGFP. Spiked samples reduced at roughly the same rate, perhaps faster, than N48-roGFP without carboxysomes, proving that carboxysomes are not acting as a "redox sink" decreasing the measured rate of reduction (Figure 10C).

3.4.4 Measured rates of reduction between soluble N48-roGFP and carboxysomal N48-roGFP are similar across reducing agents

Values quantifying the initial rate of the reduction curves (K_{fast}) in Figure 8 show that there is very little difference between rates of reduction for soluble N48-roGFP and carboxysomal N48-roGFP. Rates were quantified for the highest concentration condition for each reducing agent: 50 mM BME, 150 uM DTT, 150 uM TCEP, and 50 uM thioredoxin. Statistical significance could not be calculated because K_{fast} is not a mean; it was calculated as a nonlinear regression parameter from the plots in Figure 8. Error is reported as a 95% confidence interval of K_{fast} derived from the nonlinear regression.

None of the results showed the expected theoretical outcome. As the smallest molecule, BME was expected to have roughly equal rates of reduction, yet results showed a higher rate for soluble N48-roGFP (Figure 11A). In contrast, DTT showed higher rates for carboxysomal N48-roGFP. TCEP and thioredoxin had roughly equal rates between samples. Again, contrary to theory, the more porous Δpentamer carboxysomes had similar rates of reduction to WT carboxysomes (Figure 11B). A comparison of rate ratios further confirmed that the assay could not detect meaningful permeability differences between samples (Figure 11C, D). Sometimes the opposite result occurred than what was expected, such as DTT reducing carboxysomal N48-roGFP at a faster rate than soluble protein, or Δpentamer not having any effect on kinetics. Disappointingly, the largest reducing agent, thioredoxin, encountered no permeability barrier to carboxysomal N48-roGFPs. This suggested that purified carboxysomes were likely damaged and porous to molecules of a large range of sizes. The only way to look at carboxysomes and assess this is TEM. Unfortunately, carboxysomes can get damaged during the TEM preparation process, so it is impossible to know if a TEM result showing broken carboxysomes reflects their true state during the assay.

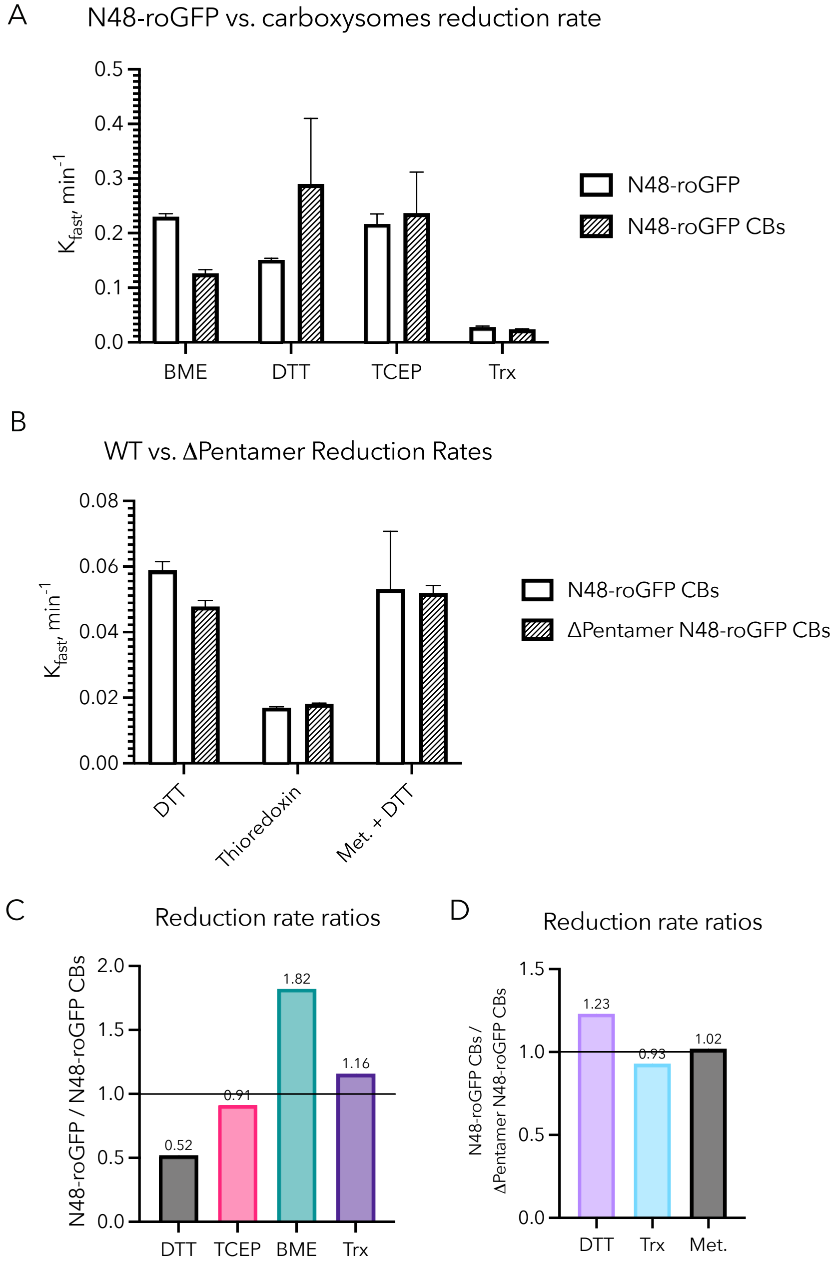


Figure 11. Quantification of soluble N48-roGFP and carboxysomal N48-roGFP reduction. Measured rates are derived from the plots in Figure 8. A) Reduction rates of soluble and carboxysomal N48-roGFP with BME, DTT, TCEP, and thioredoxin (trx). B) Reduction rates of N48-roGFP in WT and Δpentamer carboxysomes with DTT, trx, and DTT + metabolites. C) Ratio of soluble over carboxysomal N48-roGFP reduction rates. D) Ratio of WT over Δpentamer N48-roGFP reduction rates.

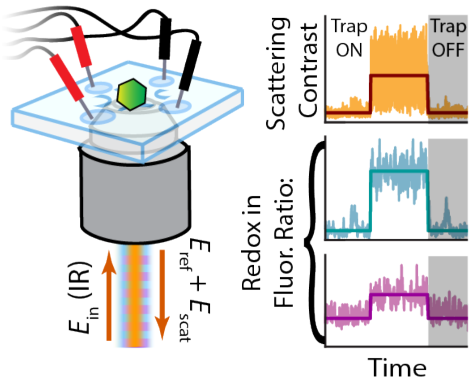
3.5 From a bulk assay to single particle studies

While a bulk assay to test carboxysomal redox and permeability was a good idea in theory, technical limitations to obtaining verifiably whole carboxysomes precluded meaningful data collection.

Fortuitously, the Moerner Lab had a perfect setup to measure the permeability and redox properties of single carboxysomes using their Interferometric Scattering Anti-Brownian ELectrokinetic (ISABEL) trap (described further below). Measuring single carboxysomes offered many advantages over the bulk assay: 1) Determination of basic properties of carboxysomes, such as mass and diameter (mostly reported in Lavania et al. (15)), 2) Visibility into the spread of carboxysome properties, and 3) The potential to allow us to distinguish whole carboxysomes from broken carboxysomes or aggregates based on permeability scores.

The following is the abstract of the published paper “Ratiometric Sensing of Redox Environments Inside Individual Carboxysomes Trapped in Solution” by Carpenter, William B., Abhijit A. Lavania, Julia S. Borden, Luke M. Oltrogge, Davis Perez, Peter D. Dahlberg, David F. Savage, and W. E. Moerner, published in 2022 in the Journal of Physical Chemistry Letters 13 (20): 4455–62 and reprinted with permission. (16)

Diffusion of biological nanoparticles in solution impedes our ability to continuously monitor individuals and measure their physical and chemical properties. To overcome this, we previously developed the Interferometric Scattering Anti-Brownian ELectrokinetic (ISABEL) trap, which uses scattering to localize a particle and applies electrokinetic forces which counteract Brownian motion, thus enabling extended observation. Here, we present an improved ISABEL trap that incorporates a near-infrared scatter illumination beam and rapidly interleaves 405 and 488 nm fluorescence excitation reporter beams. With the ISABEL trap, we monitor the internal redox environment of individual carboxysomes labeled with the ratiometric redox reporter roGFP2. Carboxysomes widely vary in scattering contrast (reporting on size) and redox-dependent ratiometric fluorescence. Further, we used redox sensing to explore the chemical kinetics within intact carboxysomes, where bulk measurements may contain unwanted contributions from aggregates or interfering fluorescent proteins. Overall, we demonstrate the ISABEL trap’s ability to sensitively monitor nanoscale biological objects, enabling new experiments on these systems.



3.6 Ratiometric Sensing of Redox Environments Inside Individual Carboxysomes Trapped in Solution

The following text and figures are adapted from Carpenter, William B., Abhijit A. Lavania, Julia S. Borden, Luke M. Oltrogge, Davis Perez, Peter D. Dahlberg, David F. Savage, and W. E. Moerner. 2022. "Ratiometric Sensing of Redox Environments Inside Individual Carboxysomes Trapped in Solution." Journal of Physical Chemistry Letters 13 (20): 4455–62.

3.6.1 Introduction

For nanoscale biological objects in solution, Brownian fluctuations dominate their translational dynamics. Due to their stochastic trajectories and fast diffusion, individual objects are commonly immobilized for extended study (20, 21), which may undesirably perturb them from their native states (20–24). One approach is to use stage motion to rapidly track and follow single particles in solution, demonstrated on particles with diffusion coefficients as large as approximately 10 $\mu m^2/s$ (25–27). To make extended measurements without tethering, Anti-Brownian ELectrokinetic (ABEL) traps have also been developed (28) to apply electrokinetic positional feedback on single fluorescent objects and thus greatly reduce Brownian motion (29–31). These traps have been used to directly measure the dynamics of single enzymes (32), photosynthetic complexes (33–36), and even single organic fluorophores (37). Typically, these molecules can be held for seconds, when photobleaching or blinking interrupts continuous positional monitoring.

To overcome the need for fluorescence to estimate position, our lab recently developed the Interferometric Scattering ABEL (ISABEL) trap, which tracks a nanoparticle's position by its scattering interfered with a local oscillator arising from the back reflection off a water-quartz interface in the sample cell (38). The interference between the scattered and reflected light enhances sensitivity by producing a signal that scales linearly with particle polarizability, which may be interpreted as mass for objects with fixed composition (39–41). Our initial study demonstrated trapping of gold nanoparticles as small as 20 nm, and that polymer nanoparticles down to 50 nm diameter could be held for more than 30 seconds. Fluorescently labeled particles could also be trapped far beyond the time of photobleaching. Interferometric positional monitoring, independent of fluorescence, opens the door to studying weakly fluorescent biological objects or introducing complex fluorescence excitation protocols.

One such nanoscale biological object is the carboxysome, a proteinaceous microcompartment ~100 nm in diameter, which is responsible for the fixation of CO_2 into organic carbon in many autotrophic bacteria (42–45). The chemoautotroph *Halothiobacillus neapolitanus* contains α-carboxysomes, whose operon encodes 10 proteins that collectively assemble into carboxysomes, including: rubisco large (CbbL) and small subunits (CbbS), a disordered scaffolding protein (CsoS2), carbonic anhydrase (CsoSCA), two pentameric shell protein paralogs (CsoS4AB), three hexamer shell protein paralogs (CsoS1ABC), and a pseudo-hexameric shell protein (CsoS1D) (43, 46). The self-assembled accumulation of rubisco and carbonic anhydrase inside the roughly icosahedral (47) shell (Fig. 12a) has evolved to create a high local concentration of rubisco and CO_2 to overcome rubisco's slow turnover rate and outcompete deleterious oxygenation reactions (48). Functional carboxysomes can also be recombinantly grown in *E. coli* (42), which aids in inserting fluorescent reporters but also increases the diversity of shapes, sizes, and integrity of the shell (Fig. 12b). This paper reports exclusively on *E. coli*-derived carboxysomes.

Structural and simulation studies posit that the proteinaceous shell preferentially allows the bidirectional diffusion of metabolically important species such as HCO_3^-, and ribulose-1,5-bisphosphate (8, 49), and therefore is expected to support a distinct chemical environment from the surrounding cytosol (3, 50). The protein shell is also thought to establish a distinctly oxidizing redox environment within the carboxysome relative to the known reducing environment of the cytosol (2–5, 48). These hypotheses remain unconfirmed, however, due to the lack of direct measurements on selective shell permeability and redox dynamics. Because of the variation of carboxysome size, shape, and integrity, and to mitigate contamination from purification byproducts, it would be highly beneficial to study carboxysomes at the single-particle level. Our goal is to not only trap, but also to sense the redox chemical environment inside individual carboxysomes using a local fluorescent protein reporter, roGFP2, which encodes redox information in its fluorescence excitation spectrum (Fig. 12c). We have genetically targeted approximately 3-15 copies of roGFP2 inside individual carboxysomes (Fig. S1). The ratio of fluorescence brightness from 405 nm and 488 nm excitation is related to the concentration of reducing species in solution (Fig. 12d), and gives a readout that does not rely on GFP copy number.

To enable fluorescence excitation spectroscopy of roGFP2 inside carboxysomes, we have redesigned the ISABEL trap (38). In the new configuration (*vide infra*), the scatter illumination beam has been red-shifted to 800 nm in the near-IR to open up the visible region for fluorescence reporters without photobleaching them. As well, we have introduced two rapidly interleaving excitation beams at 405 and 488 nm to measure the fluorescence emission from roGFP2 (17).

In this paper, we directly measure the redox-dependent ratiometric fluorescence of single trapped carboxysomes, where air-oxidized carboxysomes show much more heterogeneous ratiometric fluorescence than reduced carboxysomes. Despite this heterogeneity, we also observe reduction kinetics in carboxysomes after mixing with reductant, in a first step towards measuring shell permeability. Together, these measurements demonstrate the ability of the ISABEL trap to go beyond synthetic nanoparticles to make extended measurements on single biological objects, and that single-particle measurements on individual carboxysomes provide a new avenue for measuring their physical and chemical properties.

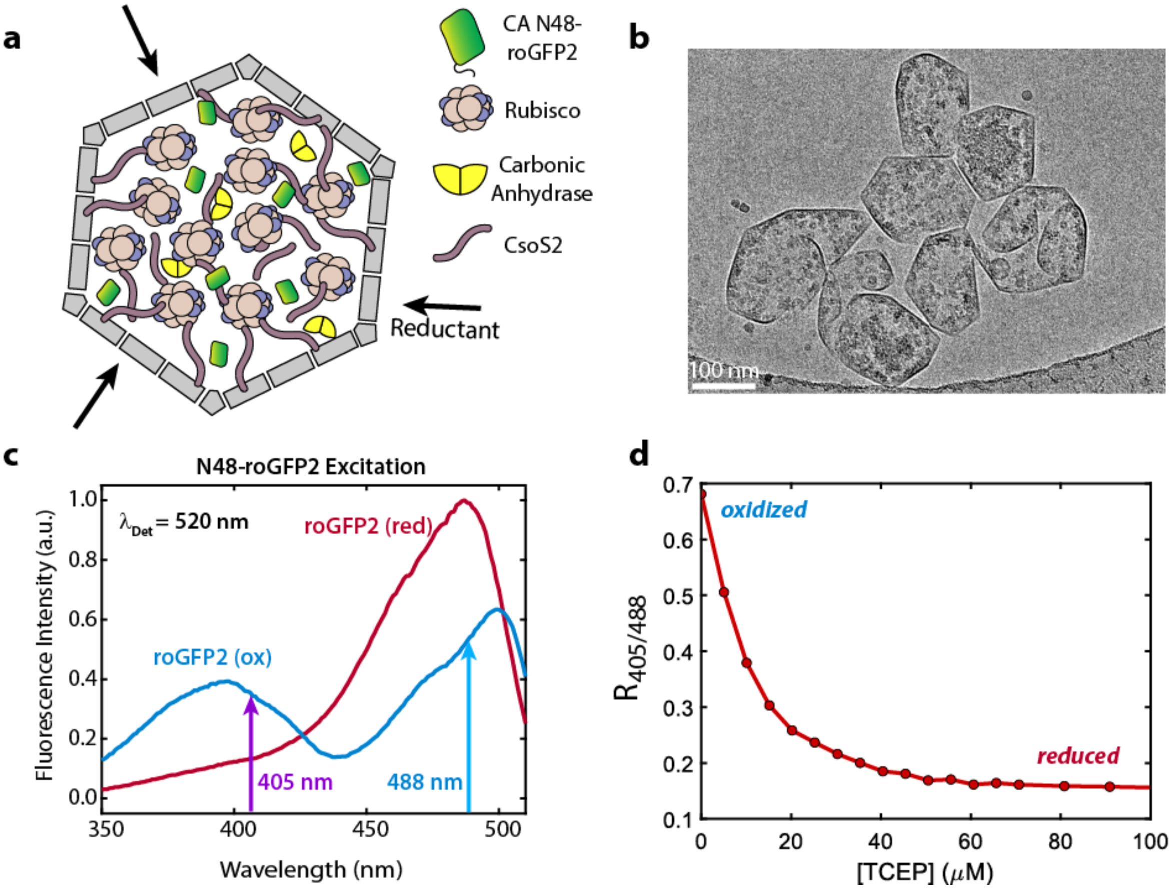


Figure 12. Visualizations of carboxysomes and characteristics of the redox-sensitive GFP mutant roGFP2. (a) The carboxysome consists of a porous proteinaceous shell and the internal cargo rubisco, carbonic anhydrase, and the scaffolding protein CsoS2. roGFP2 is targeted inside the carboxysome using the N-terminal sequence from carbonic anhydrase. (b) A cryo-TEM image of a cluster of α-carboxysomes recombinantly expressed in *E. coli*, demonstrating the variety of shapes, sizes, and integrity. (c) Changes in the fluorescence excitation spectrum of roGFP2 enable ratiometric readout of the redox environment. The fully oxidized spectrum (blue) is bimodal and gives a high fluorescence ratio $R_{405/488}$, while the fully reduced spectrum (red) consists of one peak and produces a low fluorescence ratio. The vertical arrows indicate the excitation wavelengths used in this study for ratiometric measurements. (d) Ratiometric fluorescence from roGFP2 decreases when the reductant TCEP is added.

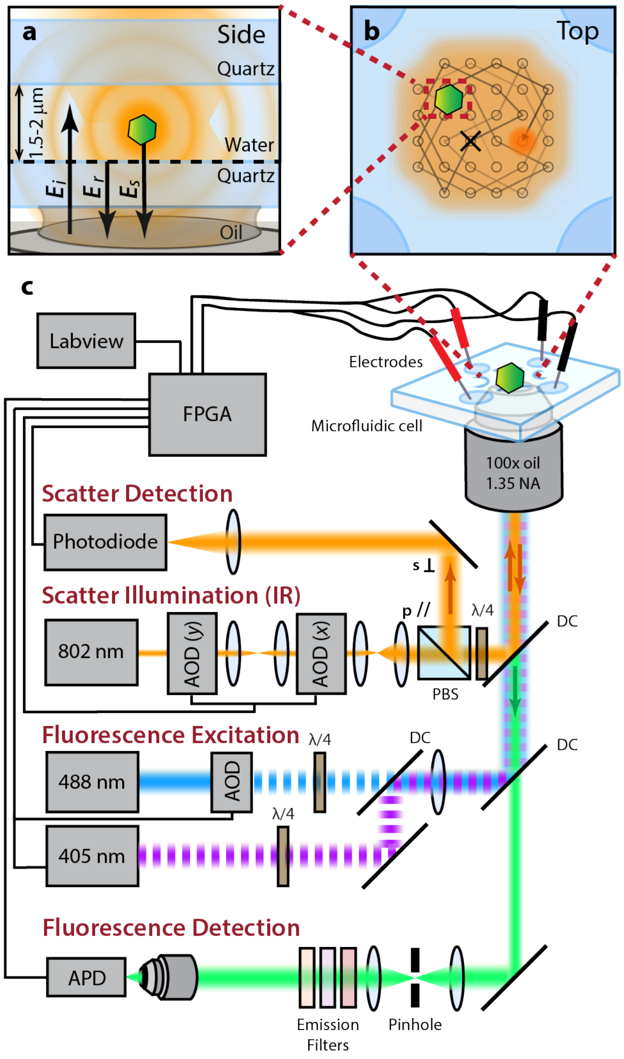


Figure 13. Overview of ISABEL Trap with interleaved fluorescence excitation. (a) A focused incident field E_i illuminates a carboxysome, which radiates the scattered electric field E_s that interferes with E_r, the reflection from the quartz-water interface. (b) Top view of the microfluidic trap, where the incident beam is scanned in a 32-spot Knight's tour pattern. Electrokinetic feedback in two dimensions is applied to the particle to push the object toward a designated location near the center of the illumination scan pattern (marked "x"). (c) Optical paths of scatter and fluorescence beams, described in more detail in Note S2 and Fig. S2. The scatter illumination beam is deflected by two acousto-optic deflectors (AODs) controlled by the FPGA; it is linearly polarized at the polarizing beam splitter (PBS), then converted to circular polarization with a quarter-wave plate (λ/4) to be back-reflected, converted back to the orthogonal linear polarization, and separated for detection on a photodiode. Position is monitored and feedback voltages are calculated on the FPGA, then applied to the solution with platinum electrodes. Simultaneously, the FPGA digitally modulates two CW fluorescence excitation lasers, alternating each ms. Fluorescence emission spanning 500-570nm is collected on an avalanche photodiode (APD) after spatial filtering with a 75µm pinhole. Detected photons are time tagged on the FPGA and labeled with the identity of the corresponding excitation laser. AOD: acousto-optic deflectors, AOM: acousto-optic modulator, PBS: polarizing beam splitter, DC: dichroic beamsplitter.

3.6.2 Results

Figure 13 shows the quartz microfluidic trapping cell and the optical layout of the ISABEL trap. Nanoparticles in aqueous solution are allowed to diffuse to the center of the cell's two crossed channels in a region 1.5-2 µm deep (Fig. 13a). When an infrared electric field E_i is incident on the particle, the back-scattered field E_s interferes with the quartz-water interfacial reflection, E_r, producing a detected intensity I_{det} given by

$$I_{det} \propto |E_r + E_s|^2 = |E_r|^2 + |E_s|^2 + 2|E_r||E_s|\cos\theta \qquad (1)$$

where θ is the phase between the reflected and scattered fields. The last term on the right-hand side of Eq. 1 represents the interferometric contribution to the intensity, which is linear in scattered field and thus linear in polarizability and mass for proteinaceous objects (51). For small particles, $|E_s|^2$ is negligible, and $|E_r|^2$ can be obtained by a background measurement when there are no particles in the trapping area. This allows for on-the-fly determination of the absolute fractional scattering contrast c

$$c = \frac{\left|I_{det} - |E_r|^2\right|}{|E_r|^2} = \frac{2|E_s\cos\theta|}{|E_r|} \qquad (2)$$

The particle position is detected by the location of the maximum contrast in the sample plane, obtained from a "Knight's tour" scan pattern of the near-IR beam steered by two acousto-optic deflectors (29). Rapid positional feedback forces are applied to the solution via voltages calculated on a Field-Programmable Gate Array (FPGA) and two pairs of platinum electrodes placed at the ends of the crossed microfluidic channels. The particle is directed to the trap center in two dimensions by electroosmosis.

In addition to the IR trapping beam, the second major difference from previous work is the addition of two visible lasers in wide-field illumination for spectroscopic measurements of the trapped object (Fig. 13c). The FPGA digitally modulates each laser power with a 2-ms alternating square wave, so that emitted photons from GFP fluorescence can be separated into two excitation channels (Note S2).

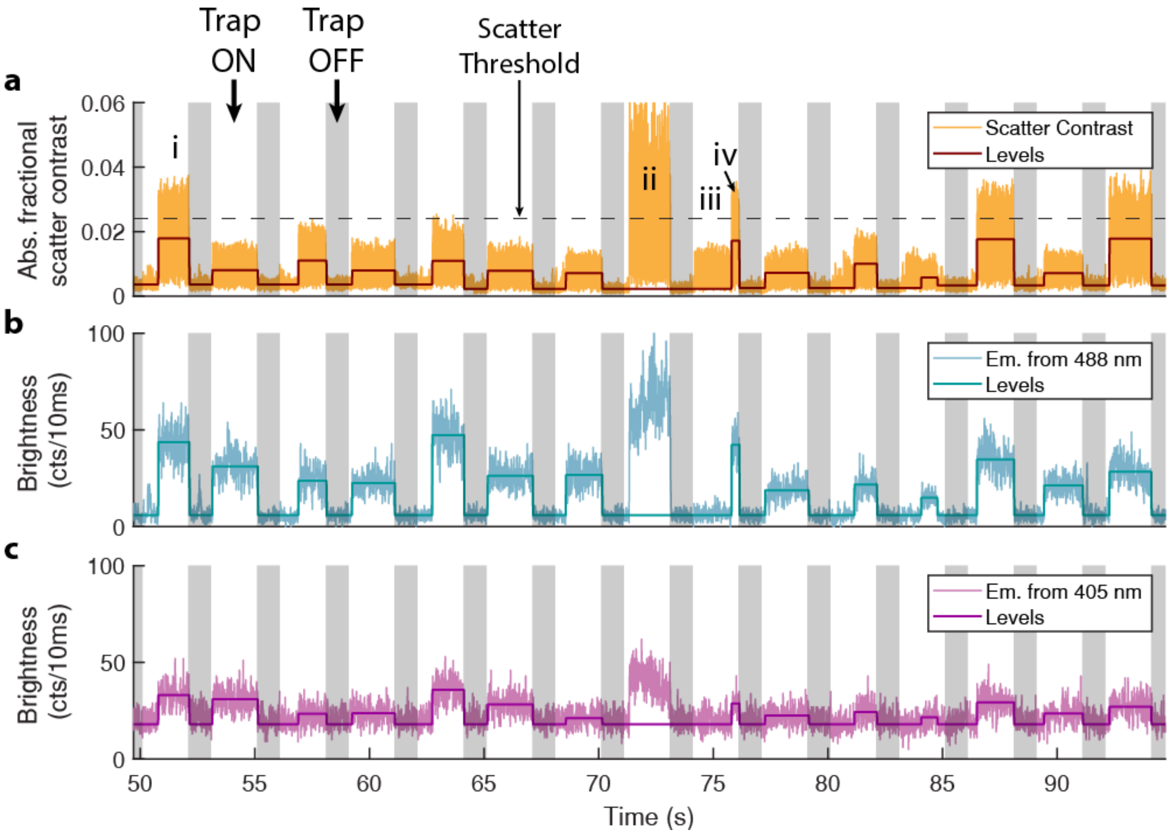


Figure 14. Representative multi-channel time traces from a carboxysome trapping experiment in air-oxidized buffer. When positional feedback is turned on (white regions), single detected particles are held at the trap center until feedback is toggled off (gray regions). Trapped roGFP2-labeled carboxysomes display signal in all three channels, such as in Event i. (a) Absolute fractional scatter contrast trace, with individual measurements for each timepoint (yellow) and average levels for an event (red) determined by the changepoints found on the 488 nm trace (b). Black dashed line indicates the scatter threshold used to reject large aggregates from analysis (Note S3). (b)-(c) 488 and 405 nm excitation channels, respectively, and corresponding average levels determined from a changepoint finding algorithm on the 488 nm trace.

Three simultaneous variables are monitored in time for individual trapped carboxysomes: absolute fractional scattering contrast (Note S3), emission from 405 nm excitation, and emission from 488 nm excitation (Fig 14). A step change in the fractional scattering contrast trace at $t \approx 51$ s (Fig. 14a, Event i) shows that when trapping turns on, a diffusing particle becomes trapped for >1s, then leaves when the feedback is turned off. For these experiments, we toggled feedback on for 2s and off for 1s to collect statistics from additional single particles, although we can trap carboxysomes for tens of seconds if desired (Fig. S3). The scattering trace of Event i shows a sudden increase in scattering contrast about a mean value (dark red line) determined by a changepoint algorithm described below, with wide fluctuations due to the evolving phase θ between E_r and E_s as the particle diffuses in the axial direction. Subsequent trapping events display various mean scattering values, indicating a range of particle sizes, including an exceptionally large particle (Event ii), which is likely an aggregate of multiple carboxysomes and is excluded from further analysis. At $t \approx 75$s (Event iii), a somewhat small object is trapped, which is then replaced by an object of higher contrast (Event iv), since anti-Brownian feedback can be applied to only one object at a time.

Simultaneously, we monitor fluorescence from each carboxysome via the interleaved 405 and 488 nm excitation (Figs. 14b and 14c, respectively). Low intensity (< 50 W/cm^2) is necessary at 405 nm to balance the roGFP2 emission rate with the light-induced photoconversion (52) of roGFP2 chromophores over extended trapping times (Figs. S3 and S4). In Event i, a steady fluorescence level is present in both channels over the trapping time, but particles are brighter and background is lower in the 488 nm channel. Fluorescence changepoints and mean fluorescence brightness were determined with a changepoint finding algorithm (53) on the 488 nm excitation trace, which provided changepoints to also find the average levels in the scattering and 405 traces. Like the various scattering levels, the fluorescence traces show a distribution of brightnesses, indicating variation in roGFP2 loading between carboxysomes. The highly scattering object in Event ii is accompanied with high brightness in both channels. The object in Event iii is non-fluorescent, possibly being an emptied carboxysome shell, a protein aggregate that remained after purification, or a dislodged piece of the polyelectrolyte passivation layer (Note S1). Because this particle is non-fluorescent, it is not identified by the algorithm. Conversely, the object in Event iv shows signal in all three channels, implying that it is a carboxysome.

The simultaneously measured levels from scattering and fluorescence provide correlated data from individual trapping events, yielding multidimensional statistics measured from carboxysomes in reducing or air-oxidized buffers (Fig. 15). Figs. 15a-15c show 2D scatter plots from carboxysomes internally reduced by 1 mM TCEP in buffer. In these plots, each point represents the average level found from an individual trapping event, and its color reflects the local density of points (23). In the Fig. 15a marginal histogram, the 488 nm fluorescence distribution is centered at ~60 counts/10ms but spans two orders of magnitude. The 2D scatter plot shows correlation between fluorescence levels and fractional scattering contrast, with subpopulations within the spread. The 2D scatter plot in Fig. 15b relating 405 nm brightness to scattering shows similar trends, though with lower brightnesses values centered at ~5 counts/10ms. The scattering contrast histogram (Fig. 15c right) appears bimodal, peaked at 0.008 and 0.015. In contrast, cryo-TEM imaging reveals that the distribution of carboxysome diameters is unimodal ($\mu \pm \sigma = 141 \pm 31$ nm, Fig. S5). Because of the approximate doubling in scattering contrast and unimodal size distribution from cryo-TEM, the higher contrast peak likely arises from trapped carboxysome dimers. The monomer and dimer populations are not readily separable in a 1D measurement, though they are better separated with our multi-dimensional correlated measurements.

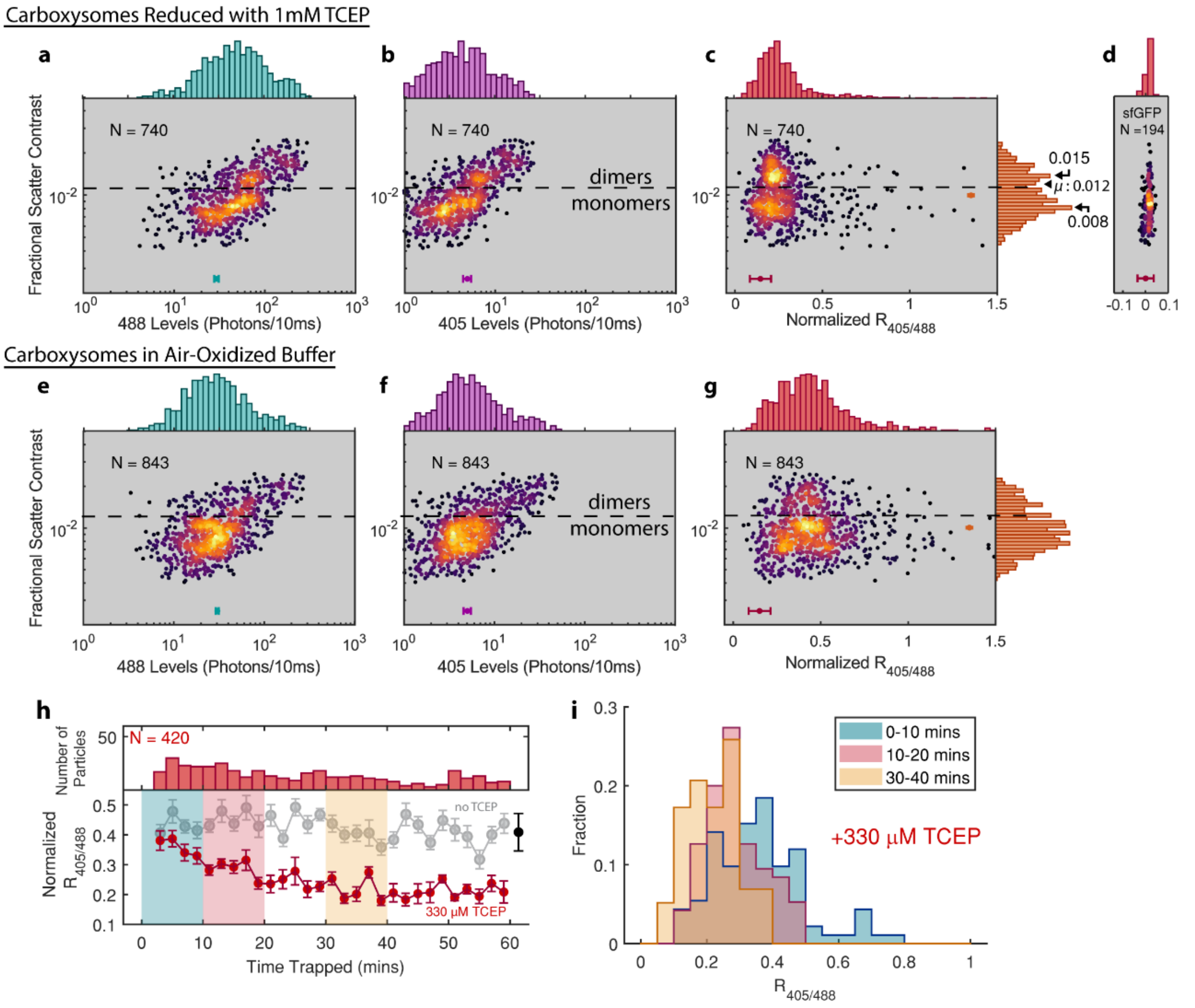


Figure 15. Multidimensional statistics of measurements on individual carboxysomes. (a)-(b) Scatter plots and marginal histogram between 488 and 405 nm fluorescence levels, respectively, with fractional scatter contrast, presented on logarithmic axes in both dimensions, due to the considerable range measured across carboxysomes. Each point corresponds to a single trapping event, colored by local density of points. The teal and magenta horizontal error bars denote the RMS standard error over all brightness levels. (c) Fractional scattering contrast *versus* ratiometric fluorescence $R_{405/488}$ for carboxysomes in 1mM TCEP reducing buffer, with scattering contrast marginal histogram to the far right. Mean is denoted by μ and the positions of the two peaks are indicated. Horizontal error bars denote the experimental RMS standard error in ratio uncertainty due to brightness fluctuations about each mean level (Supplementary Note S4 and Fig. S6). (d) The ratio-contrast scatter plot for sfGFP carboxysomes, demonstrating the narrow distribution measured for a reporter independent of redox. (e)-(g) Fluorescence, scatter contrast, and ratio scatter for carboxysomes in air-oxidized buffer, demonstrating ratios with a higher mean value and wider distribution than reduced carboxysomes. (h) Ratio kinetics measured after mixing with TCEP from individual carboxysomes. Red points correspond to ratios averaged for the number of carboxysomes shown in the top panel in 2 min windows after mixing with 330 µM TCEP, and gray points correspond to carboxysomes without TCEP reductant. Trapping starts approximately two minutes after mixing to load the sample into the cell, thus starting the

Figure 15 cont. experiment. Red and gray error bars indicate standard errors on the mean ratio measured in each time bin. Black error bar indicates the same RMS ratio error as in (g). (i) Ratio histograms from the reducing experiment, plotted in 10-minute intervals. At early times, the distribution is broad and centered at higher values, but gradually narrows and shifts to lower values over time.

Turning to redox ratios, Fig. 15c shows the 2D scatter plot correlating reduced roGFP2 fluorescence ratios $R_{405/488}$ with the fractional scattering contrast of each carboxysome. The mean of the ratio distribution is 0.25 ($\sigma = 0.16$), a comparatively low value consistent with reducing conditions (Fig. S7). To test the ratio uncertainty due to measurement error, we also trapped *E. coli* carboxysomes labeled with superfolder GFP (sfGFP), whose ratiometric fluorescence is not redox-dependent (Fig. 15d, see also Fig. S8). The tight ratio distribution from sfGFP-labeled carboxysomes indicates that the larger ratio spread in roGFP2 carboxysomes arises from ratio variation between particles. To quantify the measured ratio uncertainty, we propagated the standard errors on the two fluorescence levels into their ratio (Note S4) and present the RMS standard error over all ratios (red error bars in Fig. 15c). In sfGFP carboxysomes, the ratio spread is comparable to the RMS uncertainty, indicating that measurement uncertainty dominates the spread. However, for reduced roGFP2 carboxysomes, the ratio distribution exceeds the bounds of the RMS error, indicating other contributions to the ratio spread, discussed further below. As well, the ratio distribution for roGFP2 narrows with larger scattering contrast, attributed to increased brightness from higher roGFP2 loading in some carboxysomes. While shot noise dictates that brighter fluorescence increases the absolute noise, the relative noise influencing ratio uncertainty is decreased.

When roGFP2-carboxysomes are left in air-oxidized buffer (Figs. 15e-15g), they display higher ratios on average ($\mu = 0.46$) and a broader distribution ($\sigma = 0.23$). The mean ratio value is consistent with bulk redox ratio (Fig. S7), but the distribution shows an unexpectedly large spread. The RMS standard error on the ratios is comparable to the reduced case, but is distinctly smaller than the spread of the oxidized carboxysome ratio distribution, indicating additional heterogeneity. The ratio spread is insensitive to pH, added HCO_3^-, or added oxidant (1mM diamide, Fig S9). In particular, roGFP2-labeled carboxysomes are already in fully oxidizing environments when exposed to air. The wide spread of redox ratios likely arises from kinetics of individual roGFP2s which may occur on multiple timescales: sub-ms-timescale protonation/deprotonation kinetics (54), 10s of ms blinking into dark states (55), and the likely slower kinetics (56) of the binding and unbinding of the engineered disulfide bridge on the roGFP2 β-barrel. The capabilities of the ISABEL trap combined with additional biological constructs will allow further investigation of the heterogeneity in oxidized samples in future work.

Figs. 15h and 15i demonstrate an hour-long reduction kinetics measurement of the ISABEL trap, where fluorescence ratios are measured on individual particles after mixing air-equilibrated samples into reducing buffer (330 µM TCEP). This measurement can be employed on dilute samples or where it is important to exclude ruptured fragments, large aggregates, and free roGFP2. Here, ratios are collected from individual carboxysomes as in Fig. 14 and averaged over 2-minute intervals, thus pooling roughly

~15 carboxysomes for each time point (numbers in upper panel of Fig. 15h). After mixing carboxysomes with reducing buffer, ratios decrease on a ~15-minute timescale (red trace in Fig. 15h) and settle at the reduced ratio mean of 0.22. This measurement recapitulates the ensemble reduction kinetics (Fig. S7), indicating that the bulk measurement is not dominated by external roGFP2. Along with the mean values in each time interval, the single-particle measurements allow us to measure the ratio distribution over time (Fig. 15i). In this case, the ratios first show a broader spread ($\sigma = 0.21$ between 0-10 mins), reflecting both the initially oxidized population and its partial reduction over 10 min. The redox state of the carboxysome population shifts over time to a more equilibrated narrower distribution of reduced carboxysomes ($\sigma = 0.08$ between 30-40 mins). The capability to select individual particles in a heterogeneous sample, and the ability to measure spread of the redox ratio over time demonstrates the distinct advantages of single-particle over ensemble measurements.

3.6.3 Conclusion

In summary, we demonstrated that individual carboxysomes can be trapped in solution with active feedback using interferometric detection of their optical scattering from a near-infrared laser. We introduced rapidly interleaving 405 and 488 nm excitation lasers to monitor the ratiometric fluorescence from individual roGFP2-labeled carboxysomes, decoupling fluorescence channels from the positional monitoring to measure intermittent signals with low excitation intensity. Carboxysomes recombinantly expressed in *E. coli* display wide distributions of scattering contrasts and fluorescence brightness, which can be directly monitored by trapping individual particles. Reduced and oxidized carboxysomes show low and high values of the average redox ratios $R_{405/488}$, respectively. Controlling for chemical environment, single-carboxysome roGFP2 ratios display a wide range of values, indicative of the small numbers ($N \approx$ 3-15) of roGFP2 per carboxysome and other sources of heterogeneity, particularly evident in oxidized environments. We can observe minutes-timescale redox kinetics over the population of carboxysomes, which enables kinetic measurements for biological samples that are highly dilute or contain unwanted contributors to signal such as free roGFP2. Taken together, these experiments demonstrate the ability of the ISABEL trap to monitor nanoscale biological objects like carboxysomes, viruses, and exosomes for extended times and to expand the range of local reporter experiments that can be done in these systems.

3.6.4 Supplementary Material

See supplementary material for notes on sample preparation and characterization, ISABEL electronics and optics, and analysis and error notes. Supplementary figures show quantification of roGFP2 loading in carboxysomes, long trapping events, the impact of 405 nm illumination on ratiometric fluorescence, carboxysome sizing with cryo-EM, uncertainty statistics from trapping events, bulk reduction kinetic traces, 2D scatter plots for sfGFP carboxysomes, insensitivity of ratio distribution with buffer chemistry changes, fluorescence spectra of roGFP2 and dye solution standards, and characterization of the ISABEL PSF.

3.6.5 Acknowledgments

This work was supported in part by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences, & Biosciences Division, Physical Biosciences Program, under Award Numbers DE-FG02-07ER15892 (W.E.M.) and DE-SC0016240 (D.F.S.). P.D.D. was supported in part by the Panofsky Fellowship at the SLAC National Accelerator Laboratory and by grant 2021-234593 from the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation. D.D.P. was supported by a Stanford Graduate Fellowship.

3.7 Materials and Methods

3.7.1 *Purifications*

Standard cell growth and induction

All roGFPs used are the roGFP2 variant. Unless otherwise noted, all cells were inoculated into 50 ml of LB medium and grown overnight at 37°C to saturation. The following morning cells were back-diluted to OD600 ~0.05 in 1L LB and grown for a couple of hours until mid-log phase (~OD600 0.4-6). At mid-log cells were inoculated with 500 uM IPTG (unless another noted inducer was used) and grown overnight at 18°C to allow for protein expression. Cells were harvested the next day via centrifugation at 5000 xg for 15 minutes, the supernatant discarded, and the pellets frozen at -20°C until needed for purification.

Standard lysis and clarification

The following protocol describes standard lysis and clarification procedure for all purifications, unless otherwise noted. Cell pellets were thawed at RT until they turned to liquid sludge. Buffer was added to ~25 ml total, plus 1 mM phenylmethanesulfonyl fluoride (PMSF), 0.1 mg/ml lysozyme, and 0.1 ul/ml benzonase. Cells were lysed on an Avestin EmulsiFlex-C3 homogenizer, passing the resuspension through 3 times until it looked more translucent than opaque. Lysed cells were clarified at 27,000 xg for 30 minutes and the supernatant collected for further processing.

Standard nickel gravity column procedure for His-tagged proteins

~2-5 ml of Ni-NTA slurry was added to a gravity column and equilibrated with 50/300mM Tris/NaCl + 20 mM imidazole. Clarified protein was added, then washed with 1-5 volumes of 50/300mM Tris/NaCl + 60 mM imidazole. Protein was eluted with 50/300mM Tris/NaCl + 300 mM imidazole and collected in fractions. Pooled fractions were concentrated and desalted into 50/150mM Tris/NaCl pH 7.4.

Standard carboxysome purification

The following protocol describes the standard carboxysome purification method. The buffer used throughout is TEMB (10 mM Tris-HCl, pH8, 10 mM $MgCl_2$, 1 mM EDTA, adjusted to pH 8.0). Cell pellets were lysed in BPER (volume per package instructions) with added $MgCl_2$, EDTA, PMSF, lysozyme, and benzonase at the concentrations listed above for 45 min – 1 hour at room temperature. Lysed cells were spun at 12,000 xg for 15-30 minutes to pellet cell debris and the supernatant saved for subsequent steps. Supernatant was spun at 40,000 xg for 30 minutes to pellet carboxysomes, and the supernatant discarded. A pipette was used to remove what didn't fully pour out. 1500 ul TEMB was added to the pellet and the tube gently shaken on ice for 30 minutes to overnight. Pellets were resuspended by pipetting gently up and down. Resuspended pellets were spun at 900 xg for 3 minutes to pellet any remaining insoluble junk and the supernatant saved. Supernatant was loaded onto a 10-50% TEMB sucrose gradient (gradients were stacked 5 ml layers of 10, 20, 30, 40, and 50% sucrose, with 50% on the bottom and 10% on top). Sucrose gradients were spun at 105,000 xg for 35 minutes. Fractions were collected 1 ml at a time and run on an SDS-PAGE gel to assess purification quality and location in the gradient. Carboxysome fractions, usually ~13-24, were pooled and spun at 100,000 xg for 90 minutes. The pellet, usually a bit gel-like and glassy looking, was resuspended in 500-1500 ul TEMB buffer and stored at 4°C.

His-roGFP (pJB17)
BL21-AI cells were transformed with pJB17 and induced with 0.1% arabinose at mid-log phase. Cells were lysed and clarified per the standard methods. Protein was purified via the standard nickel gravity column method.

N48-roGFP-FLAG-6xHis (pJB34)
BL21 cells were transformed with pJB34 and induced, lysed, and clarified per the standard methods. Cells were induced with 100 nM aTc. Protein was purified via the standard nickel gravity column method.

Thioredoxin and Thioredoxin Reductase (TrxA; pJB19 and TrxB; pJB21)
BL21 cells were transformed with pJB19 or 21 and induced, lysed, and clarified per the standard methods. Protein was purified via the standard nickel gravity column method.

N48-roGFP carboxysomes (pJB31 + pHnCB10)
E. coli BW 25113 cells were transformed with pJB31 and pHnCB10 to allow for separately inducible expression of N48-roGFP and carboxysomes. Cells were induced with 500 uM IPTG (for carboxysomes) and 100 nM anhydrotetracycline (aTc). Carboxysomes were purified via the standard method.

ΔPentamer N48-roGFP carboxysomes (pJB31 + pJB33)
Protocol is the same as for N48-roGFP carboxysomes except cells were transformed with pJB31 and pJB33 (pHnCB10 with pentamers deleted).

3.7.2 Assays

roGFP fluorescence assay – control experiments
All roGFP assays were run on flat black, clear bottom 96 well plates (Corning Costar 3631) in a Tecan Infinite M1000 plate reader. Plate reader settings were 5 nm bandwidth, 50 flash, 10 ms settle time, and 110 gain. Figure 3A used 2 uM roGFP with 1 mM of each indicated redox agent in 50 mM Tris, 150 mM NaCl buffer. Figure 3B used 200 nM roGFP with the indicated concentrations of DTT in 50 mM Tris, 150 mM NaCl buffer. Figure 3C and D used 150 nM roGFP with 150 uM Trx, 2.3 uM TrxR, and 150 uM NADPH in 50 mM Tris, 150 mM NaCl buffer. Figure 5 used 1 mg/ml N48-roGFP carboxysomes, 0.15 uM roGFP, and 0.2 mM DTT. Figure 6 used 0.15 uM roGFP, 0.2 mM DTT, and 1 mg/ml BSA. Figure 7 used 0.15 uM roGFP, 0.15 uM N48-roGFP, 0.15

mM DTT, 0.15 mM Diamide, and 1 mg/ml BSA in 50/150mM Tris/NaCl. Figure 10C used 0.15 uM N48-roGFP, 0.2 mM DTT, and 2 mg/ml carboxysomes.

roGFP fluorescence assays.

All assays were run on flat black, clear bottom 96 well plates (Corning Costar 3631) in a Tecan Infinite M1000 plate reader. Plate reader settings were 5 nm bandwidth, 50 flash, 10 ms settle time, and 110 gain. Each well had 100 ul of sample. All reducing agents were prepared fresh (or thawed fresh, for Trx). The concentration of N48-roGFP was 0.15 uM with 1 mg/ml BSA added. The concentration of N48-roGFP carboxysomes and Δpentamer carboxysomes was 2 mg/ml (no 340 nm baseline correction). For the Trx sample, NADPH was N2 sparged and stored frozen at 0.05M; tubes of this were thawed before adding it to the assay at a concentration of 250 uM. For the Δpentamer thioredoxin assay, 50 uM Trx and 125 uM NADPH was used. NaOH was added to glutathione solubilized in buffer until the pH was brought up to 8. For the Δpentamer + metabolite assay, the following concentrations of metabolites were used: 276 uM ATP, 2.5 mM RuBP, 2.5 mM αKG, 1.3 mM GTP, 740 uM ADP, 500 uM NADPH. For all assays, buffer was 50/150mM Tris/NaCl.

Broken carboxysome assay (Figure 10)

Carboxysomes were spun at 20,000 xg for 10 minutes and the supernatant removed. Pellets were frozen at -20°C for 30 minutes. Pellets were thawed at room temperature, then resuspended in 200 ul 50/150mM Tris/NaCl. The concentration of DTT was 150 uM, and of carboxysomes was 2 mg/ml (no A280 baseline correction).

roGFP *in vivo* experiments (Figure 4)

Experiments were run on a Tecan Infinite M1000 plate reader. Plate reader settings were 5 nm bandwidth, 50 flash, 10 ms settle time, and 110 gain. Saturated culture was back-diluted 100x to ~OD600 0.3. 150 ul of culture was added per well, x6 technical replicates. Cells were allowed to grow for 2 hours before adding inducer (500 uM IPTG, 100 nM aTc).

3.7.3 Data analysis

All data were plotted in Prism. The extra sum-of-squares F test was used to compare a one-phase exponential decay model with a two-phase exponential decay model to determine which one best fit the data. In brief, the sum-of-squares (sum of the square of the difference between the points and the model curve) improves with more parameters; the F test determines if the improvement in the sum-of-squares with a two-phase decay is within the amount expected by chance, or if it is outside the expected chance interval, in which case the two-phase decay would be a better fit than the simpler one-phase model. Fitted and reported curves were chosen based on the results of the F test. K_{fast}, half life, and span were reported by Prism. For Figure 11, K_{fast} values were measured for the following conditions: 150 uM DTT, 150 uM TCEP, 50 mM BME, and 50 uM Trx. Error bars are 95% confidence intervals reported by Prism.

3.7.4 Sequences

His-roGFP (pJB17)
MHHHHHHSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFISTTGK
LPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTR
AEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNCHNVYIMADKQKNGIKVNFKIRH
NIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTCSALSKDPNEKRDHMVLLEFVTAA
GITHGMDELYK*

N48-roGFP-FLAG (pJB31)
MNTRNTRSKQRAPFGVSSSVKPRLDLIEQAPNPAYDRHPACITLPERTCRHPLGSSSK
GEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFISTTGKLPVPWPTLVT
TLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL
VNRIELKGIDFKEDGNILGHKLEYNYNCHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLA
DHYQQNTPIGDGPVLLPDNHYLSTCSALSKDPNEKRDHMVLLEFVTAAGITHGMDELY
KDYKDDDDK*

N48-roGFP-FLAG-6xHis (pJB34)
MNTRNTRSKQRAPFGVSSSVKPRLDLIEQAPNPAYDRHPACITLPERTCRHPLGSSSK
GEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFISTTGKLPVPWPTLVT
TLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTL
VNRIELKGIDFKEDGNILGHKLEYNYNCHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLA
DHYQQNTPIGDGPVLLPDNHYLSTCSALSKDPNEKRDHMVLLEFVTAAGITHGMDELY
KDYKDDDDKHHHHHH*

His-TrxA (pJB19)
MHHHHHHSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGK
LTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLA*

His-TrxB (pJB21)
MHHHHHHGTTKHSKLLILGSGPAGYTAAVYAARANLQPVLITGMEKGGQLTTTTEVEN
WPGDPNDLTGPLLMERMHEHATKFETEIIFDHINKVDLQNRPFRLNGDNGEYTCDALII
ATGASARYLGLPSEEAFKGRGVSACATCDGFFYRNQKVAVIGGGNTAVEEALYLSNIA
SEVHLIHRRDGFRAEKILIKRLMDKVENGNIILHTNRTLEEVTGDQMGVTGVRLRDTQN
SDNIESLDVAGLFVAIGHSPNTAIFEGQLELENGYIKVQSGIHGNATQTSIPGVFAAGDV
MDHIYRQAITSAGTGCMAALDAERYLDGLADAK*

3.8 References

1. Price, G. D., and Badger, M. R. (1989) Expression of Human Carbonic Anhydrase in the Cyanobacterium Synechococcus PCC7942 Creates a High CO(2)-Requiring Phenotype : Evidence for a Central Role for Carboxysomes in the CO(2) Concentrating Mechanism. *Plant Physiol.* **91**, 505–513
2. Peña, K. L., Castel, S. E., de Araujo, C., Espie, G. S., and Kimber, M. S. (2010) Structural basis of the oxidative activation of the carboxysomal γ-carbonic anhydrase, CcmM. *Proceedings of the National Academy of Sciences*. **107**, 2455–2460
3. Chen, A. H., Robinson-Mosher, A., Savage, D. F., Silver, P. A., and Polka, J. K. (2013) The bacterial carbon-fixing organelle is formed by shell envelopment of preassembled cargo. *PLoS One*. **8**, e76127
4. Heinhorst, S., Williams, E. B., Cai, F., Murin, C. D., Shively, J. M., and Cannon, G. C. (2006) Characterization of the carboxysomal carbonic anhydrase CsoSCA from Halothiobacillus neapolitanus. *J. Bacteriol.* **188**, 8087–8094
5. Wang, H., Yan, X., Aigner, H., Bracher, A., Nguyen, N. D., Hee, W. Y., Long, B. M., Price, G. D., Hartl, F. U., and Hayer-Hartl, M. (2019) Rubisco condensate formation by CcmM in Î²-carboxysome biogenesis. *Nature*. **566**, 131–135
6. Yanai, I., and Lercher, M. (2020) The two languages of science. *Genome Biol.* **21**, 147
7. Li, T., Jiang, Q., Huang, J., Aitchison, C. M., Huang, F., Yang, M., Dykes, G. F., He, H.-L., Wang, Q., Sprick, R. S., Cooper, A. I., and Liu, L.-N. (2020) Reprogramming bacterial protein organelles as a nanoreactor for hydrogen production. *Nat. Commun.* **11**, 5448
8. Mahinthichaichan, P., Morris, D. M., Wang, Y., Jensen, G. J., and Tajkhorshid, E. (2018) Selective permeability of carboxysome shell pores to anionic molecules. *J. Phys. Chem. B*. **122**, 9110–9118
9. Mangan, N. M., Flamholz, A., Hood, R. D., Milo, R., and Savage, D. F. (2016) pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5354-62
10. Tsai, Y., Sawaya, M. R., Cannon, G. C., Cai, F., Williams, E. B., Heinhorst, S., Kerfeld, C. A., and Yeates, T. O. (2007) Structural analysis of CsoS1A and the protein shell of the Halothiobacillus neapolitanus carboxysome. *PLoS Biol.* **5**, e144
11. Tanaka, S., Sawaya, M. R., Phillips, M., and Yeates, T. O. (2009) Insights from multiple structures of the shell proteins from the beta-carboxysome. *Protein Sci.* **18**, 108–120
12. Kinney, J. N., Axen, S. D., and Kerfeld, C. A. (2011) Comparative analysis of carboxysome shell proteins. *Photosynth. Res.* **109**, 21–32
13. Badger, M. R., Bassett, M., and Comins, H. N. (1985) A model for HCO(3) accumulation and photosynthesis in the cyanobacterium synechococcus sp: theoretical predictions and experimental observations. *Plant Physiol.* **77**, 465–471
14. Reinhold, L., Kosloff, R., and Kaplan, A. (1991) A model for inorganic carbon fluxes and photosynthesis in cyanobacterial carboxysomes. *Can. J. Bot.* **69**, 984–988

15. Lavania, A. A., Carpenter, W. B., Oltrogge, L. M., Perez, D., Turnšek, J. B., Savage, D. F., and Moerner, W. E. (2022) Exploring Masses and Internal Mass Distributions of Single Carboxysomes in Free Solution Using Fluorescence and Interferometric Scattering in an Anti-Brownian Trap. *J. Phys. Chem. B.* 10.1021/acs.jpcb.2c05939
16. Carpenter, W. B., Lavania, A. A., Borden, J. S., Oltrogge, L. M., Perez, D., Dahlberg, P. D., Savage, D. F., and Moerner, W. E. (2022) Ratiometric Sensing of Redox Environments Inside Individual Carboxysomes Trapped in Solution. *J. Phys. Chem. Lett.* **13**, 4455–4462
17. Hanson, G. T., Aggeler, R., Oglesbee, D., Cannon, M., Capaldi, R. A., Tsien, R. Y., and Remington, S. J. (2004) Investigating mitochondrial redox potential with redox-sensitive green fluorescent protein indicators. *J. Biol. Chem.* **279**, 13044–13053
18. Blikstad, C., Dugan, E. J., Laughlin, T. G., Turnšek, J. B., Liu, M. D., Shoemaker, S. R., Vogiatzi, N., Remis, J. P., and Savage, D. F. (2023) Identification of a carbonic anhydrase–Rubisco complex within the alpha-carboxysome. *Proceedings of the National Academy of Sciences*. **120**, e2308600120
19. Cai, F., Menon, B. B., Cannon, G. C., Curry, K. J., Shively, J. M., and Heinhorst, S. (2009) The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO2 leakage barrier. *PLoS One*. **4**, e7521
20. Okumus, B., Wilson, T. J., Lilley, D. M. J., and Ha, T. (2004) Vesicle encapsulation studies reveal that single molecule ribozyme heterogeneities are intrinsic. *Biophys. J.* **87**, 2798–2806
21. Rasnik, I., McKinney, S. A., and Ha, T. (2005) Surfaces and orientations: much to FRET about? *Acc.Chem.Res.* **38**, 542–548
22. Goldsmith, R. H., and Moerner, W. E. (2010) Watching conformational- and photodynamics of single fluorescent proteins in solution. *Nature Chemistry*. **2**, 179–186
23. Squires, A. H., Dahlberg, P. D., Liu, H., Magdaong, N. C. M., Blankenship, R. E., and Moerner, W. E. (2019) Single-molecule trapping and spectroscopy reveals photophysical heterogeneity of phycobilisomes quenched by Orange Carotenoid Protein. *Nature Communications*. 10.1038/s41467-019-09084-2
24. Friedel, M., Baumketner, A., and Shea, J. E. (2006) Effects of surface tethering on protein folding mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8396–8401
25. Cang, H., Wong, C. M., Xu, C. S., Rizvi, A. H., and Yang, H. (2006) Confocal three dimensional tracking of a single nanoparticle with concurrent spectroscopic readouts. *Appl. Phys. Lett.* **88**, 223901
26. McHale, K., Berglund, A. J., and Mabuchi, H. (2007) Quantum dot photon statistics measured by three-dimensional particle tracking. *Nano Lett.* **7**, 3535–3539
27. Hou, S., Exell, J., and Welsher, K. (2020) Real-time 3D single molecule tracking. *Nature Communications*. **11**, 3607
28. Squires, A. H., Cohen, A. E., and Moerner, W. E. (2018) Anti-Brownian Traps. in *Encyclopedia of Biophysics* (Roberts, G. C. K., and Watts, A. eds), Springer, Berlin, Heidelberg

29. Wang, Q., and Moerner, W. E. (2010) Optimal strategy for trapping single fluorescent molecules in solution using the ABEL trap. *Applied Physics B-Lasers and Optics*. **99**, 23–30
30. Su, B., Düser, M. G., Zarrabi, N., Heitkamp, T., Ilka Starke, I., and Börsch, M. (2015) Observing conformations of single F0F1-ATP synthases in a fast anti-Brownian electrokinetic trap. *Proc. SPIE*
31. Cohen, A. E., and Moerner, W. E. (2006) Suppressing Brownian motion of individual biomolecules in solution. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4362–4365
32. Goldsmith, R. H., Tabares, L. C., Kostrz, D., Dennison, C., Aartsma, T. J., Canters, G. W., and Moerner, W. E. (2011) Redox cycling and kinetic analysis of single molecules of solution-phase nitrite reductase. *Proceedings of the National Academy of Sciences of the United States of America*. **108**, 17269–17274
33. Schlau-Cohen, G., Yang, H.-Y., Krueger, T. P. J., Xu, P., Gwizdala, M., van Grondelle, R., Croce, R., and Moerner, W. E. (2015) Single-molecule identification of quenched and unquenched states of LHCII. *The Journal of Physical Chemistry Letters*. **6**, 860–867
34. Schlau-Cohen, G. S., Bockenhauer, S., Wang, Q., and Moerner, W. E. (2014) Single-molecule spectroscopy of photosynthetic proteins in solution: exploration of structure–function relationships. *Chem. Sci.* **5**, 2933–2939
35. Schlau-Cohen, G. S., Wang, Q., Southall, J., Cogdell, R. J., and Moerner, W. E. (2013) Single-molecule spectroscopy reveals photosynthetic LH2 complexes switch between emissive states. *Proceedings of the National Academy of Sciences of the United States of America*. **110**, 10899–10903
36. Squires, A. H., Wang, Q., Dahlberg, P. D., and Moerner, W. E. (2022) A bottom-up perspective on photodynamics and photoprotection in light-harvesting complexes using anti-Brownian trapping. *Journal of Chemical Physics*. **156**, 070901
37. Wang, Q., and Moerner, W. E. (2013) Lifetime and spectrally resolved characterization of the photodynamics of single fluorophores in solution using the Anti-Brownian Electrokinetic Trap. *Journal of Physical Chemistry B*. **117**, 4641–4648
38. Squires, A. H., Lavania, A. A., Dahlberg, P. D., and Moerner, W. E. (2019) Interferometric scattering enables fluorescence-free electrokinetic trapping of single nanoparticles in free solution. *Nano Letters*. **19**, 4112–4117
39. Gemeinhardt, A., McDonald, M. P., König, K., Aigner, M., Mackensen, A., and Sandoghdar, V. (2018) Label-free imaging of single proteins secreted from living cells via iSCAT microscopy. *JoVE (Journal of Visualized Experiments)*
40. Lindfors, K., Kalkbrenner, T., Stoller, P., and Sandoghdar, V. (2004) Detection and spectroscopy of gold nanoparticles using supercontinuum white light confocal microscopy. *Physical Review Letters*. **93**, 037401
41. Taylor, R. W., and Sandoghdar, V. (2019) Interferometric scattering microscopy: seeing single nanoparticles and molecules via Rayleigh scattering. *Nano Letters*. **19**, 4827

42. Bonacci, W., Teng, P. K., Afonso, B., Niederholtmeyer, H., Grob, P., Silver, P. A., and Savage, D. F. (2012) Modularity of a carbon-fixing protein organelle. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 478–483
43. Oltrogge, L. M., Chaijarasphong, T., Chen, A. W., Bolin, E. R., Marqusee, S., and Savage, D. F. (2020) Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation. *Nat. Struct. Mol. Biol.* **27**, 281–287
44. Shively, J. M., Ball, F., Brown, D. H., and Saunders, R. E. (1973) Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of Thiobacillus neapolitanus. *Science*. **182**, 584–586
45. Heinhorst, S., Cannon, G. C., and Shively, J. M. (2006) Carboxysomes and Carboxysome-like Inclusions. in *Complex intracellular structures in prokaryotes* (Shively, J. M. ed), pp. 141–165, Microbiology Monographs, Springer Berlin Heidelberg, Berlin, Heidelberg, **2**, 141–165
46. Yeates, T. O., Kerfeld, C. A., Heinhorst, S., Cannon, G. C., and Shively, J. M. (2008) Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nature Reviews Microbiology*. **6**, 681–691
47. Iancu, C. V., Ding, H. J., Morris, D. M., Dias, D. P., Gonzales, A. D., Martino, A., and Jensen, G. J. (2007) The structure of isolated Synechococcus strain WH8102 carboxysomes as revealed by electron cryotomography. *J. Mol. Biol.* **372**, 764–773
48. Dou, Z., Heinhorst, S., Williams, E. B., Murin, C. D., Shively, J. M., and Cannon, G. C. (2008) CO2 fixation kinetics of Halothiobacillus neapolitanus mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO2. *J. Biol. Chem.* **283**, 10377–10384
49. Faulkner, M., Szabo, I., Weetman, S. L., Sicard, F., Huber, R. G., Bond, P. J., Rosta, E., and Liu, L.-N. (2020) Molecular simulations unravel the molecular principles that mediate selective permeability of carboxysome shell protein. *Scientific Reports*. **10**, 17501
50. Borden, J. S., and Savage, D. F. (2021) New discoveries expand possibilities for carboxysome engineering. *Curr. Opin. Microbiol.* **61**, 58–66
51. Young, G., Hundt, N., Cole, D., Fineberg, A., Andrecka, J., Tyler, A., Olerinyova, A., Ansari, A., Marklund, E. G., Collier, M. P., Chandler, S. A., Tkachenko, O., Allen, J., Crispin, M., Billington, N., Takagi, Y., Sellers, J. R., Eichmann, C., Selenko, P., Frey, L., Riek, R., Galpin, M. R., Struwe, W. B., Benesch, J. L. P., and Kukura, P. (2018) Quantitative mass imaging of single biological macromolecules. *Science*. **360**, 423–427
52. Chattoraj, M., King, B. A., Bublitz, G. U., and Boxer, S. G. (1996) Ultra-fast excited state dynamics in green fluorescent protein: Multiple states and proton transfer. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8362–8367
53. Watkins, L. P., and Yang, H. (2005) Detection of intensity change points in time-resolved single-molecule measurements. *J. Phys. Chem. B.* **109**, 617–628
54. Oltrogge, L. M., Wang, Q., and Boxer, S. G. (2014) Ground-state proton transfer kinetics in green fluorescent protein. *Biochemistry*. **53**, 5947–5957
55. Schwille, P., Kummer, S., Heikal, A. A., Moerner, W. E., and Webb, W. W. (2000) Fluorescence correlation spectroscopy reveals fast optical excitation-driven

intermolecular dynamics of yellow fluorescent proteins. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 151–156

56. Cannon, M., and Remington, S. J. (2005) Re-engineering redox-sensitive green fluorescent protein for improved response rate. *Protein Science*. **15**, 45–57

Chapter 4:
Conclusion

4.1 Summary and Future Directions

It's possible to feel like we've discovered almost everything there is to know about carboxysomes. Every protein in the operon has an assigned role, even the long puzzling CsoS2 (1–6). We more or less know how the components interact (6–8). We've checked the Feynman "what I cannot create I do not understand" box with functional carboxysome reconstitution in *E. coli* (9, 10). Cryo-EM studies have allowed us to visualize the carboxysome shell and Rubisco on the interior (11–13), and even managed to give us a picture of just how stringy CsoS2 looks when bound to shell (14).

But do we really know everything? Prior to this work, we didn't know the role of the CsoS2 Middle Region and its highly conserved residues; we now know that CsoS2 binds to shell proteins in a highly sequence-specific manner to coordinate carboxysome assembly (Chapter 2). We didn't know if α-carboxysomes could be reconstituted *in vitro*; we now know that it's possible to form phase-separated carboxysome condensates (Chapter 2). We didn't know how to quantify and measure the permeability and redox of carboxysomes; we now have established two distinct methods to evaluate these important parameters (Chapter 3).

Though I found answers, along the way I uncovered even more questions. The literature pile is high, but there is undeniably still a wealth of knowledge waiting to be discovered. The following list lays out what I think are the most interesting outstanding questions in the carboxysome field.

4.1.1 Permeability

We still don't know how carboxysome permeability works! And it's arguably the most crucial part of the whole carbon concentrating mechanism! For the Calvin cycle to function, the carboxysome must be permeable to HCO_3^-, RuBP, and 3-PG, and likely ATP to power rubisco activase (15, 16). It must also be impermeable to CO_2 (4, 17, 18). Almost every shell protein has a structure, but no relevant metabolite has ever been visualized in the center of the pore; the closest has been a sulfate ion in the crystal structures of major hexameric proteins CsoS1A (α-carboxysomes) and CcmK1 (β-carboxysomes) (3, 19). Various mathematical models and molecular dynamics simulations of permeability exist (18, 20–22), yet assumptions must be made to construct such models.

Chapter 3 of this thesis explores an *in vitro* approach to measuring carboxysomal permeability, both at the bulk level and single-particle level. Despite being able to detect and measure *single carboxysomes*, the results bucked long-held hypotheses: the reducing agent TCEP was able to enter and reduce roGFP inside carboxysomes. In bulk assays, even the protein thioredoxin was able to access roGFP, which is far too large to enter via the hexameric pores. There are three possible explanations: (1) purified carboxysomes used in the assay were broken, allowing entry of larger reducing agents, or (2) carboxysomes were not broken, and the results indicated a mechanism of entry for larger molecules, or (3) the assay contained a mixture of broken and unbroken carboxysomes. Option 2 is unlikely based on the results of Cai et al. 2009, in which they demonstrated that a pentamer deletion, which makes porous carboxysomes, gives a high-CO_2 phenotype (4). Option 3 is less likely as the single-particle analysis should have been able to distinguish broken from unbroken carboxysomes, and no reduced vs. oxidized double population was observed. So option 1 remains the most likely, which is disappointing because it only proves that the method doesn't work.

Given the results of Chapter 3, future research on carboxysome permeability will be most fruitful if undertaken *in vivo*, and if possible in an organism that natively makes carboxysomes and requires them for growth. This will ensure that all carboxysomes studied are assembled correctly and have natural permeability properties. A similar Goldilocks-style approach could be taken, though targeted carboxysome reporter proteins would require a degradation tag to eliminate unencapsulated protein. One possible reporter is the Halo protein, which fluoresces upon binding to an array of small molecules of different sizes that would have to traverse the pores.

Carboxysomes are smaller than the diffraction limit, and this experiment would require super-resolution microscopy to visualize carboxysomes in *H. neapolitanus*. A long-term approach would be to establish a new model organism that produces carboxysomes at the upper size limit of ~400 nm in diameter, such as in *Anabaena variabilis* M3 (see Fig. 2D in Rae et al. 2013 (23)) and could be imaged with more accessible light microscopy.

Research on this topic is more relevant than ever given the recent discovery of shell proteins surrounding the diatom pyrenoid (24, 25). Though they have no structural or sequence relationship to carboxysome proteins, it establishes that a protein-based barrier has convergently evolved in eukaryotes that is necessary for a functioning biophysical CCM. Future work on permeability in both systems will lead to meaningful advances on how protein barriers control the flux of metabolites and enable biophysical CO_2 concentration.

4.1.2 Redox

The role of redox reactions in carboxysome assembly and function is still only partially understood. Many, many papers will tell you that the carboxysome has an oxidizing interior, and plenty of evidence points towards this being true (see basically this whole thesis for a more in-depth discussion on this) (26, 27). However, the biggest counterpoint to this, with α-carboxysomes, at least, is that mutating the CsoS2 cysteines to serines *in vivo* had no effect on the growth of the native organism. It's also contradictory that though the carboxysome is supposedly impermeable to oxygen (as many papers claim), it has an oxidizing interior. Is there a way for both to simultaneously be true? I'll venture a somewhat creative hypothesis, but where else to do that than in a thesis? It's been shown that phase-separated droplets have an electric potential gradient at the liquid-liquid interface, which can drive spontaneous redox reactions without the presence of cellular electron carriers (28). This can result in the oxidation of hydroxide ions, ultimately leading to H_2O_2 production. This may be a way for carboxysomes to, well, have their oxygen and eat it to. Or rather, not eat it, since there's no O_2 around to compete with CO_2 for the rubisco active site.

I think the best future approach to this problem will be an in-cell assay with an encapsulated redox reporter, either roGFP or alkaline phosphatase. Both will need to have a degradation tag, to eliminate unencapsulated protein, and super-resolution microscopy will be necessary to visualize individual carboxysomes. But this may give more reliable data than *in vitro* methods (see Ch. 3) and can be done in the host organism, guaranteeing that carboxysomes are assembling correctly. Cell-permeable redox dyes can also be used, such as PL-1, which detects H_2O_2 production (28). Another approach to understanding redox would be to cap free cysteines using a reagent such as iodoacetamide, and map the bonded vs. free cysteines using mass spectrometry. This could confirm whether there is specific or nonspecific disulfide bonding occurring in the carboxysome, especially given the results of Ch. 2, where CsoS2 displayed characteristics of intra-molecular disulfide bonding.

4.1.3 Structure

It goes without saying that it would be super dope to have a full structure of a native, wild-type carboxysome. Again, many advancements have been made in this area: we can see shell, Rubisco, and even the NTD and CTD of CsoS2, which is already pretty unbelievable (11–14). Seeing the MR is the next frontier, as well as obtaining structures for β-carboxysomes and achieving the dream of seeing metabolites transiting pores.

4.1.4 Plant engineering

People have been talking about this for over a decade (29–32), and slowly but surely folks are doing the work to make it happen (33–38). But it involves whole carbon concentrating mechanism (CCM) engineering such as introducing bicarbonate transporters into the chloroplast membrane and knocking out stromal carbonic anhydrases, in addition to carboxysome expression and function. Though we're still quite far from achieving carboxysome-based biophysical CO_2 fixation in plants, this sub-field has perhaps had an outsize impact on carboxysome researchers as the most digestible answer to the question "why study carboxysomes?" "So, what is the *application*?", asks a grant, or maybe your mom at Thanksgiving. We'll put it in a plant, make it better, and solve global climate change, obviously!

4.1.5 In vitro carboxysomes

Producing carboxysomes *in vitro* would be the ultimate confirmation that we understand how all the protein components associate and the necessary reaction conditions to spark the assembly process. This work demonstrates the most complete reconstitution of an α-carboxysome to date (so far, as of this writing), though condensates are hardly carboxysomes as I note in the Discussion of Ch. 2; that I was able to make condensates, but not carboxysomes, indicates that we are still missing pieces of crucial information about how carboxysomes assemble. One orthogonal approach would be to do *in vitro* transcription-translation (txtl) with DNA of the carboxysome operon as input. If ratiometric control of protein concentration emerges on the transcriptional level, as appears to be true (39), then txtl would incorporate that native gene regulation. If kinetics or order of operations is important, where 5' operonic proteins (Rubisco, CsoS2) associate first before the production of 3' operonic proteins (shells), txtl would also control for this. It would also be interesting to test the functionality of *in vitro* carboxysomes, both for Rubisco activity and CO_2 concentration.

4.1.6 Carboxysome disassembly

This thesis talks a lot about carboxysome assembly, but very little is known about carboxysome *disassembly*. It's well documented that they last an average of 3-5 generations, but some have been observed to persist up to 10 generations or more (40, 41). Understanding carboxysome disassembly brings up important questions of carboxysome maintenance and persistence. Are there signaling pathways or certain environmental conditions that initiate disassembly? What cellular components and proteins participate in disassembly? What makes certain carboxysomes more prone to disassembly than others, or the opposite, more stable than others? While interesting questions on their own, answering these will inform carboxysome engineering efforts both *in vivo* and *in vitro*.

4.1.7 Regulation of carboxysome expression

Unlike plants, cyanobacteria and other carboxysome-containing bacteria can turn expression of their biophysical CCM on and off. We exploit this fact in our selection assays; if an essential carboxysome component is mutated, *H. neapolitanus* can still grow just fine in high CO_2 (see Ch. 2). Luke has unpublished TEM data showing that cells grown in high CO_2 do not have carboxysomes. This means that there is an upstream signaling pathway that senses the CO_2 (or bicarbonate) concentration and regulates a cellular response. Understanding this pathway would be very relevant for designing synbio-style engineered organisms to have a specific output in response to the environmental CO_2 concentration. The transcription factors that likely regulate operon expression are probably the ones identified in Jack's screen – two LysR and one Crp/Fnr type regulators (42). I think it's pretty obvious someone should work on this, and lucky for you, I'm done with my PhD, so there's no competition!

4.1.8 Hybrid carboxysomes with faster Rubiscos

Engineering carboxysomes to encapsulate faster Rubiscos would enable biological carbon fixation to happen at a much higher rate in atmospheric concentrations of CO_2 (~400 ppm). Rubiscos have a k_{cat} vs. specificity ($S_{c/o}$) tradeoff whereby the faster Rubiscos are less able to distinguish between CO_2 and O_2 as their substrates (43). These Rubiscos tend to be found in organisms that live in high CO_2 environments, or that create one with a biophysical CCM. Six of the seven fastest Rubiscos known come from organisms that grow under low-O_2 or variable O_2 conditions (44). The fastest Rubiscos known today have a k_{cat} of ~15-22 s^{-1} and $S_{C/O}$ of ~12-20, while the average plant Rubisco has a k_{cat} of ~3.3 s^{-1} and $S_{C/O}$ of 80-120 uM (44). Encapsulating these fastest Rubiscos in carboxysomes would greatly increase the catalytic efficiency of carbon fixation at atmosphere. The tricky part is engineering these Rubiscos to take the place of a native carboxysomal Rubisco. Most of the fast Rubiscos discovered are Form II, as opposed to all carboxysomal Rubiscos which are Form I. While a carboxysomal targeting tag would suffice to get some Rubisco encapsulated (45, 46), a full replacement of native Rubisco would likely require engineering of the CsoS2 or CcmM binding site. These binding sites span the large and small subunits, complicating engineering efforts (6, 26). I look forward to seeing what creative solutions will overcome these problems.

4.2 Final Thoughts

Reflecting on the question of why one should study carboxysomes, the broad answer is three-fold. One, for their applications in biotechnology as cellular tools to increase carbon fixation. They are modular, genetically encoded compartments with the potential to increase the efficiency of carbon capture, or introduce it *de novo*, into a wide variety of organisms. When a family member, friend, or colleague asks why I do what I do, I often gravitate towards this answer. Usually what they're actually asking (and sometimes they say this directly) is, what is the application? This answer has a clear purpose and noble end goal. This answer bothers me, though. As Michael Anderson notes in a summary of Santiago Ramón y Cajal's *Advice for a Young Investigator*, "The call to specify in advance the use of an investigation represents a lack of faith in the scientific enterprise, and of necessity limits creativity… Good science makes for good application, not the other way around" (47).

So the second, and truer, reason to study carboxysomes is simply for the pure curiosity of understanding how they work. Pursuing questions out of curiosity is one of the greatest luxuries of graduate school. You can't choose what the data will tell you, or the outcome of an experiment. Forcing biology to work a certain way, without deeply understanding it first, often fails. I entered into the work presented here with more of an engineering mindset, and exited with more of a curiosity mindset. I think to be a scientist you ultimately need a mix of both, but that curiosity is like a secret ingredient; you may not know it's there, but everything just feels more exciting for some inexplicable reason.

And thirdly, none of this work is actually about carboxysomes. It's about the broader important questions for which the humble carboxysome is a model system. How do organisms concentrate CO_2 and increase the efficiency of carbon fixation? How does a protein sequence direct the replicable self-assembly of thousands of other proteins? How does disorder lead to order? These are the grand and powerful questions driving this research.

4.3 References

1. Shively, J. M., Ball, F., Brown, D. H., and Saunders, R. E. (1973) Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of Thiobacillus neapolitanus. *Science*. **182**, 584–586
2. Baker, S. H., Williams, D. S., Aldrich, H. C., Gambrell, A. C., and Shively, J. M. (2000) Identification and localization of the carboxysome peptide Csos3 and its corresponding gene in Thiobacillus neapolitanus. *Arch. Microbiol.* **173**, 278–283
3. Tsai, Y., Sawaya, M. R., Cannon, G. C., Cai, F., Williams, E. B., Heinhorst, S., Kerfeld, C. A., and Yeates, T. O. (2007) Structural analysis of CsoS1A and the protein shell of the Halothiobacillus neapolitanus carboxysome. *PLoS Biol.* **5**, e144
4. Cai, F., Menon, B. B., Cannon, G. C., Curry, K. J., Shively, J. M., and Heinhorst, S. (2009) The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO2 leakage barrier. *PLoS One*. **4**, e7521
5. Cai, F., Dou, Z., Bernstein, S. L., Leverenz, R., Williams, E. B., Heinhorst, S., Shively, J., Cannon, G. C., and Kerfeld, C. A. (2015) Advances in understanding carboxysome assembly in prochlorococcus and synechococcus implicate csos2 as a critical component. *Life*. **5**, 1141–1171
6. Oltrogge, L. M., Chaijarasphong, T., Chen, A. W., Bolin, E. R., Marqusee, S., and Savage, D. F. (2020) Multivalent interactions between CsoS2 and Rubisco mediate \alpha-carboxysome formation. *Nat. Struct. Mol. Biol.* **27**, 281–287
7. Melnicki, M. R., Sutter, M., and Kerfeld, C. A. (2021) Evolutionary relationships among shell proteins of carboxysomes and metabolosomes. *Curr. Opin. Microbiol.* **63**, 1–9
8. Blikstad, C., Dugan, E. J., Laughlin, T. G., Turnšek, J. B., Liu, M. D., Shoemaker, S. R., Vogiatzi, N., Remis, J. P., and Savage, D. F. (2023) Identification of a carbonic anhydrase–Rubisco complex within the alpha-carboxysome. *Proceedings of the National Academy of Sciences*. **120**, e2308600120
9. Bonacci, W., Teng, P. K., Afonso, B., Niederholtmeyer, H., Grob, P., Silver, P. A., and Savage, D. F. (2012) Modularity of a carbon-fixing protein organelle. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 478–483
10. Flamholz, A. I., Dugan, E., Blikstad, C., Gleizer, S., Ben-Nissan, R., Amram, S., Antonovsky, N., Ravishankar, S., Noor, E., Bar-Even, A., Milo, R., and Savage, D. F. (2020) Functional reconstitution of a bacterial CO2 concentrating mechanism in Escherichia coli. *Elife*. 10.7554/eLife.59882
11. Ni, T., Sun, Y., Burn, W., Al-Hazeem, M. M. J., Zhu, Y., Yu, X., Liu, L.-N., and Zhang, P. (2022) Structure and assembly of cargo Rubisco in two native α-carboxysomes. *Nat. Commun.* **13**, 4299
12. Metskas, L. A., Ortega, D., Oltrogge, L. M., Blikstad, C., Lovejoy, D. R., Laughlin, T. G., Savage, D. F., and Jensen, G. J. (2022) Rubisco forms a lattice inside alpha-carboxysomes. *Nat. Commun.* 10.1038/s41467-022-32584-7
13. Evans, S. L., Al-Hazeem, M. M. J., Mann, D., Smetacek, N., Beavil, A. J., Sun, Y., Chen, T., Dykes, G. F., Liu, L.-N., and Bergeron, J. R. C. (2023) Single-particle

cryo-EM analysis of the shell architecture and internal organization of an intact α-carboxysome. *Structure*. **31**, 677-688.e4

14. Ni, T., Jiang, Q., Ng, P. C., Shen, J., Dou, H., Zhu, Y., Radecke, J., Dykes, G. F., Huang, F., Liu, L.-N., and Zhang, P. (2023) Intrinsically disordered CsoS2 acts as a general molecular thread for α-carboxysome shell assembly. *Nat. Commun.* **14**, 5512
15. Tsai, Y.-C. C., Liew, L., Guo, Z., Liu, D., and Mueller-Cajar, O. (2021) The CbbQO-type Rubisco activases encoded in carboxysome gene clusters can activate carboxysomal Form IA Rubiscos. *J. Biol. Chem.*
16. Chen, T., Fang, Y., Jiang, Q., Dykes, G. F., Lin, Y., Price, G. D., Long, B. M., and Liu, L.-N. (2021) Incorporation of Functional Rubisco Activases into Engineered Carboxysomes to Enhance Carbon Fixation. *ACS synthetic biology [electronic resource]*. 10.1021/acssynbio.1c00311
17. Price, G. D., and Badger, M. R. (1989) Expression of Human Carbonic Anhydrase in the Cyanobacterium Synechococcus PCC7942 Creates a High CO(2)-Requiring Phenotype : Evidence for a Central Role for Carboxysomes in the CO(2) Concentrating Mechanism. *Plant Physiol.* **91**, 505–513
18. Mangan, N. M., Flamholz, A., Hood, R. D., Milo, R., and Savage, D. F. (2016) pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5354-62
19. Tanaka, S., Sawaya, M. R., Phillips, M., and Yeates, T. O. (2009) Insights from multiple structures of the shell proteins from the beta-carboxysome. *Protein Sci.* **18**, 108–120
20. Mahinthichaichan, P., Morris, D. M., Wang, Y., Jensen, G. J., and Tajkhorshid, E. (2018) Selective permeability of carboxysome shell pores to anionic molecules. *J. Phys. Chem. B*. **122**, 9110–9118
21. Badger, M. R., Bassett, M., and Comins, H. N. (1985) A model for HCO(3) accumulation and photosynthesis in the cyanobacterium synechococcus sp: theoretical predictions and experimental observations. *Plant Physiol.* **77**, 465–471
22. Reinhold, L., Kosloff, R., and Kaplan, A. (1991) A model for inorganic carbon fluxes and photosynthesis in cyanobacterial carboxysomes. *Can. J. Bot.* **69**, 984–988
23. Rae, B. D., Long, B. M., Badger, M. R., and Price, G. D. (2013) Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO2 fixation in cyanobacteria and some proteobacteria. *Microbiol. Mol. Biol. Rev.* **77**, 357–379
24. Shimakawa, G., Demulder, M., Flori, S., Kawamoto, A., Tsuji, Y., Nawaly, H., Tanaka, A., Tohda, R., Ota, T., Matsui, H., Morishima, N., Okubo, R., Wietrzynski, W., Lamm, L., Righetto, R. D., Uwizeye, C., Gallet, B., Jouneau, P.-H., Gerle, C., Kurisu, G., Finazzi, G., Engel, B. D., and Matsuda, Y. (2023) Diatom pyrenoids are encased in a protein shell that enables efficient CO2 fixation. *bioRxiv*. 10.1101/2023.10.25.564039
25. Nam, O., McKenzie, C., Dowle, A., Dowson, M., Barrett, J., and Mackinder, L. C. M. (2023) A Protein Blueprint of the Diatom CO2-Fixing Organelle. *bioRxiv*. 10.1101/2023.10.26.564148

26. Wang, H., Yan, X., Aigner, H., Bracher, A., Nguyen, N. D., Hee, W. Y., Long, B. M., Price, G. D., Hartl, F. U., and Hayer-Hartl, M. (2019) Rubisco condensate formation by CcmM in \beta-carboxysome biogenesis. *Nature*. **566**, 131–135
27. Chen, A. H., Robinson-Mosher, A., Savage, D. F., Silver, P. A., and Polka, J. K. (2013) The bacterial carbon-fixing organelle is formed by shell envelopment of preassembled cargo. *PLoS One*. **8**, e76127
28. Dai, Y., Chamberlayne, C. F., Messina, M. S., Chang, C. J., Zare, R. N., You, L., and Chilkoti, A. (2023) Interface of biomolecular condensates modulates redox reactions. *Chem*. **9**, 1594–1609
29. Price, G. D., Pengelly, J. J. L., Forster, B., Du, J., Whitney, S. M., von Caemmerer, S., Badger, M. R., Howitt, S. M., and Evans, J. R. (2013) The cyanobacterial CCM as a source of genes for improving photosynthetic CO2 fixation in crop species. *J. Exp. Bot.* **64**, 753–768
30. Zarzycki, J., Axen, S. D., Kinney, J. N., and Kerfeld, C. A. (2013) Cyanobacterial-based approaches to improving photosynthesis in plants. *J. Exp. Bot.* **64**, 787–798
31. Hanson, M. R., Lin, M. T., Carmo-Silva, A. E., and Parry, M. A. J. (2016) Towards engineering carboxysomes into C3 plants. *Plant J.* **87**, 38–50
32. Rae, B. D., Long, B. M., Förster, B., Nguyen, N. D., Velanis, C. N., Atkinson, N., Hee, W. Y., Mukherjee, B., Price, G. D., and McCormick, A. J. (2017) Progress and challenges of engineering a biophysical CO2-concentrating mechanism into higher plants. *J. Exp. Bot.* **68**, 3717–3737
33. Lin, M. T., Occhialini, A., Andralojc, P. J., Parry, M. A. J., and Hanson, M. R. (2014) A faster Rubisco with potential to increase photosynthesis in crops. *Nature*. **513**, 547–550
34. Lin, M. T., Occhialini, A., Andralojc, P. J., Devonshire, J., Hines, K. M., Parry, M. A. J., and Hanson, M. R. (2014) \beta-Carboxysomal proteins assemble into highly organized structures in Nicotiana chloroplasts. *Plant J.* **79**, 1–12
35. Occhialini, A., Lin, M. T., Andralojc, P. J., Hanson, M. R., and Parry, M. A. J. (2016) Transgenic tobacco plants with improved cyanobacterial Rubisco expression but no extra assembly factors grow at near wild-type rates if provided with elevated CO2. *Plant J.* **85**, 148–160
36. Long, B. M., Hee, W. Y., Sharwood, R. E., Rae, B. D., Kaines, S., Lim, Y.-L., Nguyen, N. D., Massey, B., Bala, S., von Caemmerer, S., Badger, M. R., and Price, G. D. (2018) Carboxysome encapsulation of the CO2-fixing enzyme Rubisco in tobacco chloroplasts. *Nat. Commun.* **9**, 3570
37. Orr, D. J., Worrall, D., Lin, M. T., Carmo-Silva, E., Hanson, M. R., and Parry, M. A. J. (2019) Hybrid cyanobacterial-tobacco Rubisco supports autotrophic growth and pre-carboxysomal aggregation. *Plant Physiol.* 10.1104/pp.19.01193
38. Chen, T., Hojka, M., Davey, P., Sun, Y., Dykes, G. F., Zhou, F., Lawson, T., Nixon, P. J., Lin, Y., and Liu, L.-N. (2023) Engineering α-carboxysomes into plant chloroplasts to support autotrophic photosynthesis. *Nat. Commun.* **14**, 2118
39. Cai, F., Heinhorst, S., Shively, J. M., and Cannon, G. C. (2008) Transcript analysis of the Halothiobacillus neapolitanus cso operon. *Arch. Microbiol.* **189**, 141–150

40. Hill, N., Tay, J., Altus, S., Bortz, D., and Cameron, J. (2020) Life cycle of a cyanobacterial carboxysome. *Science Advances*. [online] https://advances.sciencemag.org/content/6/19/eaba1269/tab-pdf (Accessed May 6, 2020)
41. Savage, D. F., Afonso, B., Chen, A. H., and Silver, P. A. (2010) Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science*. **327**, 1258–1261
42. Desmarais, J. J., Flamholz, A. I., Blikstad, C., Dugan, E. J., Laughlin, T. G., Oltrogge, L. M., Chen, A. W., Wetmore, K., Diamond, S., Wang, J. Y., and Savage, D. F. (2019) DABs are inorganic carbon pumps found throughout prokaryotic phyla. *Nature Microbiology*. **4**, 2204–2215
43. Flamholz, A. I., Prywes, N., Moran, U., Davidi, D., Bar-On, Y. M., Oltrogge, L. M., Alves, R., Savage, D., and Milo, R. (2019) Revisiting Trade-offs between Rubisco Kinetic Parameters. *Biochemistry*. **58**, 3365–3376
44. Davidi, D., Shamshoum, M., Guo, Z., Bar-On, Y. M., Prywes, N., Oz, A., Jablonska, J., Flamholz, A., Wernick, D. G., Antonovsky, N., de Pins, B., Shachar, L., Hochhauser, D., Peleg, Y., Albeck, S., Sharon, I., Mueller-Cajar, O., and Milo, R. (2020) Highly active rubiscos discovered by systematic interrogation of natural sequence diversity. *EMBO J.* **39**, e104081
45. Tan, Y. Q., Ali, S., Xue, B., Teo, W. Z., Ling, L. H., Go, M. K., Lv, H., Robinson, R. C., Narita, A., and Yew, W. S. (2021) Structure of a Minimal \alpha-Carboxysome-Derived Shell and Its Utility in Enzyme Stabilization. *Biomacromolecules*. **22**, 4095–4109
46. Li, T., Jiang, Q., Huang, J., Aitchison, C. M., Huang, F., Yang, M., Dykes, G. F., He, H.-L., Wang, Q., Sprick, R. S., Cooper, A. I., and Liu, L.-N. (2020) Reprogramming bacterial protein organelles as a nanoreactor for hydrogen production. *Nat. Commun.* **11**, 5448
47. Anderson, M. (2016) Santiago Ramón y Cajal's Advice for a Young Investigator. *Cerebrum*

Appendix

Contents

A.1 Chapter 2 Supplemental Figures & Tables

Supplementary Figures

Supplementary Tables

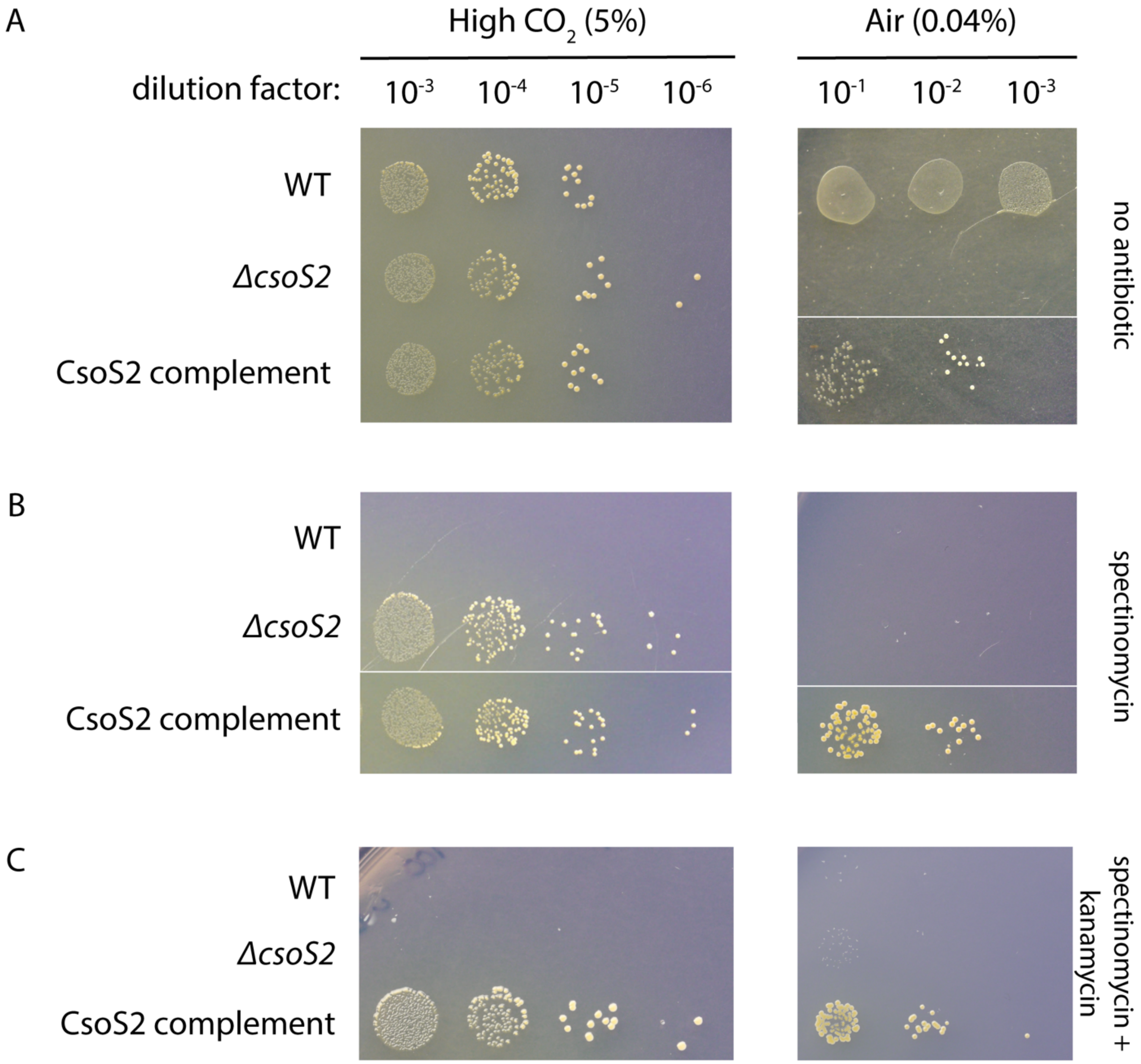


Figure S1. CsoS2 knockouts in *H. neapolitanus* are partially complemented in air. The *ΔcsoS2* strain is a knockout of CsoS2 via spectinomycin ORF insertion. The CsoS2 complement strain contains a genomic insertion of IPTG-inducible CsoS2 and kanamycin targeted to a neutral site on the genome in the *ΔcsoS2* background strain. WT, *ΔcsoS2*, and CsoS2 complement strains were plated in a serial dilution and grown in high CO_2 (5% CO_2) or air (0.04% CO_2) on DSMZ68-agar plates + 100 µM IPTG with (A) no antibiotic, (B) 10 µg/ml spectinomycin, and (C) 10 µg/ml spectinomycin + 2 µg/ml kanamycin. Some images are a composite from different rows on the same plate that were not immediately next to each other, as marked by a thin white line.

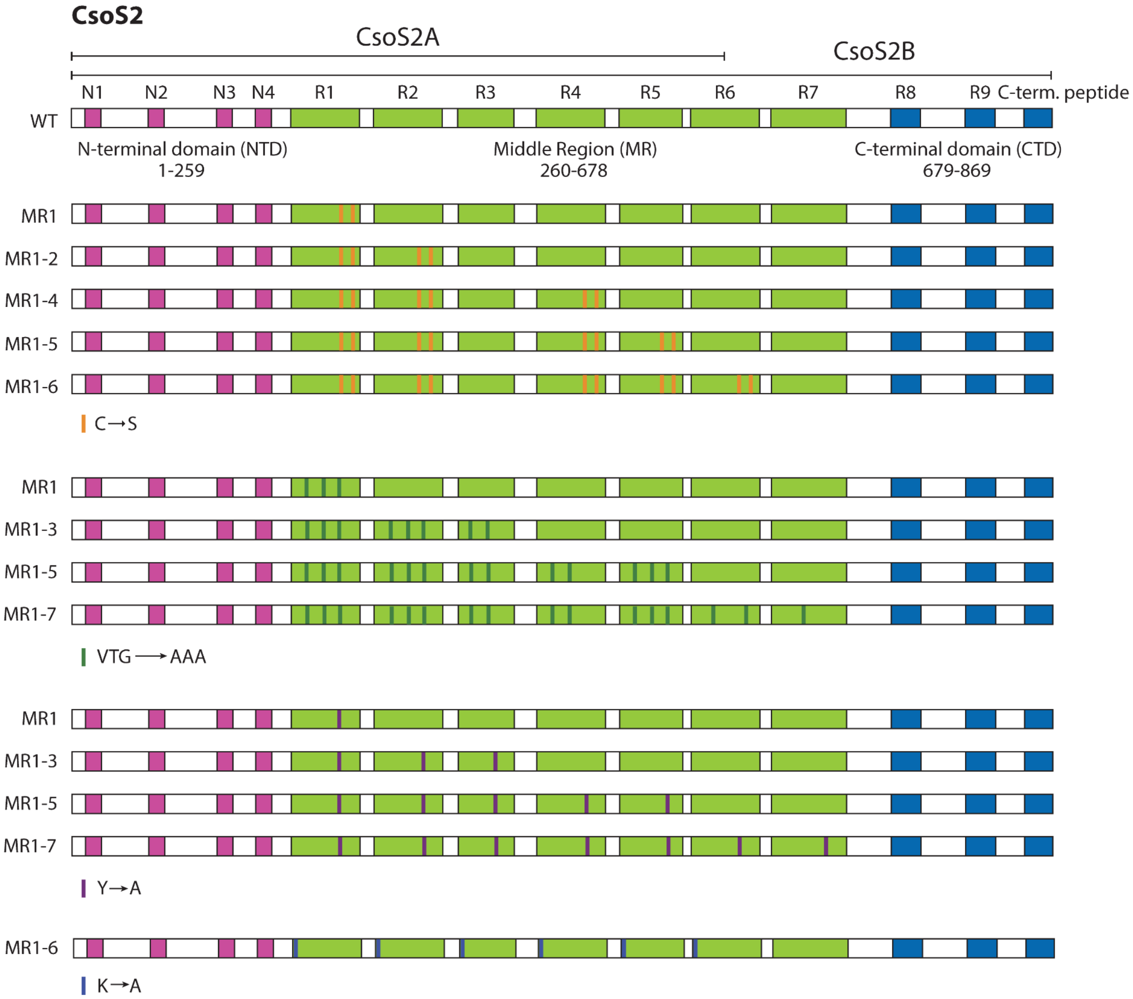
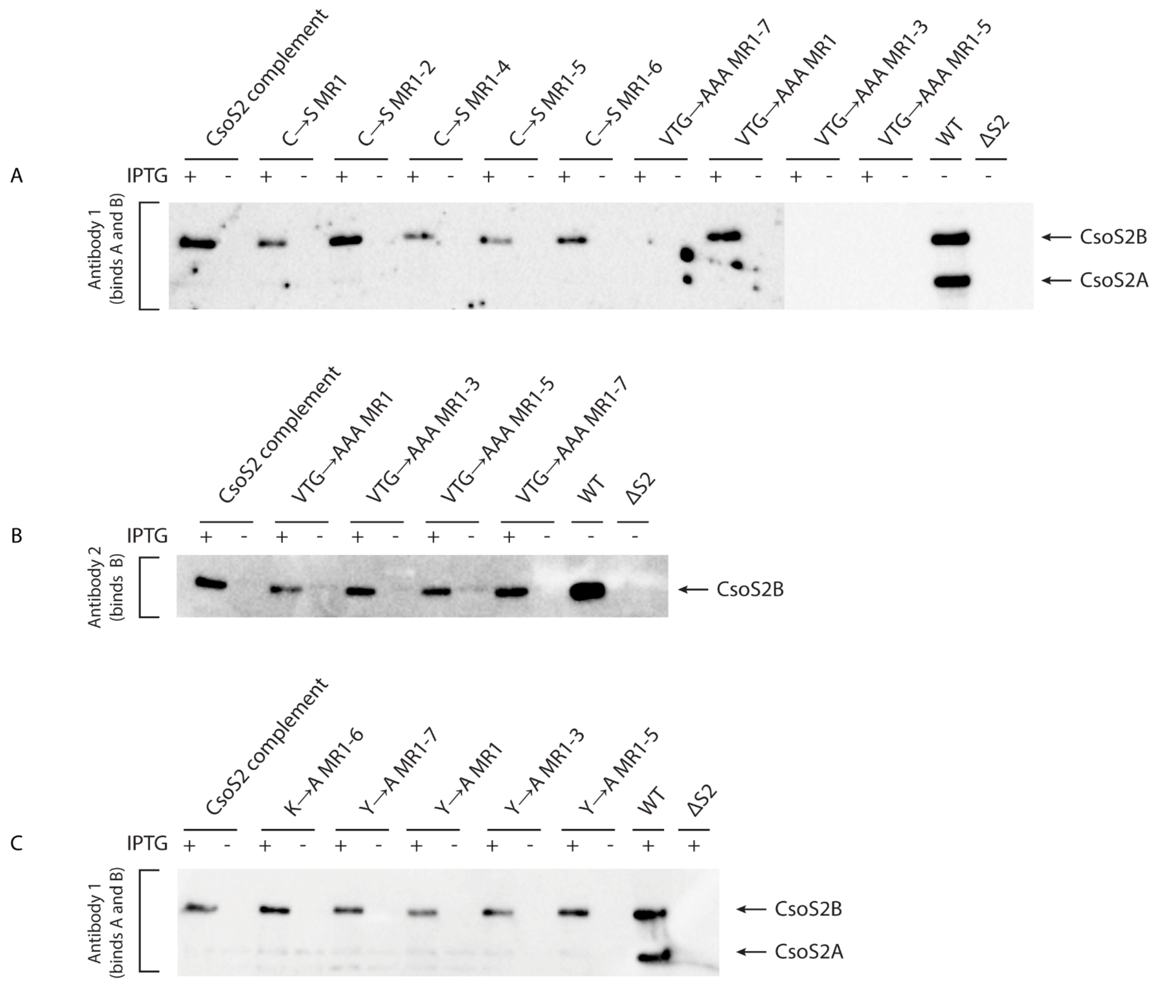


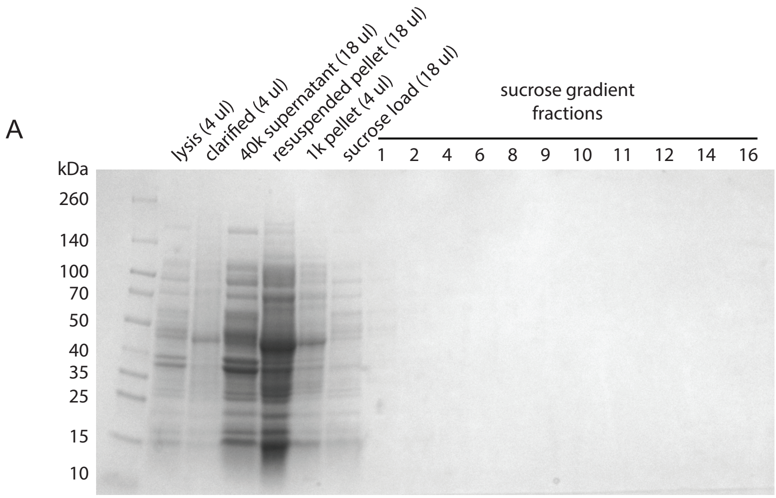
Figure S2. Illustration of all CsoS2 variants expressed *in vivo* in *H. neapolitanus*. Note for the C→S variants, there are no cysteines in R3. Note for the VTG variants, only VTG and VSG sites were mutated.



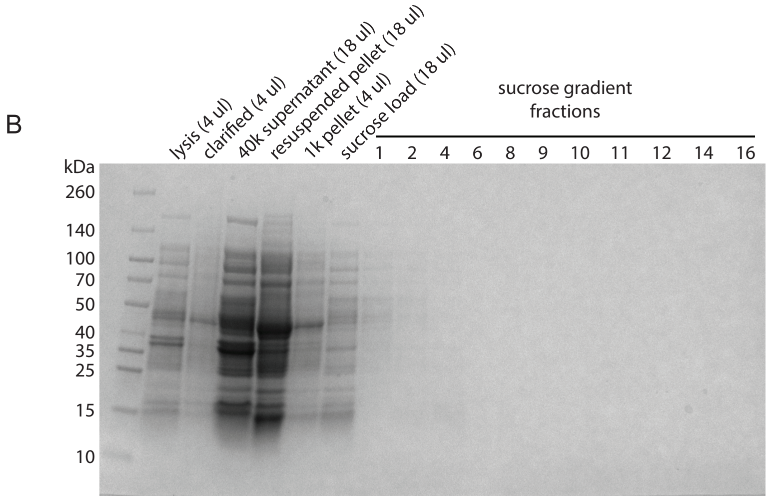
D

Anti-CsoS2 antibodies	Peptide epitope	Reacts with
Antibody 1	CGTEVSRTEKVTGNE	CsoS2A and CsoS2B
Antibody 2	CANRNVPKPEKPGSK	CsoS2B

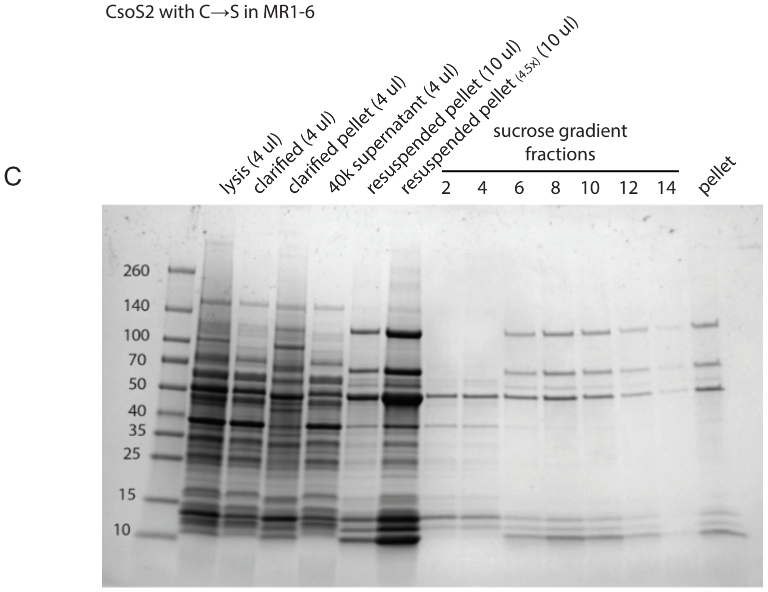
Figure S3. Western blots of complemented CsoS2 and mutant strains. Strains were grown in liquid culture at high CO_2 (5%) +/- 100 µM IPTG and appropriate antibiotics. (A) C→S and VTG→AAA mutants blotted with antibody 1; loading was normalized to ~25 µg/sample. Some antibodies do not bind due to the mutated VTG sequence in the strain. (B) VTG→AAA mutants blotted with antibody 2; loading was normalized to 50 µg/sample. WT was loaded at 12.5 µg to reduce underexposure. (C) K→A and Y→A mutants blotted with antibody 1; loading was normalized to 25 µg/sample. (D) Table of antibodies and their binding epitopes with VTG underlined.



CsoS2 with C→S in MR1



CsoS2 with C→S in MR1-6



WT *H. neapolitanus* carboxysomes purified from *E. coli*

Figure S4. Carboxysomes with C→S mutations cannot be purified from *E. coli*. The indicated CsoS2 C→S mutants were expressed in *E. coli* from a plasmid containing the *H. neapolitanus* carboxysome operon and purified using standard methods (see Experimental Procedures for details). PAGE gel of the carboxysome purification with (A) the CsoS2 variant C→S in MR1, (B) the CsoS2 variant C→S in MR1-6, and (C) wild-type carboxysomes. Carboxysome bands would be expected to appear around fractions 6-12 in (A) and (B). Both CsoS2 mutant preps were attempted a second time with similar results.

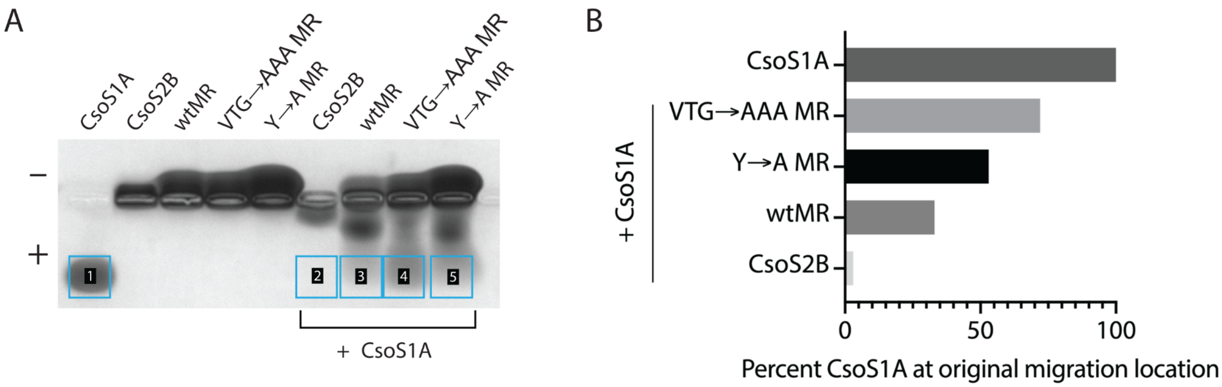


Figure S5. CsoS2B and wtMR bind to shell protein CsoS1A, and mutated MR variants attenuate binding in a native agarose gel. (A) Same agarose gel as in Figure 3B, with boxes drawn around quantified areas. (B) Percent of CsoS1A at the original migration location in (A).

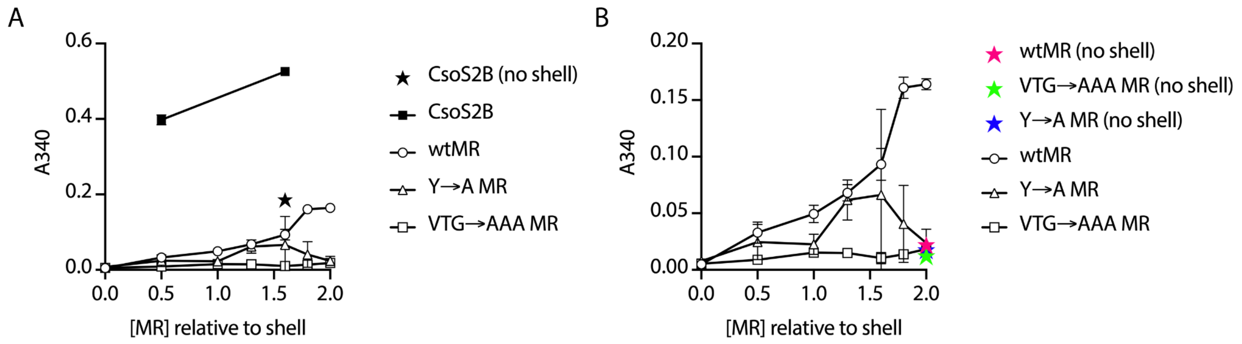


Figure S6. Shell protein CsoS1A binds CsoS2 with high affinity. Turbidity assay at 10 minutes of the indicated constructs with CsoS1A. Data are the same as in Figure 3C, but with CsoS2B shown for comparison (A) and no shell controls (A and B). Unobservable error bars are smaller than the datapoint icon. All samples contained 9 µM CsoS1A (except for the 0 µM shell sample). Concentrations of MR variants were: 0, 4.5, 9, 12, 14, 16, and 18 µM. The 0 µM shell control had 18 µM of MR. For CsoS2, concentrations tested were 4.5 and 14 µM. The 0 µM shell control had 14 µM of CsoS2.

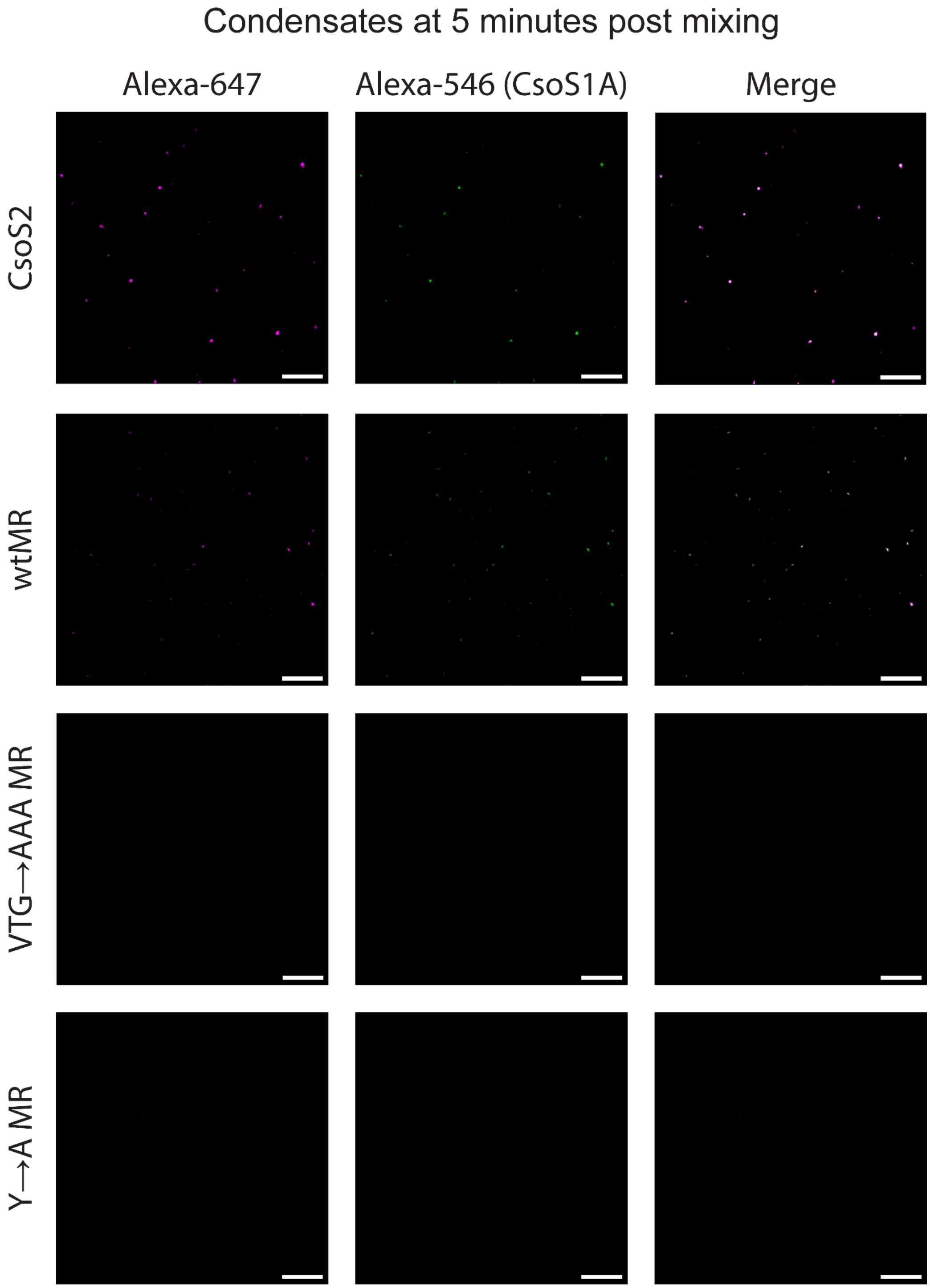


Figure S7. Smaller condensates at 5 minutes after mixing. Fluorescence microscopy of the indicated CsoS2 / MR protein variants with added CsoS1A, imaged at 5 minutes post mixing. All CsoS2 / MR variants are labeled in pink, CsoS1A is labeled in green, and the merge appears white at equally overlapping intensities. All proteins are at a final concentration of 10 µM. Scale bar is 20 µm.

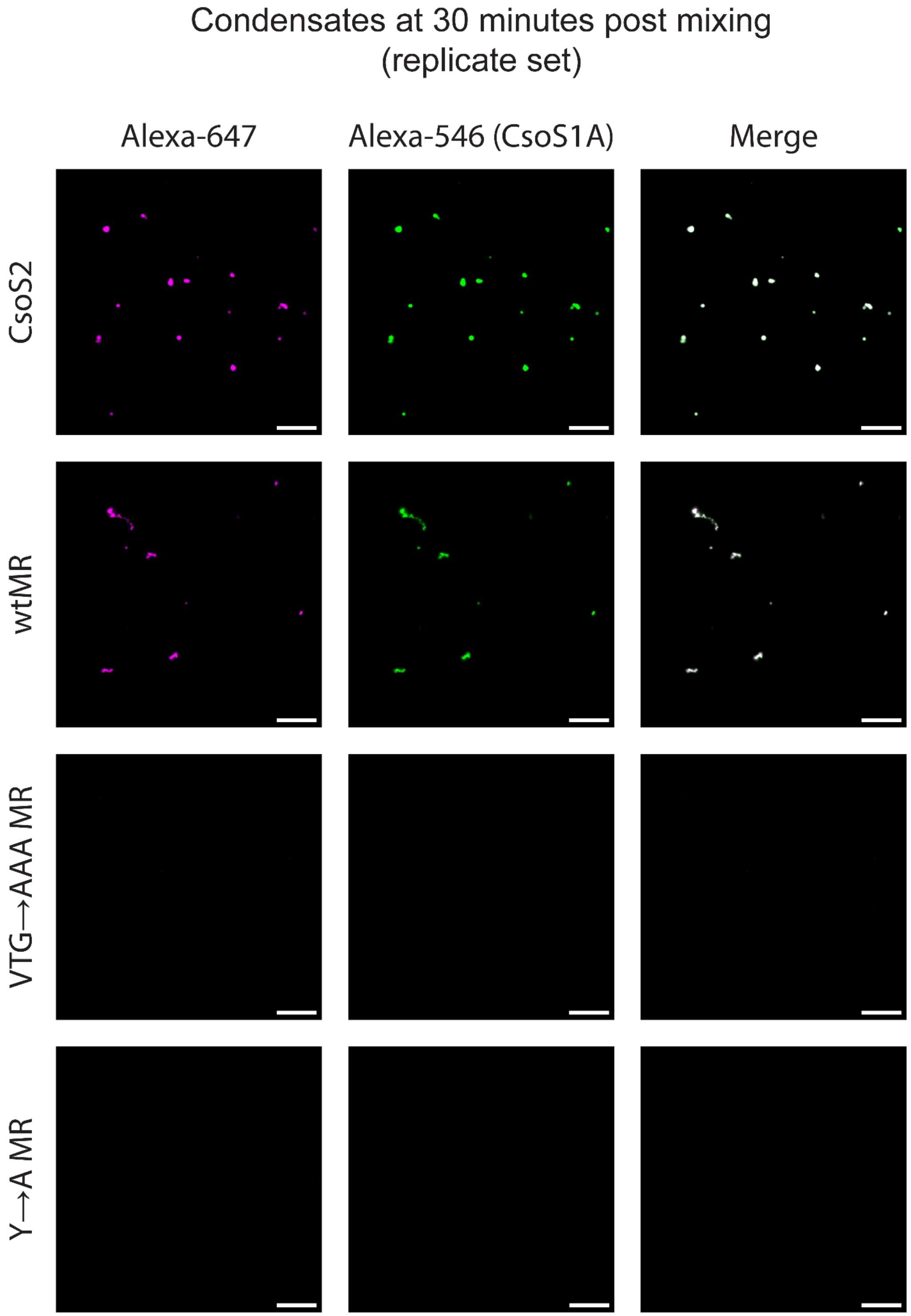


Figure S8. Larger condensates at 30 minutes after mixing (replicate set). Replicate set of micrographs in Figure 4 of the indicated CsoS2 / MR protein variants with added CsoS1A, imaged at 30 minutes post mixing. All CsoS2 / MR variants are labeled in pink, CsoS1A is labeled in green, and the merge appears white at equally overlapping intensities. All proteins are at a final concentration of 10 µM. Scale bar is 20 µm.

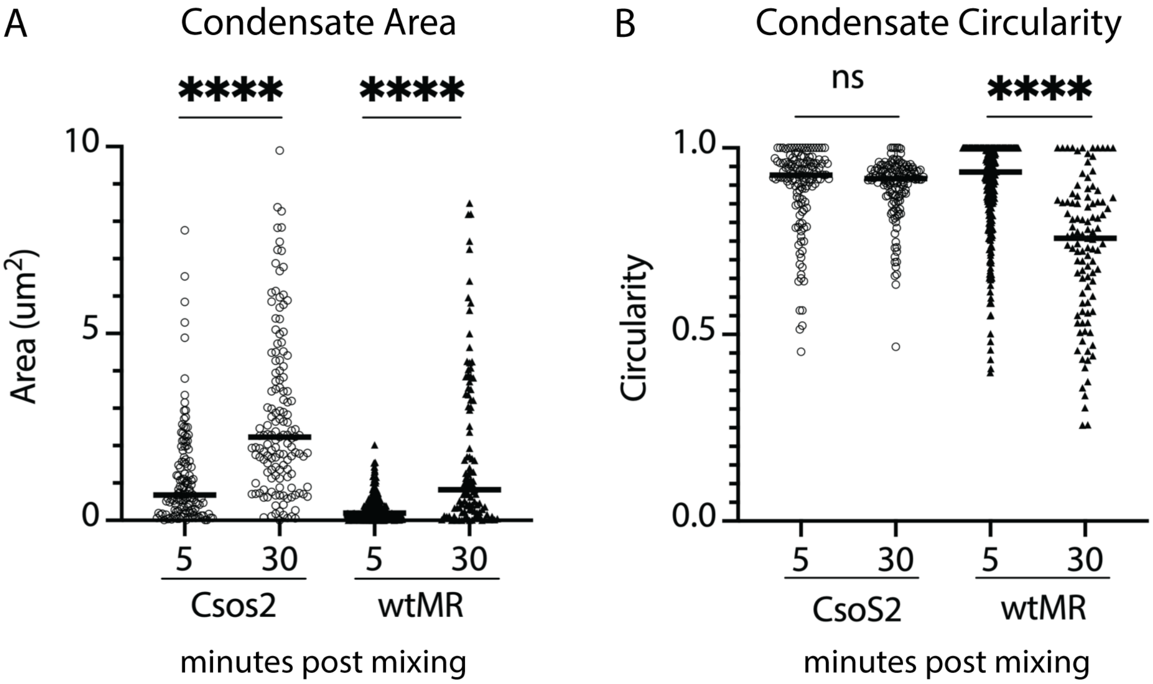
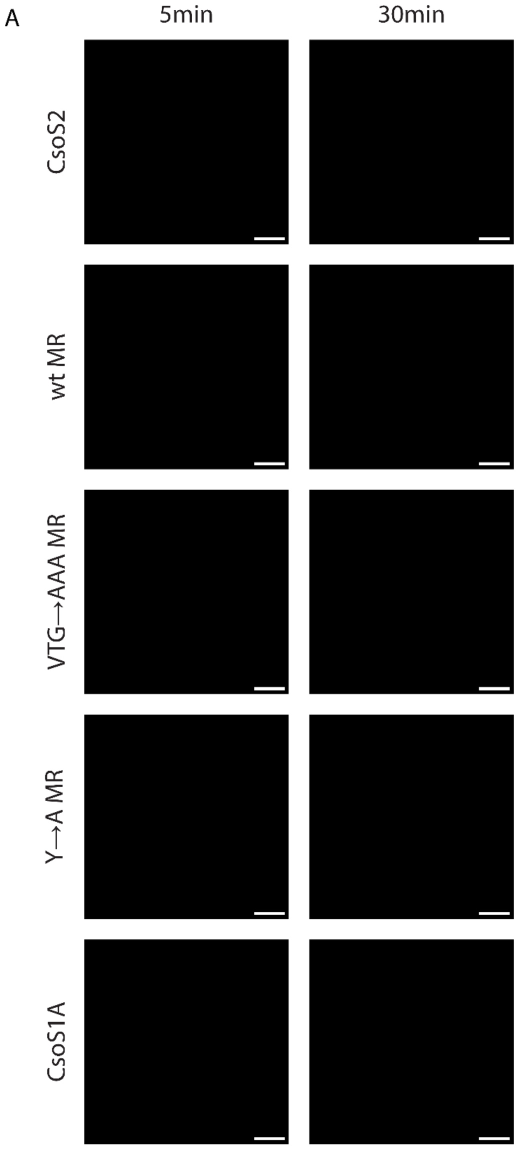


Figure S9. Over 30 minutes condensate area increases for both CsoS2 and wtMR, but condensates develop into different shapes. (A) Condensate area in µm^2 at 5 minutes and 30 minutes post mixing with CsoS1A for both CsoS2 and wtMR. (B) Condensate circularity at 5 minutes and 30 minutes post mixing with CsoS1A for both CsoS2 and wtMR. Circularity is calculated as 4π*area/perimeter^2, with 1.0 being a perfect circle and lower values indicating increasing shape elongation. The median is indicated by a black line. Each individual droplet appears as a dot on the plot. Significance of **** is $P \leq 0.0001$ in an unpaired t-test.



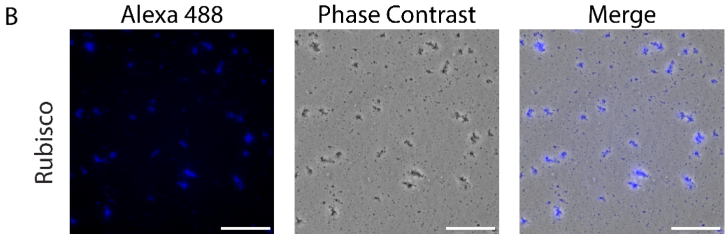


Figure S10. Individual protein controls show no condensate formation. (A) Each indicated protein construct was imaged at 5 minutes and 30 minutes. CsoS2 and MR variants were imaged on the Alexa-647 channel. CsoS1A was imaged on the Alexa-546 channel. All proteins are at a final concentration of 10 µM. All scale bars are 20 µm. (B) Rubisco imaged at 5 minutes on the Alexa-488 channel and phase contrast, and an image merge. Rubisco concentration is 10 µM. Scale bar is 20 µm.

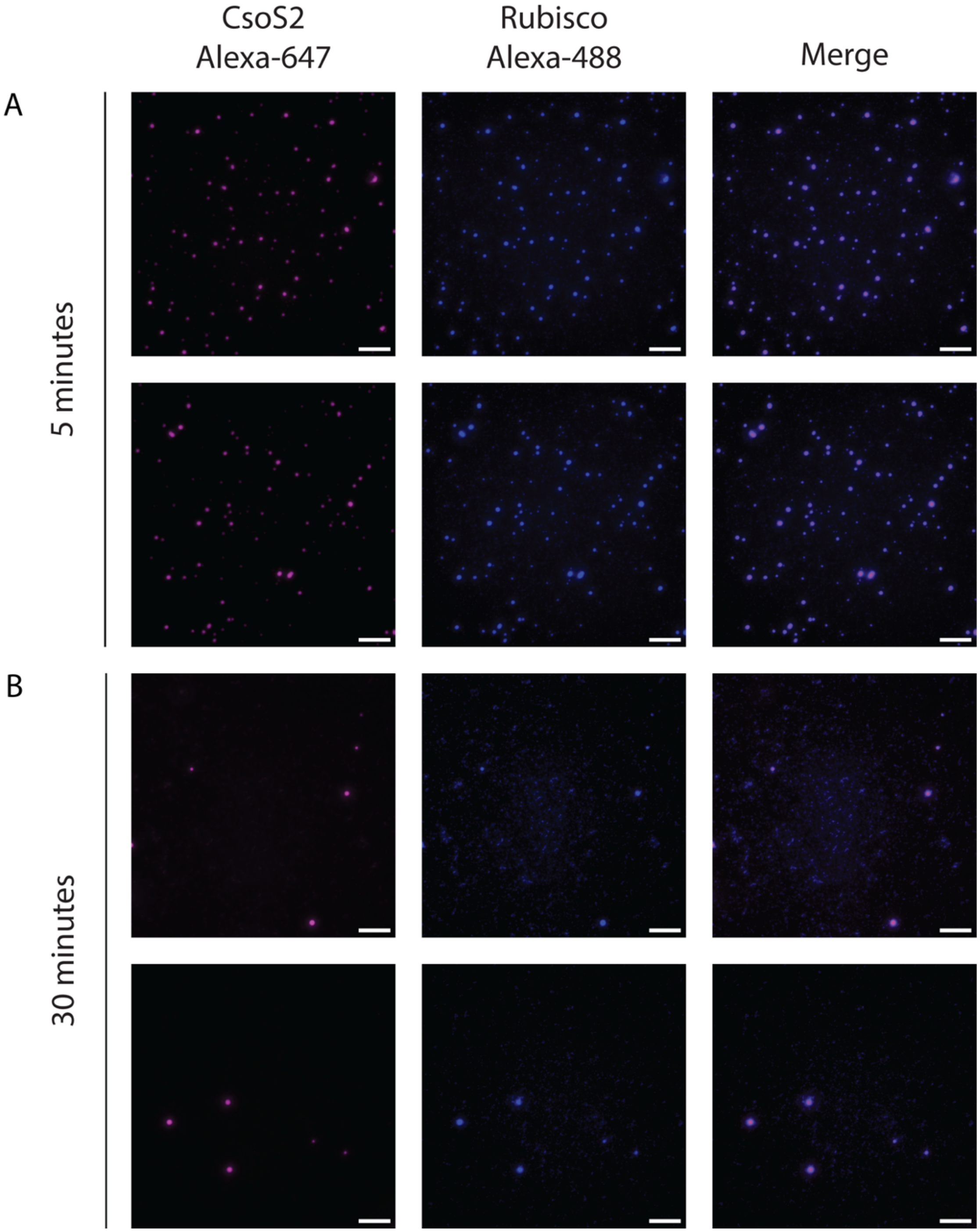


Figure S11. Rubisco and CsoS2 form condensates that dissociate over time. CsoS2 is labeled in pink, Rubisco in blue, and the merge is shown in purple. (A) 5 minutes post mixing, two replicate sets, (B) 30 minutes post mixing, two replicate sets. All proteins are at a final concentration of 10 µM. Scale bar is 10 µm.

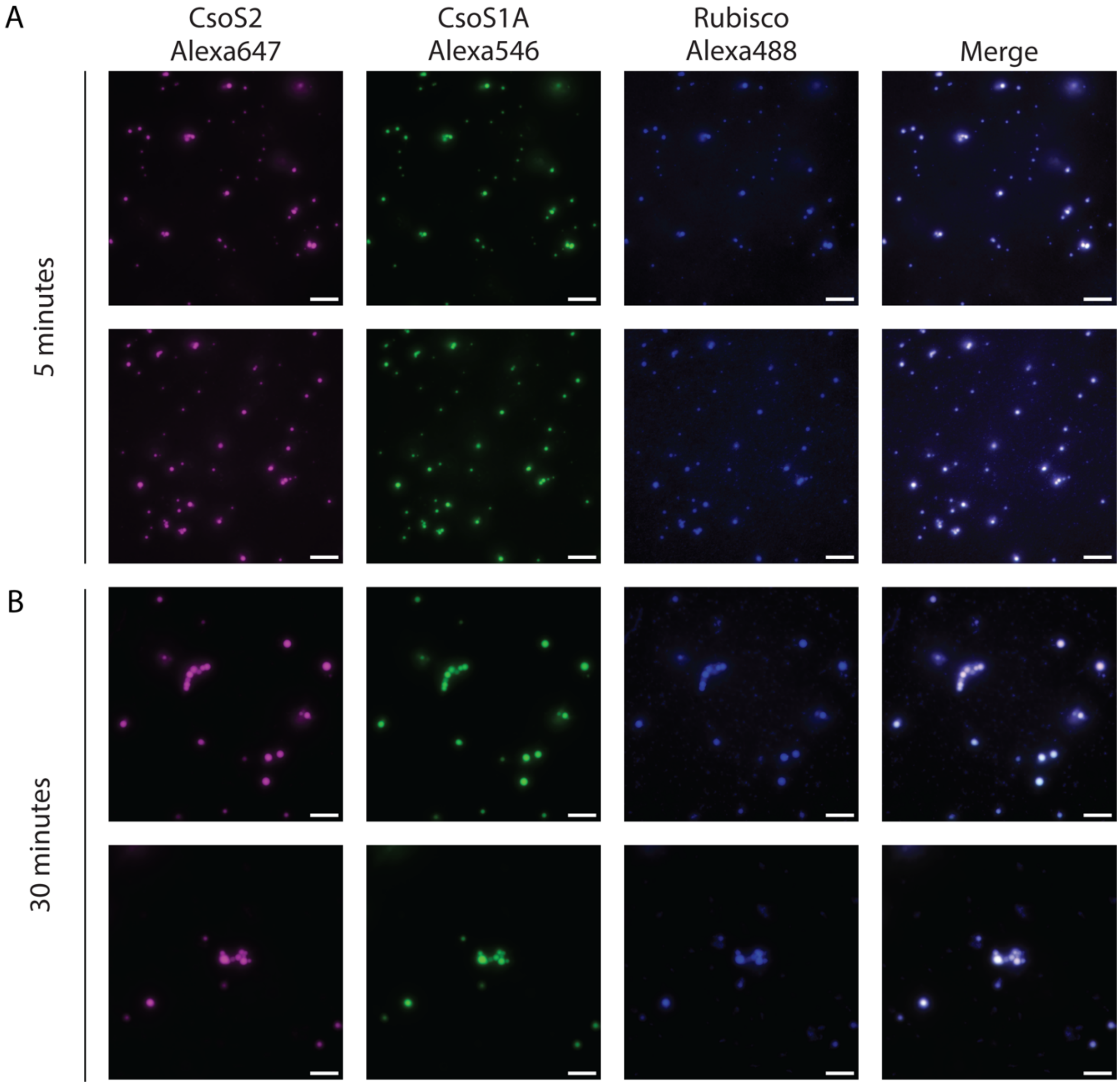


Figure S12. Rubisco, CsoS2, and CsoS1A form robust spherical condensates that grow in size over 30 minutes. CsoS2 is labeled in pink, CsoS1A in green, Rubisco in blue, and the merge appears as white. (A) 5 minutes post mixing, 2 replicate sets. (B) 30 minutes post mixing, 2 replicate sets. Final protein concentrations are 7.9 µM Rubisco, 6.1 µM CsoS2, and 17.5 µM shell. Scale bar is 10 µm.

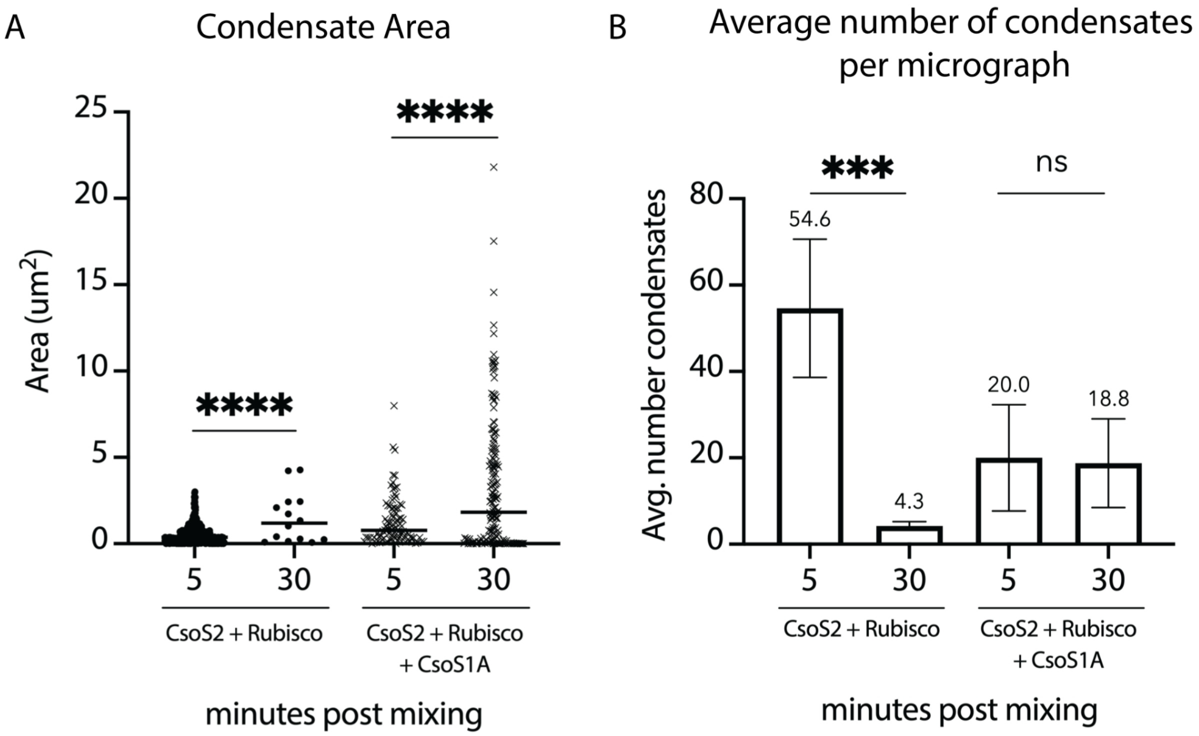


Figure S13. Addition of CsoS1A leads to an increase in droplet size. (A) Average area in µm^2 of CsoS2 + Rubisco condensates and CsoS2 + Rubisco + CsoS1A condensates at 5 minutes and 30 minutes post mixing. The median is indicated by a black line. Each individual droplet appears as a dot on the plot. Significance of **** is P ≤ 0.0001 in an unpaired t-test. (B) Average number of condensates per micrograph, which measured 83.2 x 83.2 µm. The average is written above each bar. Significance of *** is P ≤ 0.001 in an unpaired t-test. Ns, no significance.

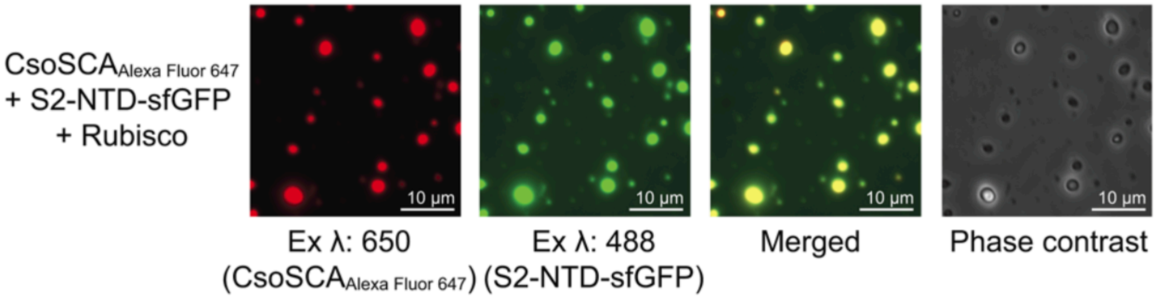


Figure S13.5. CsoSCA recruits into Rubisco–CsoS2 protein condensates. Alexa Fluor 647, sfGFP, and merged fluorescence as well as phase contrast images of protein condensates formed from a solution of Rubisco, CsoS2-NTD-sfGFP, and Alexa Fluor 647 labeled CsoSCA-MBP showing that CsoSCA recruits into Rubisco–CsoS2 protein condensates.

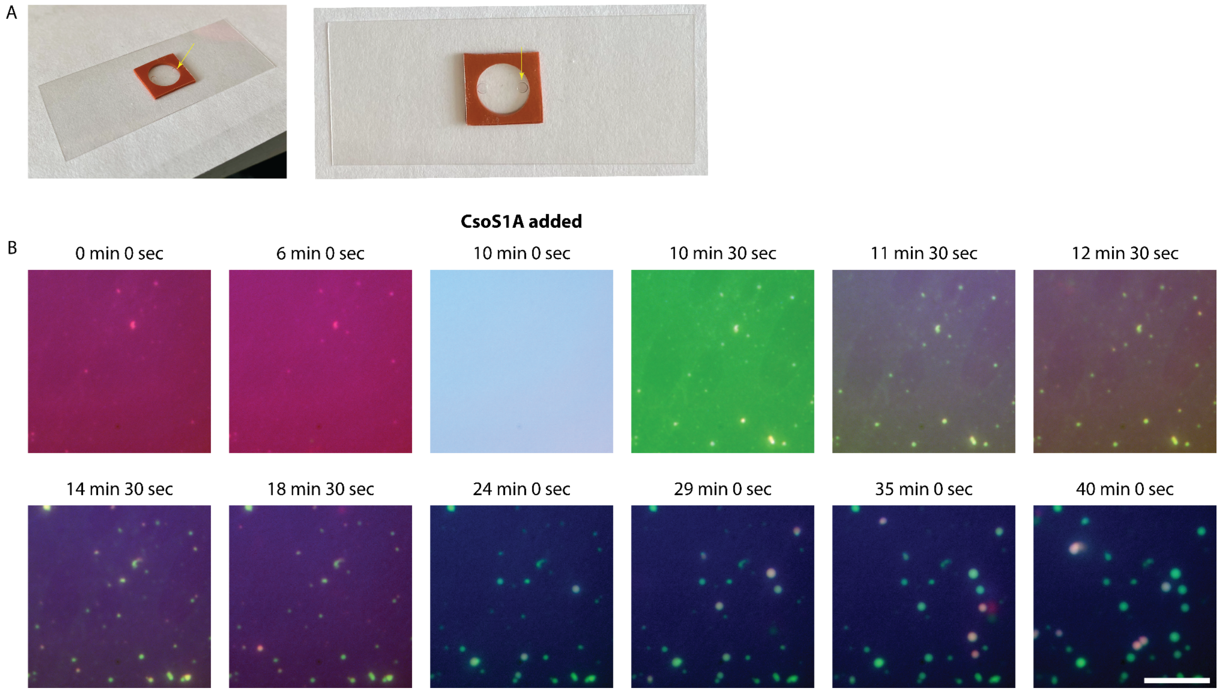


Figure S14. Adding CsoS1A to Rubisco + CsoS2 nucleates condensate formation. (A) Gasket setup; the gasket was affixed to a microscope slide, and protein solution was added with a pipette into the pore indicated by the yellow arrow. (B) 10 µl of 10 µM Rubisco (blue) and 10 µM CsoS2 (red) were premixed and added to a gasket affixed to a microscope slide. At 10 minutes, 2 µl of 105 µM CsoS1A (green) was added, giving final concentrations of 8.3 µM Rubisco, 8.3 µM CsoS2, and 17.5 µM shell in a total of 12 µl. Droplets appear in the focus plane as they condense and adhere to the slide. Times were rounded to the nearest 30 seconds. Scale bar is 10 µm.

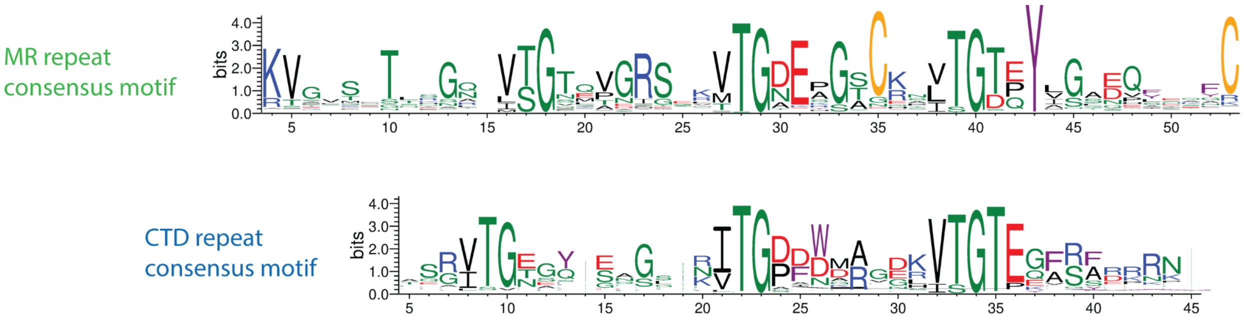


Figure S15. CTD repeat consensus motif. Comparison between the MR repeat consensus motif and the CTD repeat consensus motif. Sequence logo of the MR repeat was generated from an alignment of 1662 MR repeats identified across 272 dereplicated CsoS2 sequences. Sequence logo of the CTD repeat was generated from an alignment of 528 CTD repeats identified across the same 272 dereplicated CsoS2 sequences. Blue is basic, red is acidic, green is polar/small, black is hydrophobic, yellow is cysteine, purple is aromatic.

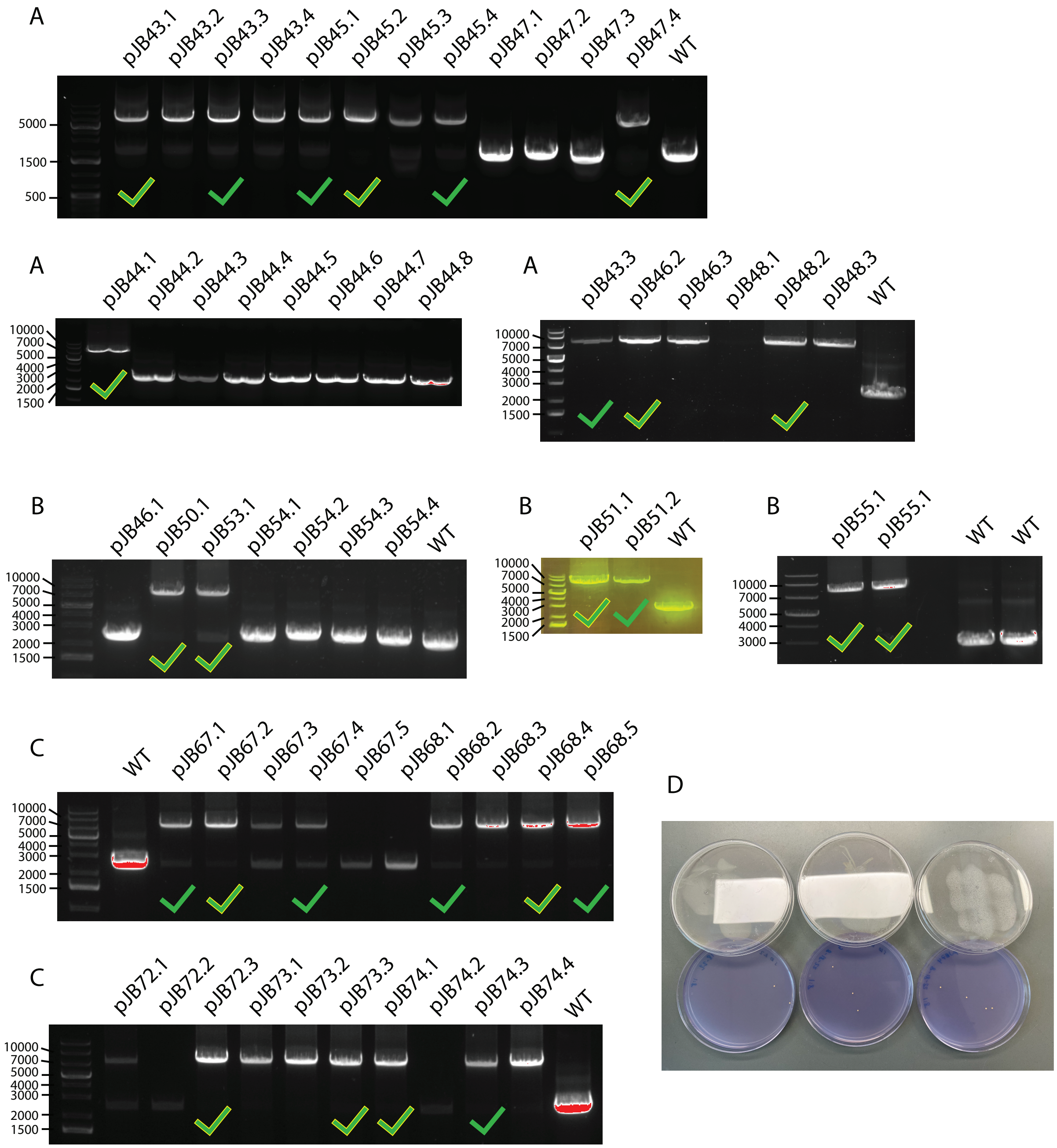


Figure S16. Confirmation of CsoS2 insertion into the *H. neapolitanus* genome. CsoS2 genomic insertion into *H. neapolitanus* was confirmed via colony PCR and sequencing (see Materials & Methods for details). This figure shows the results of colony PCR, where successful insertions appear as a band at 7820 bp, and the WT locus as a band at 2553 bp. Green check marks indicate sequence verified colonies, and green check marks with a yellow border indicate strains that were used in subsequent growth assays. (A) CsoS2 complement (pJB43) and all C →S MR mutants (pJB44, 45, 46, 47, 48). (B) VTG→AAA MR mutants (pJB50, 51, 53, 55). (C) Y →A and K→A MR mutants (pJB67, 68, 72, 73, 74). (D) Example of colonies on plates.

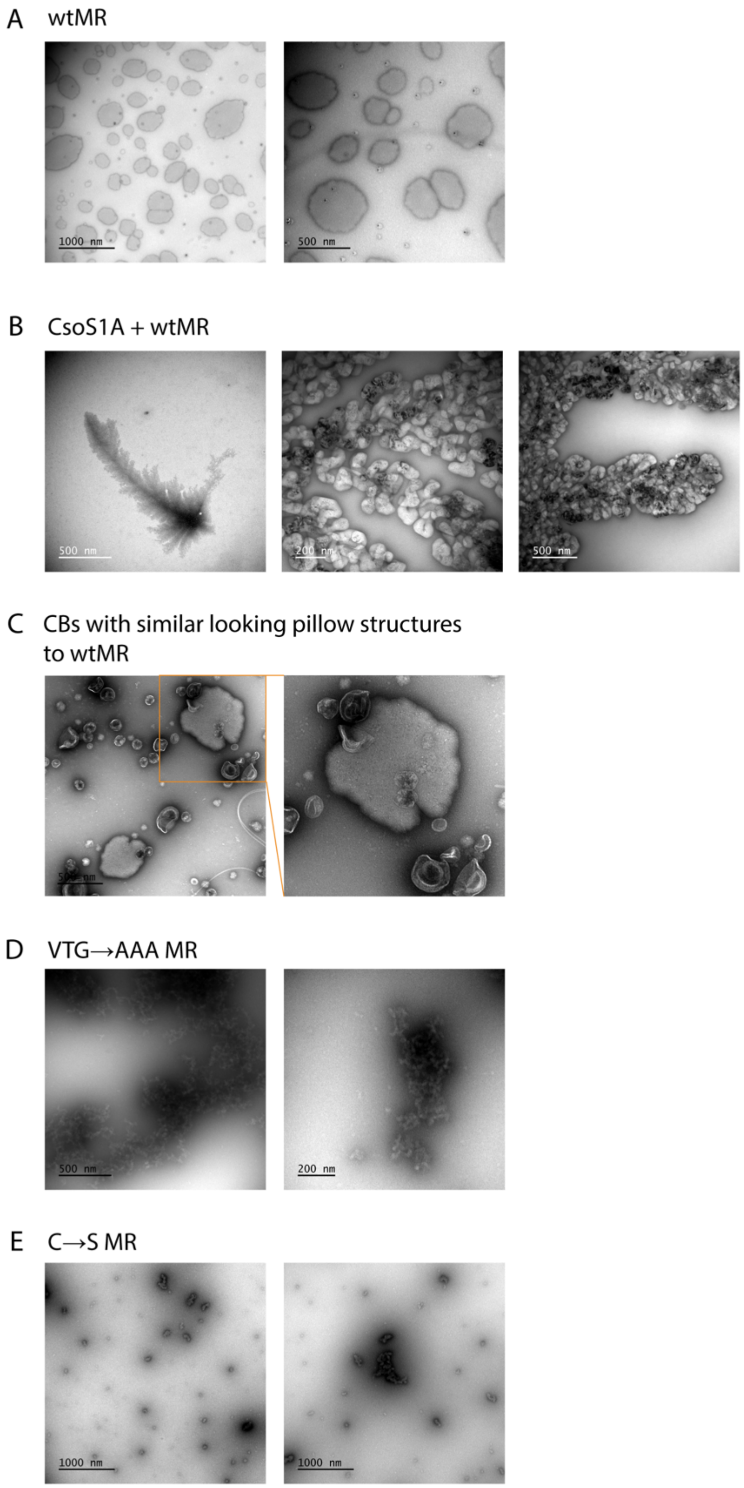


Figure S17. Purified MR variants display diverse morphotypes on TEM. (A) wtMR, (B) CsoS1A + wtMR, (C) *E.c.* carboxysomes embedded in "pillow" structures that show a similarity to wtMR under TEM, (D) VTG→AAA MR, (E) C→S MR. Unfortunately, I did not record the exact concentration used, but I diluted to A280 = 0.03, which was ~5 uM of sample.

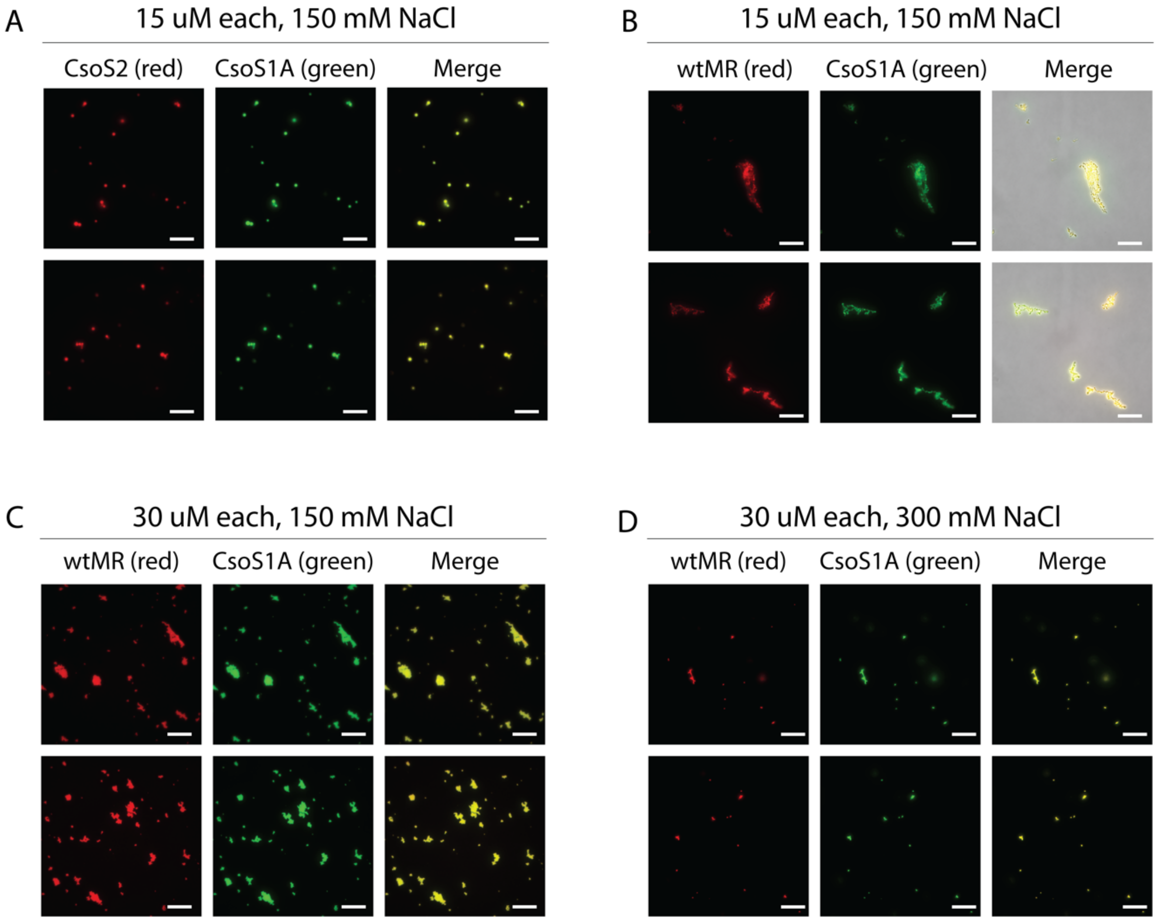


Figure S18. Additional protein and salt concentrations tested in phase separation experiments. All two-component experiments in this thesis, unless otherwise indicated, use 10 uM of each protein in 150 uM NaCl. (A) – (D) tests additional concentrations. While CsoS2 and shell do not grow significantly in size in (A), higher concentrations of wtMR leads to even larger aggregates. A higher salt concentration in (D) breaks up this interaction. Scale bar is 20 um.

Table S1. Strains used in *H. neapolitanus* growth experiments.

***H. neapolitanus* strain**	**Plasmids used to generate strain**	**Genotype**	**Strain Resistance**	**Induction**
N/A	pSAV108	Neutral site destination cloning vector, with the following features from 5'-3': *H. neapolitanus* upstream homology arm, KanR, LacIQ, pTRC promoter, BsaI site, LacZalpha, BsaI site, TrrnB terminator, *H. neapolitanus* downstream homology arm	kanamycin	N/A
WT	N/A	Halothiobacillus neapolitanus c2, ATCC 23641	N/A	N/A
ΔCsoS2	Ref. 23	Spectinomycin resistance cassette inserted into the native CsoS2 locus	spectinomycin	N/A
WT complement	pJB43	WT CsoS2, frameshift site intact	spectinomycin + kanamycin	IPTG
C→S MR1	pJB44	CsoS2 with C→S MR1 inserted into neutral site	spectinomycin + kanamycin	IPTG
C→S MR1-2	pJB45	CsoS2 with C→S CsoS2 MR1-2 inserted into neutral site	spectinomycin + kanamycin	IPTG
C→S MR1-4	pJB46	CsoS2 with C→S CsoS2 MR1-4 inserted into neutral site	spectinomycin + kanamycin	IPTG
C→S MR1-5	pJB47	CsoS2 with C→S CsoS2 MR1-5 inserted into neutral site	spectinomycin + kanamycin	IPTG
C→S MR1-6	pJB48	CsoS2 with C→S CsoS2 MR1-6 inserted into neutral site	spectinomycin + kanamycin	IPTG
VTG→AAA MR1	pJB51	CsoS2 with V(T/S)G→AAA CsoS2 MR1 inserted into neutral site	spectinomycin + kanamycin	IPTG
VTG→AAA MR1-3	pJB53	CsoS2 with V(T/S)G→AAA CsoS2 MR1-3 inserted into neutral site	spectinomycin + kanamycin	IPTG
VTG→AAA MR1-5	pJB55	CsoS2 with V(T/S)G→AAA CsoS2 MR1-5 inserted into neutral site	spectinomycin + kanamycin	IPTG
VTG→AAA MR1-7	pJB50	CsoS2 with V(T/S)G→AAA CsoS2 MR1-7 inserted into neutral site	spectinomycin + kanamycin	IPTG
Y→A MR1	pJB72	CsoS2 with Y→A CsoS2 MR1 inserted into neutral site	spectinomycin + kanamycin	IPTG
Y→A MR1-3	pJB73	CsoS2 with Y→A CsoS2 MR1-3 inserted into neutral site	spectinomycin + kanamycin	IPTG
Y→A MR1-5	pJB74	CsoS2 with Y→A CsoS2 MR1-5 inserted into neutral site	spectinomycin + kanamycin	IPTG
Y→A MR1-7	pJB68	CsoS2 with Y→A CsoS2 MR1-7 inserted into neutral site	spectinomycin + kanamycin	IPTG
K→A MR1-6	pJB67	CsoS2 with K→A CsoS2 MR1-6 inserted into neutral site	spectinomycin + kanamycin	IPTG

Table S2. Sequences used in this study.

Plasmid	**Description**	**Resistance**	**MR Domain Sequence**
	WT *H. neapolitanus* CsoS2 sequence (for reference)		(full CsoS2 sequence) Pink is NTD, Green is MR, and Blue is CTD MSPSQSGMNPADLSGLSGKELARARRAALSKQGKAAVSNKTASVNRSTKQAASSINTNQVRSSVNEVPTDYQMADQLCSTIDHADFGTESNRVRDLCRQRREALSTIGKKAVKTNGKPSGRVRPQQSVVHNDAMIENAGDTNQSSSTSLNNELSEICSIADDMPERFGSQAKTVRDICRARRQALSERGTRAVPPKPQSQGGPGRNGYQIDGYLDTALHGRDAAKRHREMLCQYGRGTAPSCKPTGRVKNSVQSGNAAPKKVETGHTLSGGSVTGTQVDRKSHVTGNEPGTCRAVTGTEYVGTEQFTSFCNTSPKPNATKVNVTTTARGRPVSGTEVSRTEKVTGNESGVCRNVTGTEYMSNEAHFSLCGTAAKPSQADKVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQYADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSCRSISGTEYLSNEQFQSFCDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSCSRVTGSQYGQSKICGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEYYGREHFEPFCTSTPEPEAQSTEQSLTCEGQIISGTSVDASDLVTGNEIGEQQLISGDAYVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQNFAPGEVMPTDFSIQTPARSAQNRITGNDIAPSGRITGPGMLATGLITGTPEFRHAARELVGSPQPMAMAMANRNKAAQAPVVQPEVVATQEKPELVCAPRSDQMDRVSGEGKERCHITGDDWSVNKHITGTAGQWASGRNPSMRGNARVVETSAFANRNVPKPEKPGSKITGSSGNDTQGSLITYSGGARG*
pJB43	WT CsoS2, frameshift site intact	kanamycin	KVETGHTLSGGSVTGTQVDRKSHVTGNEPGTCRAVTGTEYVGTEQFTSFCNTSPKPNATKVNVTTTARGRPVSGTEVSRTEKVTGNESGVCRNVTGTEYMSNEAHFSLCGTAAKPSQADKVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQYADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSCRSISGTEYLSNEQFQSFCDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSCSRVTGSQYGQSKICGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEYYGREHFEPFCTSTPEPEAQSTEQSLTCEGQIISGTSVDASDLVTGNEIGEQQLISGDAYVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQN
pJB44	CsoS2 with C→S MR1	kanamycin	KVETGHTLSGGSVTGTQVDRKSHVTGNEPGTSRAVTGTEYVGTEQFTSFSNTSPKPNATKVNVTTTARGRPVSGTEVSRTEKVTGNESGVCRNVTGTEYMSNEAHFSLCGTAAKPSQADKVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQYADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSCRSISGTEYLSNEQFQSFCDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSCSRVTGSQYGQSKICGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEYYGREHFEPFCTSTPEPEAQSTEQSLTCEGQIISGTSVDASDLVTGNEIGEQQLISGDAYVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQN
pJB45	CsoS2 with C→S CsoS2 MR1-2	kanamycin	KVETGHTLSGGSVTGTQVDRKSHVTGNEPGTSRAVTGTEYVGTEQFTSFSNTSPKPNATKVNVTTTARGRPVSGTEVSRTEKVTGNESGVSRNVTGTEYMSNEAHFSLSGTAAKPSQADKVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQYADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSCRSISGTEYLSNEQFQSFCDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSCSRVTGSQYGQSKICGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEYYGREHFEPFCTSTPEPEAQSTEQSLTCEGQIISGTSVDASDLVTGNEIGEQQLISGDAYVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQN

pJB46	CsoS2 with C→S CsoS2 MR1-4	kanamycin	KVETGHTLSGGSVTGTQVDRKSHVTGNEPGTSRAVTGTEYVGTEQFTSFSNTSPKPNATKVNVTTTARGRPVSGTEVSRTEKVTGNESGVSRNVTGTEYMSNEAHFSLSGTAAKPSQADKVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQYADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSSRSISGTEYLSNEQFQSFSDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSCSRVTGSQYGQSKICGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEYYGREHFEPFCTSTPEPEAQSTEQSLTCEGQIISGTSVDASDLVTGNEIGEQQLISGDAYVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQN
pJB47	CsoS2 with C→S CsoS2 MR1-5	kanamycin	KVETGHTLSGGSVTGTQVDRKSHVTGNEPGTSRAVTGTEYVGTEQFTSFSNTSPKPNATKVNVTTTARGRPVSGTEVSRTEKVTGNESGVSRNVTGTEYMSNEAHFSLSGTAAKPSQADKVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQYADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSSRSISGTEYLSNEQFQSFSDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSSSRVTGSQYGQSKISGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEYYGREHFEPFCTSTPEPEAQSTEQSLTCEGQIISGTSVDASDLVTGNEIGEQQLISGDAYVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQN
pJB48	CsoS2 with C→S CsoS2 MR1-6	kanamycin	KVETGHTLSGGSVTGTQVDRKSHVTGNEPGTSRAVTGTEYVGTEQFTSFSNTSPKPNATKVNVTTTARGRPVSGTEVSRTEKVTGNESGVSRNVTGTEYMSNEAHFSLSGTAAKPSQADKVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQYADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSSRSISGTEYLSNEQFQSFSDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSSSRVTGSQYGQSKISGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGSMPVTGNEYYGREHFEPFSTSTPEPEAQSTEQSLTCEGQIISGTSVDASDLVTGNEIGEQQLISGDAYVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQN
pJB51	CsoS2 with V(T/S)G→AAA CsoS2 MR1	kanamycin	KVETGHTLSGGSAAATQVDRKSHAAANEPGTCRAAAATEYVGTEQFTSFCNTSPKPNATKVNVTTTARGRPVSGTEVSRTEKVTGNESGVCRNVTGTEYMSNEAHFSLCGTAAKPSQADKVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQYADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSCRSISGTEYLSNEQFQSFCDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSCSRVTGSQYGQSKICGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEYYGREHFEPFCTSTPEPEAQSTEQSLTCEGQIISGTSVDASDLVTGNEIGEQQLISGDAYVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQN
pJB53	CsoS2 with V(T/S)G→AAA CsoS2 MR1-3	kanamycin	KVETGHTLSGGSAAATQVDRKSHAAANEPGTCRAAAATEYVGTEQFTSFCNTSPKPNATKVNVTTTARGRPAAATEVSRTEKAAANESGVCRNAAATEYMSNEAHFSLCGTAAKPSQADKVMFGATARTHQVAAASDEFRPSSAAANESGAKRTITGSQYADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSCRSISGTEYLSNEQFQSFCDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSCSRVTGSQYGQSKICGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEYYGREHFEPFCTSTPEPEAQSTEQSLTCEGQIISGTSVDASDLVTGNEIGEQQLISGDAYVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQN

pJB55	CsoS2 with V(T/S)G→AAA CsoS2 MR1-5	kanamycin	KVETGHTLSGGSAAATQVDRKSHAAANEPGTCRAAAATEYVGTEQFTSFCNTSPKPNATKVNVTTTARGRPAAATEVSRTEKAAANESGVCRNAAATEYMSNEAHFSLCGTAAKPSQADKVMFGATARTHQVAAASDEFRPSSAAANESGAKRTITGSQYADEGLARLTINGAPAKVARTHTFAGSDAAATEIGRSTRAAADESGSCRSISGTEYLSNEQFQSFCDTKPQRSPFKVGQDRTNKGQSAAANLVDRSELAAANEPGSCSRAAASQYGQSKICGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEYYGREHFEPFCTSTPEPEAQSTEQSLTCEGQIISGTSVDASDLVTGNEIGEQQLISGDAYVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQN
pJB50	CsoS2 with V(T/S)G→AAA CsoS2 MR1-7	kanamycin	KVETGHTLSGGSAAATQVDRKSHAAANEPGTCRAAAATEYVGTEQFTSFCNTSPKPNATKVNVTTTARGRPAAATEVSRTEKAAANESGVCRNAAATEYMSNEAHFSLCGTAAKPSQADKVMFGATARTHQVAAASDEFRPSSAAANESGAKRTITGSQYADEGLARLTINGAPAKVARTHTFAGSDAAATEIGRSTRAAADESGSCRSISGTEYLSNEQFQSFCDTKPQRSPFKVGQDRTNKGQSAAANLVDRSELAAANEPGSCSRAAASQYGQSKICGGGVGKVRSMRTLRGTSAAAQQLDHAPKMSGDERGGCMPAAANEYYGREHFEPFCTSTPEPEAQSTEQSLTCEGQIISGTSVDASDLAAANEIGEQQLISGDAYVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQN
pJB72	CsoS2 with Y→A CsoS2 MR1	kanamycin	KVETGHTLSGGSVTGTQVDRKSHVTGNEPGTCRAVTGTEAVGTEQFTSFCNTSPKPNATKVNVTTTARGRPVSGTEVSRTEKVTGNESGVCRNVTGTEYMSNEAHFSLCGTAAKPSQADKVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQYADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSCRSISGTEYLSNEQFQSFCDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSCSRVTGSQYGQSKICGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEYYGREHFEPFCTSTPEPEAQSTEQSLTCEGQIISGTSVDASDLVTGNEIGEQQLISGDAYVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQN
pJB73	CsoS2 with Y→A CsoS2 MR1-3	kanamycin	KVETGHTLSGGSVTGTQVDRKSHVTGNEPGTCRAVTGTEAVGTEQFTSFCNTSPKPNATKVNVTTTARGRPVSGTEVSRTEKVTGNESGVCRNVTGTEAMSNEAHFSLCGTAAKPSQADKVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQAADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSCRSISGTEYLSNEQFQSFCDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSCSRVTGSQYGQSKICGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEYYGREHFEPFCTSTPEPEAQSTEQSLTCEGQIISGTSVDASDLVTGNEIGEQQLISGDAYVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQN
pJB74	CsoS2 with Y→A CsoS2 MR1-5	kanamycin	KVETGHTLSGGSVTGTQVDRKSHVTGNEPGTCRAVTGTEAVGTEQFTSFCNTSPKPNATKVNVTTTARGRPVSGTEVSRTEKVTGNESGVCRNVTGTEAMSNEAHFSLCGTAAKPSQADKVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQAADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSCRSISGTEALSNEQFQSFCDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSCSRVTGSQAGQSKICGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEYYGREHFEPFCTSTPEPEAQSTEQSLTCEGQIISGTSVDASDLVTGNEIGEQQLISGDAYVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQN

pJB68	CsoS2 with Y→A CsoS2 MR1-7	kanamycin	KVETGHTLSGGSVTGTQVDRKSHVTGNEPGTCRAVTGTEAVGTEQFTSFCNTSPKPNATKVNVTTTARGRPVSGTEVSRTEKVTGNESGVCRNVTGTEAMSNEAHFSLCGTAAKPSQADKVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQAADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSCRSISGTEALSNEQFQSFCDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSCSRVTGSQAGQSKICGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEAYGREHFEPFCTSTPEPEAQSTEQSLTCEGQIISGTSVDASDLVTGNEIGEQQLISGDAAVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQN
pJB67	CsoS2 with K→A CsoS2 MR1-6	kanamycin	AVETGHTLSGGSVTGTQVDRKSHVTGNEPGTCRAVTGTEYVGTEQFTSFCNTSPKPNATAVNVTTTARGRPVSGTEVSRTEKVTGNESGVCRNVTGTEYMSNEAHFSLCGTAAKPSQADAVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQYADEGLARLTINGAPAAVARTHTFAGSDVTGTEIGRSTRVTGDESGSCRSISGTEYLSNEQFQSFCDTKPQRSPFAVGQDRTNKGQSVTGNLVDRSELVTGNEPGSCSRVTGSQYGQSKICGGGVGAVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEYYGREHFEPFCTSTPEPEAQSTEQSLTCEGQIISGTSVDASDLVTGNEIGEQQLISGDAYVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQN
pLz92	6xHis-CsoS2B	carbenicillin	MHHHHHHPSQSGMNPADLSGLSGKELARARRAALSKQGKAAVSNKTASVNRSTKQAASSINTNQVRSSVNEVPTDYQMADQLCSTIDHADFGTESNRVRDLCRQRREALSTIGKKAVKTNGKPSGRVRPQQSVVHNDAMIENAGDTNQSSSTSLNNELSEICSIADDMPERFGSQAKTVRDICRARRQALSERGTRAVPPKPQSQGGPGRNGYQIDGYLDTALHGRDAAKRHREMLCQYGRGTAPSCKPTGRVKNSVQSGNAAPKKVETGHTLSGGSVTGTQVDRKSHVTGNEPGTCRAVTGTEYVGTEQFTSFCNTSPKPNATKVNVTTTARGRPVSGTEVSRTEKVTGNESGVCRNVTGTEYMSNEAHFSLCGTAAKPSQADKVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQYADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSCRSISGTEYLSNEQFQSFCDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSCSRVTGSQYGQSKICGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEYYGREHFEPFCTSTPEPEAQSTEQSLTCEGQIISGTSVDASDLVTGNEIGEQQLISGDAYVGAQQTGCLPTSPRFNQTGNVQSMGFKNTNQPEQNFAPGEVMPTDFSIQTPARSAQNRITGNDIAPSGRITGPGMLATGLITGTPEFRHAARELVGSPQPMAMAMANRNKAAQAPVVQPEVVATQEKPELVCAPRSDQMDRVSGEGKERCHITGDDWSVNKHITGTAGQWASGRNPSMRGNARVVETSAFANRNVPKPEKPGSKITGSSGNDTQGSLITYSGGARG*
pBz84	6xHis-(wtMR1-6)-strep	carbenicillin	MSHHHHHHAPKKVETGHTLSGGSVTGTQVDRKSHVTGNEPGTCRAVTGTEYVGTEQFTSFCNTSPKPNATKVNVTTTARGRPVSGTEVSRTEKVTGNESGVCRNVTGTEYMSNEAHFSLCGTAAKPSQADKVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQYADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSCRSISGTEYLSNEQFQSFCDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSCSRVTGSQYGQSKICGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEYYGREHFEPFCTSTPEPEAQWSHPQFEK*
pJB64	6xHis-(VTG→AAA MR1, WT MR2-6)-strep	carbenicillin	MSHHHHHHAPKKVETGHTLSGGSAAATQVDRKSHAAANEPGTCRAAAATEYVGTEQFTSFCNTSPKPNATKVNVTTTARGRPVSGTEVSRTEKVTGNESGVCRNVTGTEYMSNEAHFSLCGTAAKPSQADKVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQYADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSCRSISGTEYLSNEQFQSFCDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSCSRVTGSQYGQSKICGGGVGKV

			RSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEYYGREHFEPFCTSTPEPEAQWSHPQFEK*
pJB65	6xHis-(VTG→AAA MR1-6)-strep	carbenicillin	MSHHHHHHAPKKVETGHTLSGGSAAATQVDRKSHAAANEPGTCRAAAATEYVGTEQFTSFCNTSPKPNATKVNVTTTARGRPAAATEVSRTEKAAANESGVCRNAAATEYMSNEAHFSLCGTAAKPSQADKVMFGATARTHQVAAASDEFRPSSAAANESGAKRTITGSQYADEGLARLTINGAPAKVARTHTFAGSDAAATEIGRSTRAAADESGSCRSISGTEYLSNEQFQSFCDTKPQRSPFKVGQDRTNKGQSAAANLVDRSELAAANEPGSCSRAAASQYGQSKICGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEYYGREHFEPFCTSTPEPEAQWSHPQFEK*
pJB75	6xHis-(Y→A MR1-6)-strep	carbenicillin	MSHHHHHHAPKKVETGHTLSGGSVTGTQVDRKSHVTGNEPGTCRAVTGTEAVGTEQFTSFCNTSPKPNATKVNVTTTARGRPVSGTEVSRTEKVTGNESGVCRNVTGTEAMSNEAHFSLCGTAAKPSQADKVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQAADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSCRSISGTEALSNEQFQSFCDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSCSRVTGSQAGQSKICGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGCMPVTGNEAYGREHFEPFCTSTPEPEAQWSHPQFEK*
pBz86	6xHis-(C→S MR1-6)-strep	carbenicillin	MSHHHHHHAPKKVETGHTLSGGSVTGTQVDRKSHVTGNEPGTSRAVTGTEYVGTEQFTSFSNTSPKPNATKVNVTTTARGRPVSGTEVSRTEKVTGNESGVSRNVTGTEYMSNEAHFSLSGTAAKPSQADKVMFGATARTHQVVSGSDEFRPSSVTGNESGAKRTITGSQYADEGLARLTINGAPAKVARTHTFAGSDVTGTEIGRSTRVTGDESGSSRSISGTEYLSNEQFQSFSDTKPQRSPFKVGQDRTNKGQSVTGNLVDRSELVTGNEPGSSSRVTGSQYGQSKISGGGVGKVRSMRTLRGTSVSGQQLDHAPKMSGDERGGSMPVTGNEYYGREHFEPFSTSTPEPEAQWSHPQFEK*
pBz12	6xHis-Rubisco	carbenicillin	CbbL MSAVKKYSAGVKEYRQTYWMPEYTPLDSDILACFKITPQPGVDREEAAAAVAAESSTGTWTTVWTDLLTDMDYYKGRAYRIEDVPGDDAAFYAFIAYPIDLFEEGSVVNVFTSLVGNVFGFKAVRGLRLEDVRFPLAYVKTCGGPPHGIQVERDKMNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENINSQPFMRWRDRFLFVQDATETAEAQTGERKGHYLNVTAPTPEEMYKRAEFAKEIGAPIIMHDYITGGFTANTGLAKWCQDNGVLLHIHRAMHAVIDRNPNHGIHFRVLTKILRLSGGDHLHTGTVVGKLEGDRASTLGWIDLLRESFIPEDRSRGIFFDQDWGSMPGVFAVASGGIHVWHMPALVNIFGDDSVLQFGGGTLGHPWGNAAGAAANRVALEACVEARNQGRDIEKEGKEILTAAAQHSPELKIAMETWKEIKFEFDTVDKLDTQNRHHHHHH* CbbS MAEMQDYKQSLKYETFSYLPPMNAERIRAQIKYAIAQGWSPGIEHVEVKNSMNQYWYMWKLPFFGEQNVDNVLAEIEACRSAYPTHQVKLVAYDNYAQSLGLAFVVYRGN*
pLz48	strep-CsoS1A	carbenicillin	MSWSHPQFEKENLYFQSADVTGIALGMIETRGLVPAIEAADAMTKAAEVRLVGRQFVGGGYVTVLVRGETGAVNAAVRAGADACERVGDGLVAAHIIARVHSEVENILPKAPQA*

Table S3. Summary of the top mass spectrometry results for the four samples submitted: VTG →AAA MR1 and VTG→AAA MR1-6 “dimer” and “trimer” bands (see Ch. 2 Fig. 6). Hit ranking is listed in the table. Keratin samples were excluded.

Input Construct	Band	Rank-ing	Locus	Seq. Count	Spectrum Count	Seq. Coverage	Length	MolWt	pI	NSAF	EMPAI	Descriptive Name
VTGAAAMR1	Dimer	1	splO85041 lCSOS2_HALNC	67	1112	43.50%	869	91934	8.8	0.96225727	1.7227013	Carboxysome assembly protein CsoS2B OS=Halothiobacillus neapolitanus (strain ATCC 23641 / c2) OX=555778 GN=csoS2 PE=1 SV=1
VTGAAAMR1	Dimer	2	gil49176302 lreflYP_026196.1l	4	4	123%	739	83478	9.1	0.004070256	0.3273945	hypothetical protein b4469 [Escherichia coli K12]
VTGAAAMR1	Dimer	3	gil16131219 lreflNP_417799.1l	3	6	6.40%	704	77581	5.4	0.006408919	0.15877736	elongation factor EF-2 [Escherichia coli K12]
VTGAAAMR1	Trimer	1	splO85041 lCSOS2_HALNC	16	150	18.60%	869	91934	8.8	0.9652143	0.53461695	Carboxysome assembly protein CsoS2B OS=Halothiobacillus neapolitanus (strain ATCC 23641 / c2) OX=555778 GN=csoS2 PE=1 SV=1
VTGAAAMR1-6	Dimer	1	splO85041 lCSOS2_HALNC	7	17	7.70%	869	91934	8.8	0.80934626	0.19398808	Carboxysome assembly protein CsoS2B OS=Halothiobacillus neapolitanus (strain ATCC 23641 / c2) OX=555778 GN=csoS2 PE=1 SV=1
VTGAAAMR1-6	Dimer	2	gil90111432 lreflNP_416925.2l	2	2	4.80%	434	47752	8	0.1906538	0.11686325	hypothetical protein b2430 [Escherichia coli K12]
VTGAAAMR1-6	Trimer	1	splO85041 lCSOS2_HALNC	39	200	30.10%	869	91934	8.8	0.4066164	0.99986184	Carboxysome assembly protein CsoS2B OS=Halothiobacillus neapolitanus (strain ATCC 23641 / c2) OX=555778 GN=csoS2 PE=1 SV=1
VTGAAAMR1-6	Trimer	2	trlD0KWZ8 lD0KWZ8_HALNC	3	51	7.40%	284	31724	9.1	0.31726816	0.18576872	Type II secretion system F domain protein OS=Halothiobacillus neapolitanus (strain ATCC 23641 / c2) OX=555778 GN=Hneap_2308 PE=4 SV=1
VTGAAAMR1-6	Trimer	3	gil16131523 lreflNP_418109.1l	2	2	5.30%	693	76430	7.3	0.005098841	0.12979591	DNA helicase, ATP-dependent resolution of Holliday junctions, branch migration [Escherichia coli K12]

A.2 Chapter 3 Supplemental Notes & Figures

Supplementary Notes

Supplementary Figures

Supplementary Notes

Note S1: Microfluidic cell and sample preparation

We performed the trapping experiments in a quartz microfluidic cell that has been previously used successfully for ABEL and ISABEL experiments.[1, 2] The cells consist of two crossed channels that cross at a ~1.5-2 µm thin trapping region and four ports to load solutions and secure platinum electrodes for feedback in *x* and *y*. The top of the cell was chemically bonded with sodium silicate to a 0.15 mm thin quartz coverslip (Esco Optics). All measurements were made with the same cell, labeled C9.

Before each experiment, the internal surfaces of the cell were washed with 1M KOH, then passivated with a polyelectrolyte multilayer (PEM) consisting of four alternating layers of poly(ethylene imine) (PEI, Aldrich) and poly(acrylic acid) (PAA, Aldrich), serving as polycations and polyanions, respectively. The PEM multilayer sequence of PEI/PAA/PEI/PAA resulted in a uniform anionic surface which prevented nonspecific adhesion between anionic carboxysomes and the microfluidic cell via electrostatic repulsion. The layer deposition protocol was followed as previously described.[3]

After cell preparation, carboxysomes were diluted into HEPES buffer (10 mM HEPES, 15 mM NaCl, pH 7.5, 0.2 µm filtered) such that less than one carboxysome would be in the trapping area at any given time. Carboxysomes were stored at 4°C and were viable for up to 6 months after purification. For dilution, carboxysomes were drawn from the top of the stock suspension so as not to retrieve any large aggregates that had precipitated out of suspension. For the air-oxidized data in Figs. 4e-g in the main text, 0.5µL of purified N48-roGFP2-labeled *E. coli* carboxysomes with A_{280} = 10.2 without A_{340} baseline subtraction were diluted 700x into 350 µL of HEPES buffer, then gently mixed four times with ~ 150 µL of the sample volume to break up aggregates. For the trapping data in Figs. 4a-4c, carboxysomes were diluted 700x into HEPES buffer with 1mM TCEP and allowed to sit at room temperature for over an hour to allow the internal roGFP2 to fully reduce. sfGFP-labeled carboxysomes were diluted 200x in HEPES buffer and gently agitated prior to trapping. The concentration of the TCEP stock solution was confirmed with Ellman's reagent.[4]

For the kinetic reduction measurements, equal volumes of 350x-diluted roGFP2-labeled carboxysomes and a solution of 660 µM TCEP in HEPES were mixed then loaded into a cell such that trapping began within 2-3 minutes of mixing. Prior to mixing, the cell was quickly dried with N_2 and loaded onto the microscope stage centered on the trapping region. Feedback voltages were kept minimal to minimize anodic oxidation of TCEP at the electrodes.[5] Signatures of oxidized TCEP reaching the trapping area were detectable after about 40 minutes, as determined by monitoring the delayed fluorescence ratio rise of ~1 µM roGFP2 solution in 200 µM TCEP in the cell with feedback voltages applied on unlabeled polystyrene beads of size comparable to carboxysomes.

Between trapping runs, quartz cells were rinsed with Nanopure water, N_2 dried then cleaned overnight in piranha solution as described in Ref. 4.

Note S2: ISABEL Trap electronics and optics

Near-IR Scatter Illumination

The new trap design incorporates a near-IR diode laser centered at 802 nm to allow the high intensities typical of interferometric scattering experiments (>100 kW/cm^2)[6] while leaving the entire visible spectrum open for fluorescent channels without risk of photobleaching. The relay line for the scatter illumination beam is shown in Fig. S2. Illumination for scattering signal was provided by a multimode 802 nm laser diode (Axcel M9-808-150-D5P), driven by a Thorlabs LDC500 laser diode driver, housed in a Thorlabs TCLDM9 mount, and temperature controlled by a Thorlabs TEC2000. To reduce the beam coherence, a bias tee in the diode mount was supplied with a sinusoidal AC wave from a function generator (Agilent 33220A) at 2.5MHz with $V_{pk\text{-}pk}$ = 2.00 V. With 50Ω impedance, this corresponds to a current modulation of 40 mA peak-peak on top of the 120 mA DC current from the laser diode driver.

The mode of the beam was circularized with a 4:1 anamorphic prism pair (Thorlabs) then focused into a single-mode fiber to produce ~ 15 mW of near-IR at the fiber output. The beam was polarized to horizontal polarization by a λ/2 plate, collimated and reduced to a spot size of 1/e^2 diameter of 0.75 mm (via lenses L_1 and L_2 in Fig. S2) before guiding into the pair of AODs (AA Opto Electronic MT110-B50A1,5-IR). The x and y AODs deflect the beam in a predetermined, 32-point "Knight's tour" scan pattern[3] controlled by the field-programmable gate array (FPGA) on a National Instruments Reconfigurable Input/Output card (NI PCIe-7856R) via a direct digital synthesizer (AA Opto Electronic DDSPA2X-D4125b-34), described in more detail below. The first-order diffraction from each AOD was taken for generating the scan pattern. Lenses L_3 and L_4 map the pivot planes of AODx and AODy to each other, and were chosen to be the same focal length to ensure an equal aspect ratio of the scan pattern in the sample plane. Lenses L_5 and L_6 magnify the beam by 2 (1/e^2 diameter of 1.5 mm) and map the back focal plane (BFP) of the objective to the AODy pivot plane, to ensure purely angular changes in the BFP and thus lateral displacement in the sample plane. L_5 was placed 125 mm down the beam from AODy to allow 400 mm between L_6 and the BFP, ensuring enough room for the remaining optics in the illumination line.

The beam was then sent through a polarizing beam splitter (PBS), followed by a zero-order λ/4 plate, and reflected off a 775 nm shortpass dichroic mirror (Chroma), such that ~8 mW is focused by an NA 1.35, 100x oil immersion objective (Olympus UPlanApo Oil-Iris) to a 500 nm 1/e^2 diameter spot in the sample plane. The beam is held at each position for 18.75 µs, such that the full 3×3 µm^2 area is scanned in 600 µs.

Scattering Detection

The reflected beam back-propagates through the quarter-wave plate, and is rotated to vertical polarization and reflected by the PBS to be collected on a photodiode (Newport 2031). The conjugate image plane at the photodiode is formed by focusing with a f = 400 mm spherical lens after the PBS. An adjustable iris before the photodiode is used to allow only the image of the scan pattern to be incident on the photodiode. The photodiode is connected to a floating analog input channel on the FPGA. For each spot in the scan pattern, the FPGA waits 10 µs, then averages eight ADC-converted analog voltage measurements from the photodiode, each spaced 1 µs apart.

Feedback

The feedback scheme adopted here is the same as previously used.[2] Feedback voltages are applied by two pairs of platinum electrodes every 600 µs, determined by the detected position of the particle on the FPGA and amplified 8x by two op-amp circuits described in detail elsewhere.[7] For feedback, the position setpoint is specified to be near the center of the scan pattern, on a single scan point. Due to interactions between the buffer and the surface charges on the passivation layer, the applied electric fields steer the particle by electroosmosis. The feedback voltages can be tuned by gains (g = 1.6 V/µm) and offsets (<160 mV) in homebuilt LabVIEW software. The FPGA calculates feedback voltage at each frame, and is linear with the displacement of the detected particle position from the trap setpoint in each dimension.

Fluorescence Excitation

The FPGA digitally modulates two collinear fluorescence excitation lasers every 1 ms for two-channel excitation of roGFP2. The 405 nm laser (Coherent Obis LX) is directly modulated by a digital output from the FPGA and spatially overlapped onto the path of the 488 laser beam (Coherent Sapphire), which is modulated via an AOD (Isomet 1205C-2, driver Isomet 222A-1) controlled by the FPGA. Both beams are circularly polarized by quarter-wave plates and attenuated such that ~30-40 µW in each beam reach the sample plane. The collinear beams are focused with an f = 400 mm Köhler lens and reflected off a multi-band dichroic (Semrock Di03-R405/488/561/635-t1-25x36) for wide-field illumination in the sample plane (1/e^2 radius ≈ 4 µm). Low intensity (< 50 W/cm^2) is necessary at 405 nm to reduce the probability of excited state proton transfer[1] and photoconversion of roGFP2 chromophores over extended trapping times (Fig. S3). For these experiments, the peak intensities of the 488 and 405 nm excitation beams are 80 and 40 W/cm^2, respectively.

Fluorescence Detection

Emitted fluorescence is spectrally separated from the scatter illumination and fluorescence excitation beams by sequential transmission through the 775 nm shortpass and multi-bandpass dichroic mirrors specified above. The fluorescence is spatially filtered by a 75 µm-diameter pinhole centered on the trapping center position (corresponding to 1.5 µm diameter at the sample) and spectrally filtered to collect the 500-570 nm emission band of roGFP2 (Fig. S9) before detection on an avalanche photodiode (PicoQuant τ-SPAD) connected to the FPGA. To distinguish photons emitted from 405 and 488 nm excitation, each time-tagged photon is additionally tagged with the identity of the excitation laser for separation of the two channels in post-processing.

Control

All of the components above are controlled and synchronized by an FPGA on a reconfigurable input-output board (NI PCIe 7856) with an 80 MHz clock and custom software written in LabVIEW, as previously implemented.[2] This control allows for the calculation of absolute fractional scattering contrast for each 600 µs frame. The FPGA then calculates and applies voltages in x and y that is proportional to the displacement of the maximum absolute fractional scattering contrast to the feedback setpoint.

Note S3: Analysis

Calculation of Interferometric Scattering Contrast

Absolute fractional scattering contrast describe in Eq. 2 of the main text is calculated in real time on the FPGA for particle localization, described in detail previously.[2] Before the measurement, we measure the ADC counts on the detector with the near-IR beam blocked as an offset to subtract from the measured ADC counts at each point. During the experiment, a 10-ms average of the ADC counts at each scan point is taken when feedback is turned off to provide the reference background intensity $|E_r|^2$ in Eq. 2. Subsequent measurements are divided by this background at that scan point to help identify the pixel of maximum fractional scattering contrast over each set of 32 beam-dwell times in a “frame”. To account for small power fluctuations in laser power between each 600 ms frame, the average of the ADC values of the outermost 16 scan points are used to normalize each background measurement.[2] For post-processing and analysis, the ADC counts (averaged from 8 consecutive measurements as described above) from each point and their corresponding background measurements are saved. The absolute fractional scattering contrast for offline analysis plotted in Fig. 3 of the main text and used in level-finding is determined from the maximum value within a 3x3 pixel area around the trap setpoint.

Determination of fluorescence and scattering contrast levels

Mean brightness and scattering contrast levels from each trapping event are determined by a changepoint-finding algorithm[8] on the 488 nm level trace due to its highest signal-to-background ratio. We empirically used a relatively strict threshold for finding changepoints (20.0 for the log likelihood ratio in the level-change hypothesis test in Eq. 6 in Ref. 8), because we wanted to maintain a level over a single trapping event even with fluctuations in fluorescence brightness due to blinking from individual roGFP2 molecules within each carboxysome. This reduces the false-positive rate during a single trapping event, but also comes at the expense of missing potential trapping events. Level values were calculated as mean values between two changepoint times. The determined changepoint times were used on the 405 nm and scattering contrast traces to determine their average values in the same time intervals. These levels can consist of trapping events or times with only background.

The determined levels are filtered and processed to ensure each point in the 2D scatter plots of Fig. 4 of the main text correspond to trapped carboxysomes. Levels that meet all of the following criteria are kept for further analysis, while levels that do not meet the criteria are rejected. Admissible levels:

- Only occur when feedback is ON, to remove stuck particles that persist when feedback is OFF.
- Only occur when feedback voltage polarity is correct, to make sure all particles have the same sign on charge.
- Span at least 200 ms, so that particles diffusing in the ROI without feedback are ignored.

- Have brightness in the 488 channel exceeding 2σ from the mean of background levels, determined as described below.
- Have brightness in the 405 channel at least 1 count/10 ms above background.
- Have absolute fractional scattering contrast between 0.0025 and 0.025, to ignore small particles and large aggregates that do not correspond to single or double carboxysomes.

Fluorescence background values for each level in the 488 and 405 nm channels are determined as the mean of the fluorescence levels during the trap OFF times immediately before and after the trapping event. These background levels are taken as their median fluorescence to mitigate the influence of fluorescence bursts from rapidly diffusing objects.

We also merge consecutive levels with absolute fractional scattering contrasts that are within 10% of each other, since fluorescence dynamics such as blinking in the 488 nm channel may give the false appearance of separate levels within a single trapping event.

In summary, particles saved for analysis reflect objects that are trapped for at least 200 ms and show scattering contrasts in an appropriate range and also show fluorescence brightness in both 405 and 488 channels. These filters are reflected in the traces shown in Fig. 4 of the main text and are used to determine the fluorescence ratios from each particle.

Determination of fluorescence ratios

Ratios from each trapping event are determined via the following expression

$$R_{405/488} = \frac{I_{405} - b_{405}}{I_{488} - b_{488}}, \qquad \text{(S1)}$$

where I_{405} and I_{488} refer to the mean fluorescence brightness over the trapping event in the 405 and 488 nm channels respectively, and b_{405} and b_{488} refer to the background levels determined for each level. To account for day-to-day changes in pointing of the fluorescence excitation beams, we normalize each ratio to the ratiometric fluorescence from Rhodamine 110 in water (R_{min}) and Coumarin 6 in absolute ethanol (R_{max}), measured after each trapping experiment, whose excitation spectra are shown in Fig. S10. For the 405 nm and 488 nm intensities used in this experiment, $R_{\text{min}} \approx 0.02$ and $R_{\text{max}} \approx 0.45$. Ratio normalizations are calculated according the following expression:

$$R_{norm} = \frac{R_{405/488} - R_{min}}{R_{max} - R_{min}}, \qquad \text{(S2)}$$

Such that the ratio values from the Rhodamine 110 and Coumarin 6 standards are 0 and 1, respectively. We initially attempted ratio normalizations with roGFP2 pushed to oxidized and reduced extremes with 1 mM H_2O_2 and 1 mM TCEP, respectively, but found that the obtained ratio values were not as reliable. This may have arisen from unspecific adsorption of roGFP2 to the microfluidic cell surfaces in a way that impacted the protonation equilibrium of the roGFP2 chromophore.

Note S4: Uncertainty Estimation

Standard errors on the two fluorescence levels and scattering contrast were determined for each trapping event via a bootstrapping function in MATLAB. For each level, 100 bootstrapped means were determined from the values within a level by sampling with replacement, and the standard error on the mean (SEM) was taken as the standard deviation on the distribution of bootstrapped means. We compare the error statistics for 488 levels, 405 levels, and ratios from oxidized roGFP2-carboxysomes, reduced roGFP2-carboxysomes, and sfGFP carboxysomes in Tables S1-S3 below. As shown in Fig. S8a-f, the uncertainties in air-oxidized carboxysomes were compared to the expected SEM with Poisson statistics (assuming a level with a mean bright *I* plus background *b*), as an “excess noise ratio” (measured SEM/shot noise SEM). For example, the RMS excess noise ratio on the 488 levels in reduced carboxysomes was 1.52x the expected value for shot noise. By contrast, the RMS excess noise ratio in the 405 channel was only 1.04, indicating closer agreement with the shot noise limit.

The SEMs and mean values were used to propagate the uncertainty to ratio measurements:

$$\left(\frac{SEM_R}{R}\right)^2 = \left(\frac{SEM_{405}}{I_{405}}\right)^2 + \left(\frac{SEM_{488}}{I_{488}}\right)^2. \quad \text{(S3)}$$

In Figs. S7g-S7i, the fractional uncertainty on the 405 channel levels are higher due to the lower brightness and larger contribution from background. The fractional uncertainty of the ratio measurements is therefore dominated by the uncertainty in the level of the 405 nm channel.

For the normalized ratios presented in the main text, the fractional uncertainties were scaled accordingly. From Eq. S2, the standard error SEM_{norm} on the normalized ratio R_{norm}, can be calculated as follows:

$$\frac{SEM_{norm}}{R_{norm}} = \frac{SEM_R}{R_{max}-R_{min}} \cdot \left(\frac{R-R_{min}}{R_{max}-R_{min}}\right)^{-1} = \frac{SEM_R}{R-R_{min}}. \qquad \text{(S4)}$$

Table S1: Error Statistics for oxidized roGFP2-carboxysomes

Channel	Mean	RMS SEM	RMS Fractional error	RMS Excess Noise Ratio
488 nm	39.9 cts/10ms	1.0 cts/10ms	0.04	1.37
405 nm	7.3 cts/10ms	0.5 cts/10ms	0.12	1.05
Norm. $R_{405/488}$	0.45	0.06	0.12	1.08

Table S2: Error Statistics for reduced roGFP2-carboxysomes

Channel	Mean	RMS SEM	RMS Fractional error	RMS Excess Noise Ratio
488 nm	57.6 cts/10ms	1.6 cts/10ms	0.04	1.50
405 nm	5.3 cts/10ms	0.5 cts/10ms	0.14	1.03
Norm. $R_{405/488}$	0.29	0.06	0.18	1.05

Table S3: Error Statistics for sfGFP-carboxysomes

Channel	Mean	RMS SEM	RMS Fractional error	RMS Excess Noise Ratio
488 nm	583.43 cts/10ms	7.0 cts/10ms	0.01	2.11
405 nm	42.3 cts/10ms	1.0 cts/10ms	0.02	1.11
Norm. $R_{405/488}$	0.01	7×10^{-4}	0.07	1.19

Note S5: Growth and purification of labeled *E. coli* carboxysomes

E. coli BW25113 harboring a plasmid containing the *H. neapolitanus* HnCB10 carboxysome operon from Bonacci et al.[9] and a plasmid containing roGFP2 with an N-terminal carbonic anhydrase "N48" tag (first 53 amino acids of CsosCA) were inoculated into 1 L of LB at half concentrations of antibiotic on each plasmid. Cells were grown to log phase and induced with 500 µM IPTG and 100 nM aTc, then grown overnight at 18°C before pelleting and freezing at -20°C. Frozen pellets were lysed in BPER reagent with the addition of 10 mM $MgCl_2$, 20 mM $NaHCO_3$, 1 mM EDTA, 0.1 mg/mL lysozyme, 1 mM PMSF, and 1 µL benzonase/25 mL lysate. Lysis occurred for 1 hour on a shaking rotor at room temperature. To clarify, cells were spun at 12,000 xG for 15 minutes, and the supernatants spun again at 40,000 xG for 30 minutes to pellet carboxysomes. The pellets were resuspended in 1.5 ml of TEMB buffer (10 mM Tris-HCl, 10 mM $MgCl_2$, 20 mM $NaHCO_3$, 1 mM EDTA, adjusted to pH 8) and gently shaken on ice from 30 minutes to overnight to loosen the pellet before resuspending. Resuspended pellets were spun at 900 xG for 3 minutes to pellet insoluble junk before loading onto a 25 ml 10-50% sucrose gradient. The sucrose gradient was spun at 105,000 xG for 35 minutes and 1 ml fractions collected and analyzed on SDS-PAGE. Carboxysome fractions were pooled and spun at 100,000 xG for 90 minutes, the supernatant dumped, and pellets resuspended in 250 ul of TEMB and stored at 4°C. The final concentration of roGFP2 labeled carboxysomes as determined by absorbance at 280nm was 10.2. The A_{280} reported here is without 340 nm baseline subtraction because carboxysomes contribute to baseline scattering. Superfolder-GFP *E. coli* carboxysomes with FLAG tag on the N-terminus of CsoS1A were prepared and purified in the same was as roGFP2-labeled *E. coli* carboxysomes. Carboxysomes were stored at 4°C and were viable for up to 6 months after purification.

Note S6: Quantifying roGFP2 brightness and roGFP2 loading per carboxysome

Copy numbers of active roGFP2-labeled carboxysomes were determined by analyzing bleachdown traces for single-step roGFP2 photobleaching. roGFP2-labeled carboxysomes were diluted 1200x in HEPES buffer and reduced for 1 hour with 1 mM TCEP, so as to push the equilibrium to the 488 nm-absorbing anionic chromophore. The reduced carboxysomes were then introduced into a hybridization cell (Grace Bio-Labs Secure Seal) stuck to a quartz coverslip (Esco Optics) and coated with PEI/PAA/PEI as described in Note S1. The polycation PEI served as the final layer to encourage electrostatic adsorption to the coverslip surface. Carboxysomes were then illuminated with 488 nm (240 W/cm^2 peak intensity, 1/e^2 radius ~ 4μm) and imaged at 20 fps on an sCMOS camera (Andor Zyla 4.2) for 2 minutes to allow for complete photobleaching. The camera was calibrated[10] to convert ADU counts to photons. Time traces were constructed from the 5x5 pixel ROI (875x875 nm^2) to collect all detected photons around each carboxysome point-spread function. Brightness step sizes due to photobleaching or blinking were determined via a changepoint-finding algorithm[8] on the time traces, then manually examined to remove spurious step changes. The illumination profile was also measured to normalize measured brightness steps and exclude carboxysomes where excitation intensity was less than 1/e of the peak. Normalized brightness steps were collected to give the histogram in Fig. S1a. Because multiple steps may occur within the 50 ms frame, there is an expected tail to higher brightnesses. Total carboxysome loading (Fig. S1b) was estimated by taking the ratio of the initial brightness level determined by the changepoint algorithm over the median single brightness step size, normalized by excitation beam intensity. Most carboxysomes appear to contain 3-15 active roGFP2 molecules, but we also observe a tail to higher loading.

Note S7: Imaging and Sizing Carboxysomes with Cryo-Electron Microscopy

3 µL of carboxysome suspension (diluted 10x in HEPES buffer from stock) was deposited onto a glow-discharged holey carbon electron microscopy grid (Quantafoil R 2/2 G200F1), blotted on both sides for 2.5 seconds, and plunge frozen (Gatan CP3). Electron micrographs were acquired on a 200-keV electron microscope (Thermo Fisher Glacios) equipped with a direct detector (Gatan K2). Images were acquired with pixel spacing of 2.43 Å and 12.7 Å, with defocus targets of -3 µm and -40 µm respectively.

Effective diameters were determined by measuring the area of the polygon traced around the carboxysome shells in Fiji,[11] then evaluated as $d_{\text{eff}} = (4A/\pi)^{1/2}$ assuming circular shape. As shown in Fig. S5, the diameters appear normally distributed with $\mu = 141$ nm and $\sigma = 31$ nm. This is in contrast to the bimodal scattering contrast histograms shown in Fig. 4 of the main text, strongly suggesting that the larger-constrast feature is due to carboxysome dimers. The distribution of effective radii (coefficient of variation $\sigma/\mu = 0.22$) also results in a spread of volumes, which generates substantial overlap between the single- and double-carboxysome features in the scattering contrast histogram.

Supplementary Figures

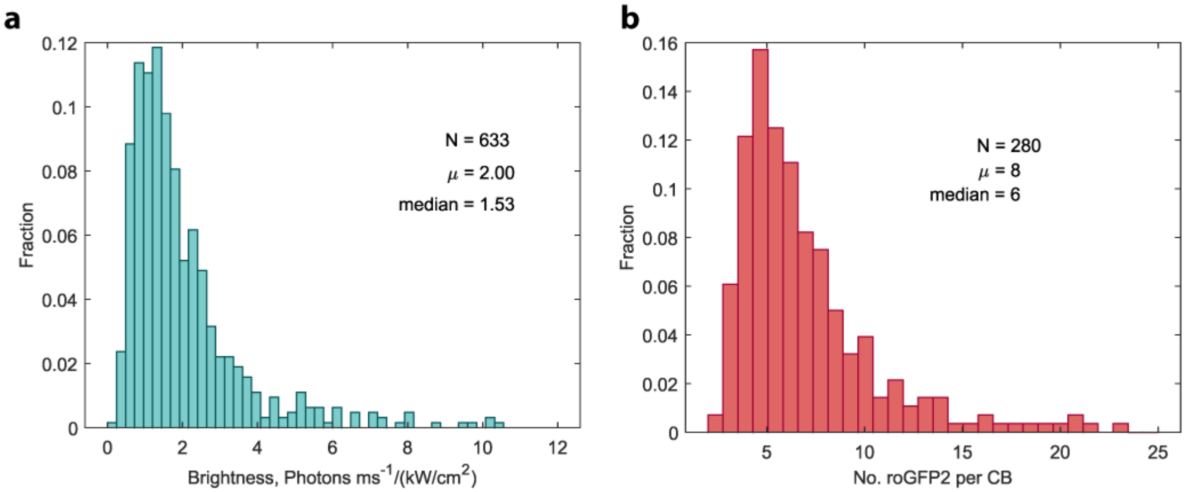


Figure S1. Quantifying roGFP2 single-molecule brightness and roGFP2 loading in individual carboxysomes. (a) Distribution of roGFP2 brightnesses measured via single-step photobleaching events in reduced carboxysomes, described in detail in Note S6 above. This histogram primarily reports on the bleaching of single roGFP2 copies, but the possible simultaneous bleaching of two or three copies are also incorporated into this histogram. (b) Estimation of number of roGFP2 copies per carboxysome, determined by dividing the initial brightness of each carboxysome by the median single-roGFP2 value measured in (a).

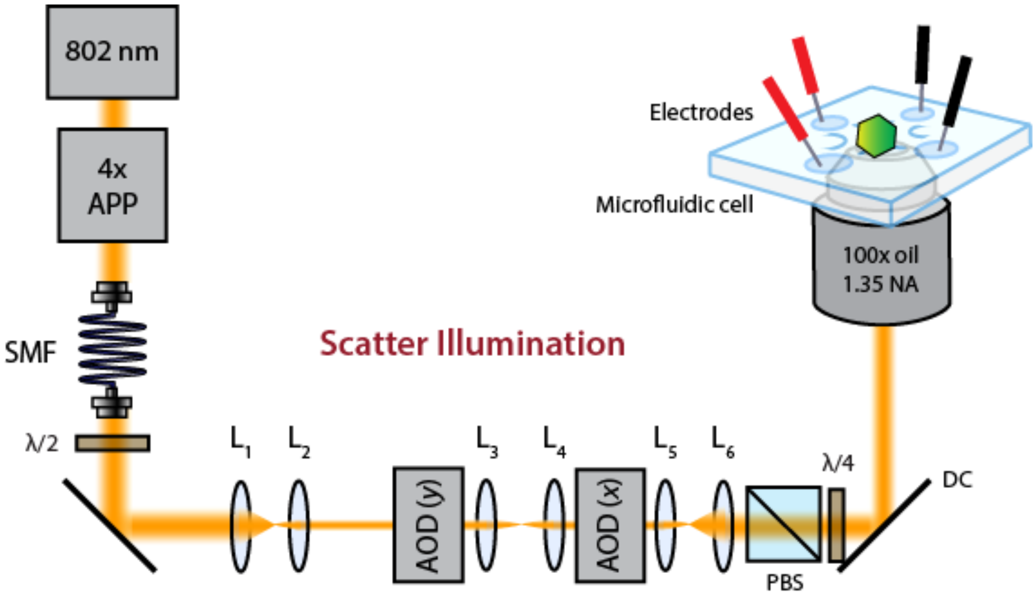


Figure S2. Scatter illumination relay line, described in detail in Note S2. 4x APP: 4x anamorphic prism pair. SMF: single-mode fiber. λ/2: half-wave plate. L_1-L_6: spherical lenses with focal lengths f_1 = 100 mm, f_2 = 50 mm, f_3 = 60 mm, f_4 = 60 mm, f_5 = 150 mm, f_6 = 300 mm. AOD: acousto-optic deflector. PBS: polarizing beam splitter. λ/4: quarter-wave plate. DC: dichroic mirror.

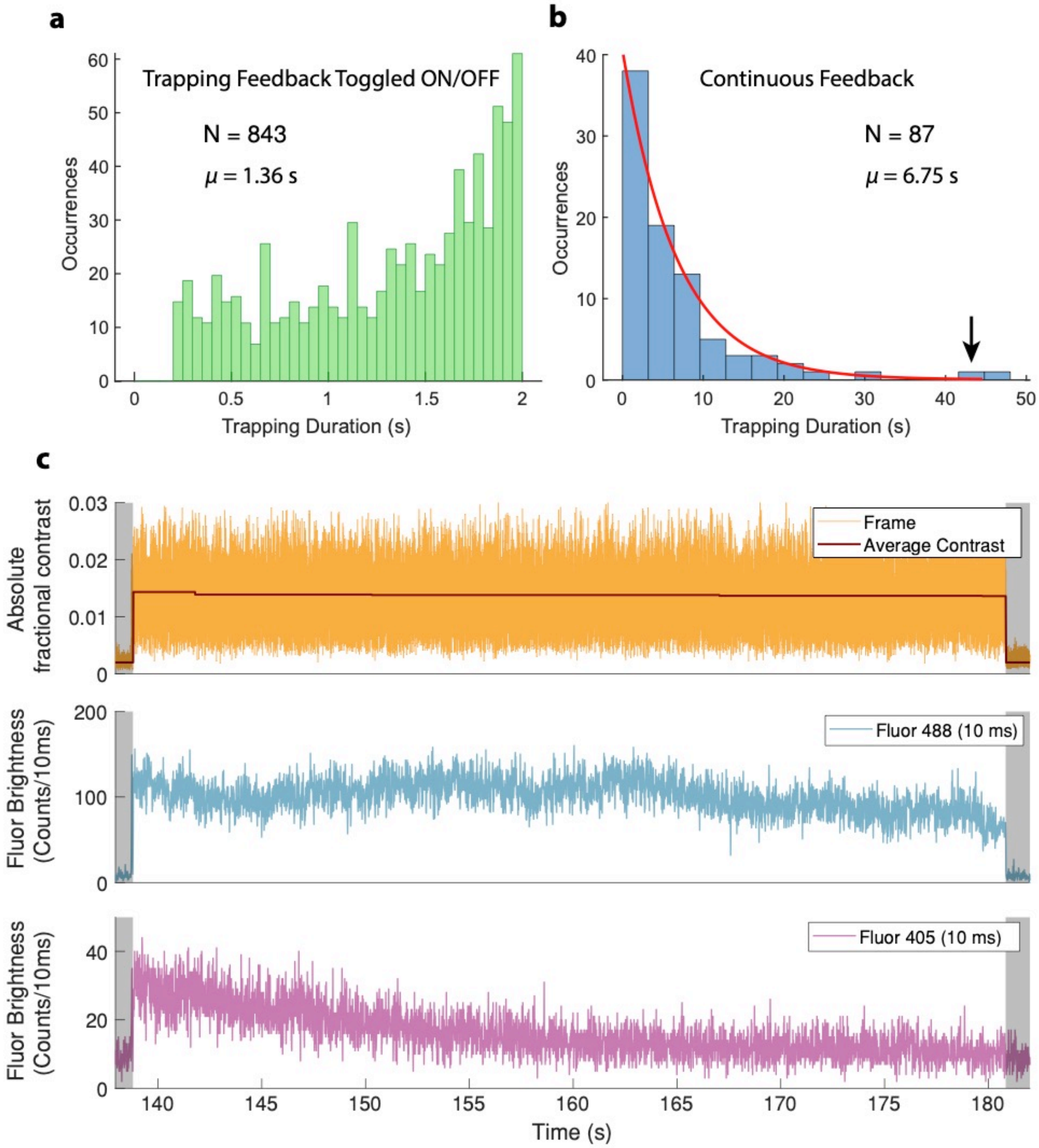


Figure S3. Statistics for carboxysome trapping durations. (a) Durations for the dataset shown in Fig. 4e-g, where feedback is toggled on and off for 2s and 1s, respectively, and minimum duration is set to 0.3s. (b) Trapping durations when feedback is not toggled, showing a longer average trapping duration. Black arrow points to the trapping event shown in (c). In this continuous feedback case, the majority of trapping events ended due to replacement by a new particle, so in principle one should be able to achieve longer trapping times by reducing the concentration of carboxysomes. (c) Carboxysomes can be trapped for more than 40 seconds via localization with interferometric scattering. Top trace: absolute fractional scattering contrast remains level while trapping feedback is ON. Middle trace: 488 nm-excited fluorescence trace shows a consistent brightness over the whole trapping event. Bottom trace: 405 nm excitation fluorescence shows a gradual reduction in brightness due to excited state proton transfer (ESPT)[12] from the chromophore.

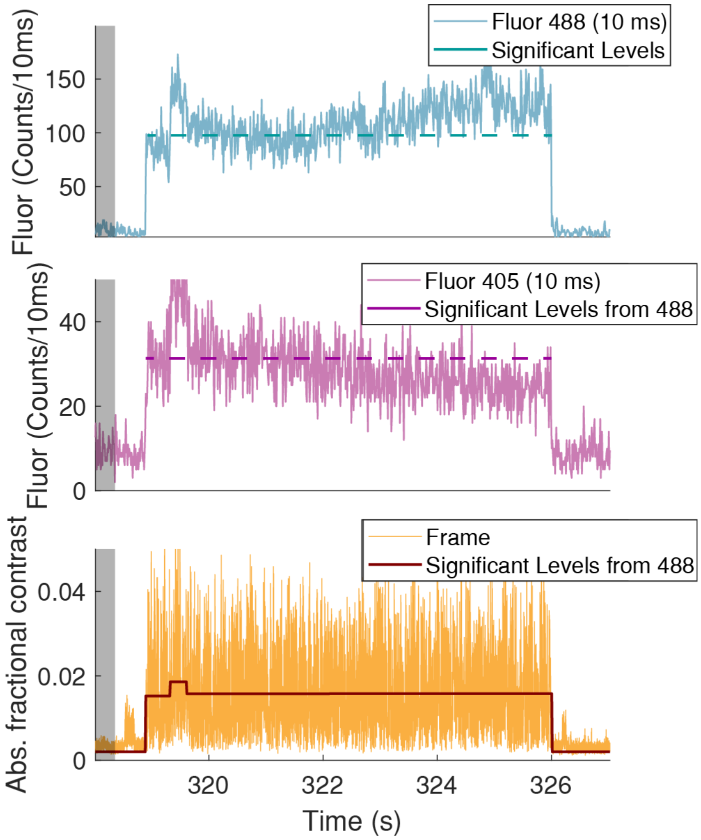


Figure S4. An example of 405 nm-induced excited state proton transfer (ESPT) introducing slow drifts into fluorescence levels. Top and middle traces: 488 fluorescence brightness increases and 405 fluorescence decreases as the roGFP2 chromophore converts to its anionic form. Dashed lines indicates mean fluorescence brightness levels from around *t* ≈ 320 s. Bottom trace: absolute fractional scatter trace with average value. From *t* = 320 s onwards, the scatter contrasts remains level, indicating the same particle is trapped until *t* = 326 s.

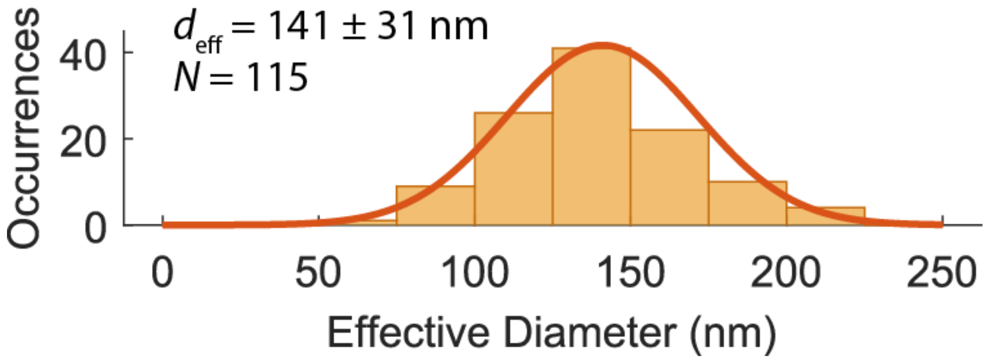


Figure S5. Distribution of effective diameters of roGFP2-labeled *E. coli* carboxysomes determined with cryo-EM imaging (see Note S7). Error on d_{eff} denotes standard deviation. Dark orange line shows a Gaussian fit to the histogram.

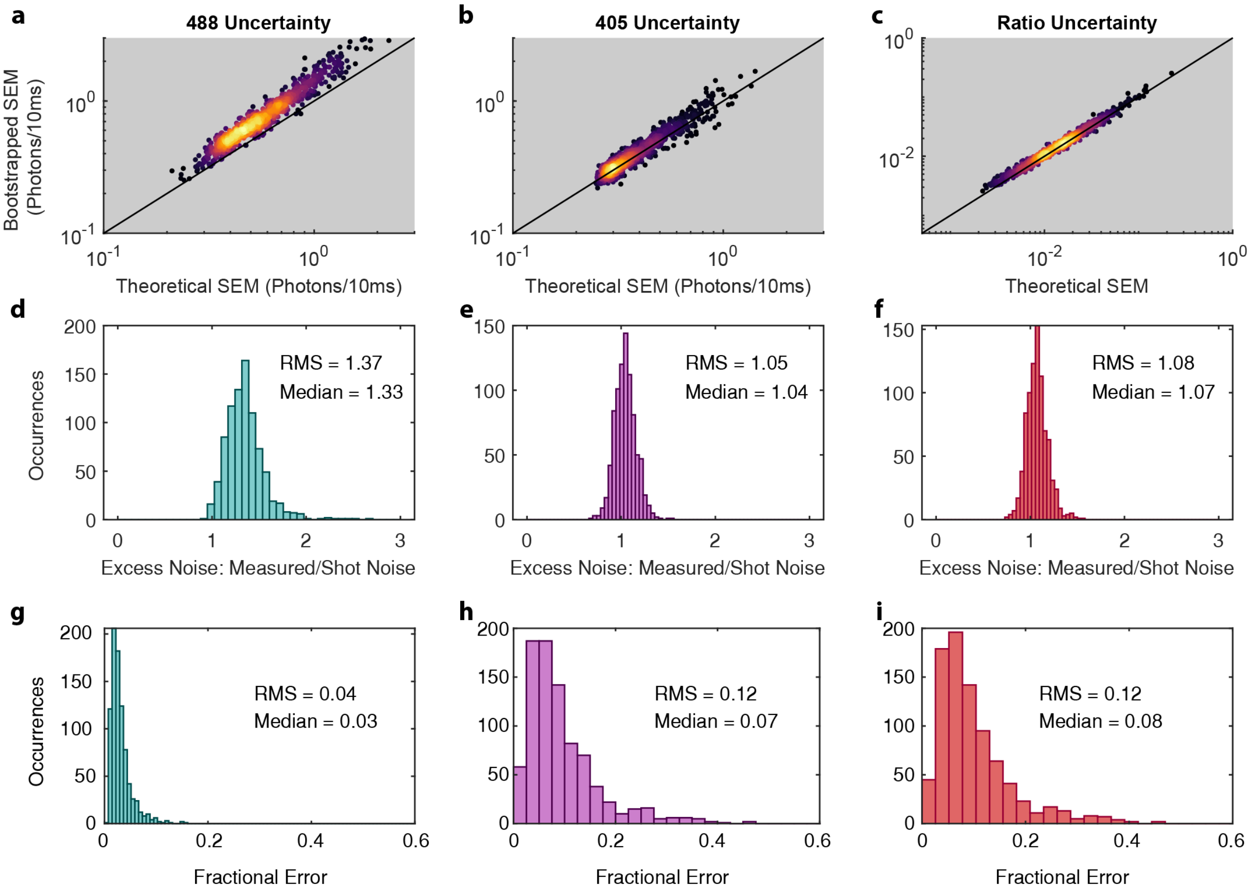


Figure S6. Statistics of standard errors on the mean (SEM) of each brightness level for carboxysomes in air-oxidized buffer. Derivations of the quantities plotted are presented in Supplemental Note S4. (a)-(c) Comparison of the theoretical and measured SEMs on 488 brightness, 405 brightness, and ratio, respectively. Theoretical SEM is based on expected uncertainty arising from shot noise, while measured SEMs are derived from bootstrapping from the brightness values within each level. The black diagonal line represents when the measured noise is equal to shot noise. Ratio SEMs in both dimensions are determined by error propagation on the 488 and 405 nm SEMs. (d)-(f) Histograms of the excess noise ratio of measured and shot noise SEMs on each level for 488, 405, and ratio uncertainties, respectively. (g)-(i) Fractional uncertainties SEM/μ for 488, 405, and ratio levels. The bottom row demonstrates that the biggest source of uncertainty in the ratio measurement arises from noise on the 405 levels. This arises from the lower brightnesses and higher backgrounds on these levels compared to 488 nm.

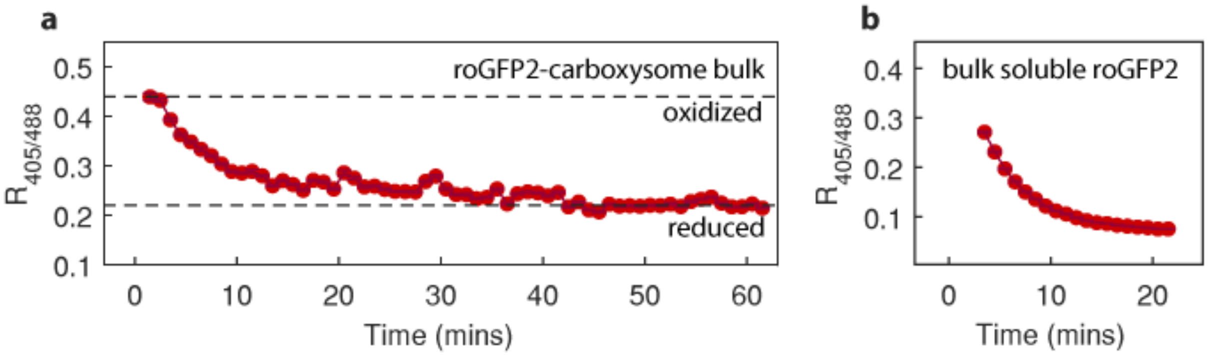


Figure S7. Ratiometric fluorescence kinetics from bulk samples after adding 330 µM TCEP: (a) roGFP2-labeled carboxysomes and (b) soluble 0.5 µM N48-roGFP2. The dashed lines in (a) show the typical ratio values in oxidizing and reduced conditions, consistent with the mean ratios measured from single carboxysomes. Ratio measurements start after approximately 2 minutes due to time needed for sample loading after mixing. Soluble roGFP2 reduces ratio more quickly in free solution and reduce to lower ratio values than roGFP2-labeled carboxysomes.

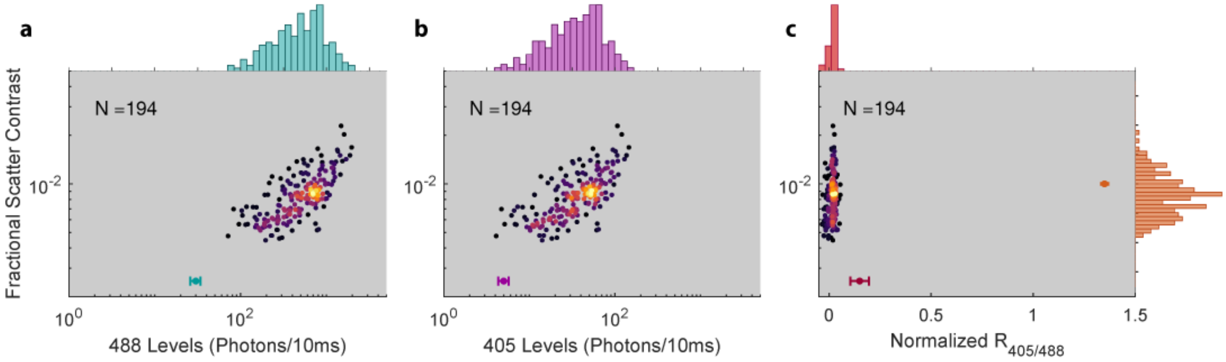


Figure S8. Scatter plots relating brightnesses, scatter contrast, and fluorescence ratio in trapped sfGFP carboxysomes.

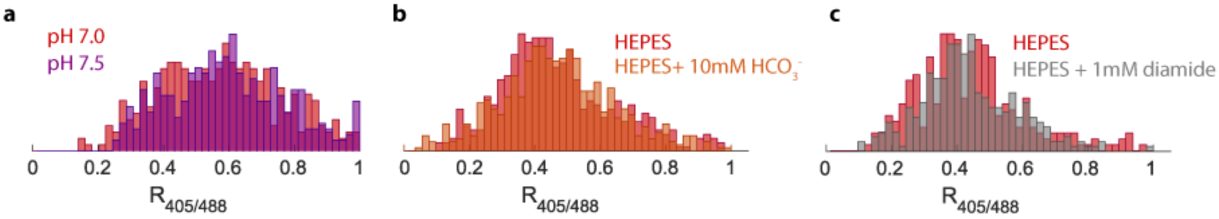


Figure S9. Changes in buffer chemistry do not significantly impact the wide ratio spread in air-oxidized carboxysomes. (a) Comparison between trapped carboxysomes in pH 7.0 and pH 7.5 citrate-phosphate buffer. (b) Comparison between trapped carboxysomes in HEPES buffer with or without 10 mM sodium bicarbonate. (c) Comparison between trapped carboxysomes in HEPES buffer with or without 1 mM diamide. This test confirms that air-oxidized carboxysomes are indeed already fully oxidized. In (b) and (c), HEPES buffer refers to 10 mM HEPES + 15 mM NaCl, pH 7.5 as used in trapping experiments.

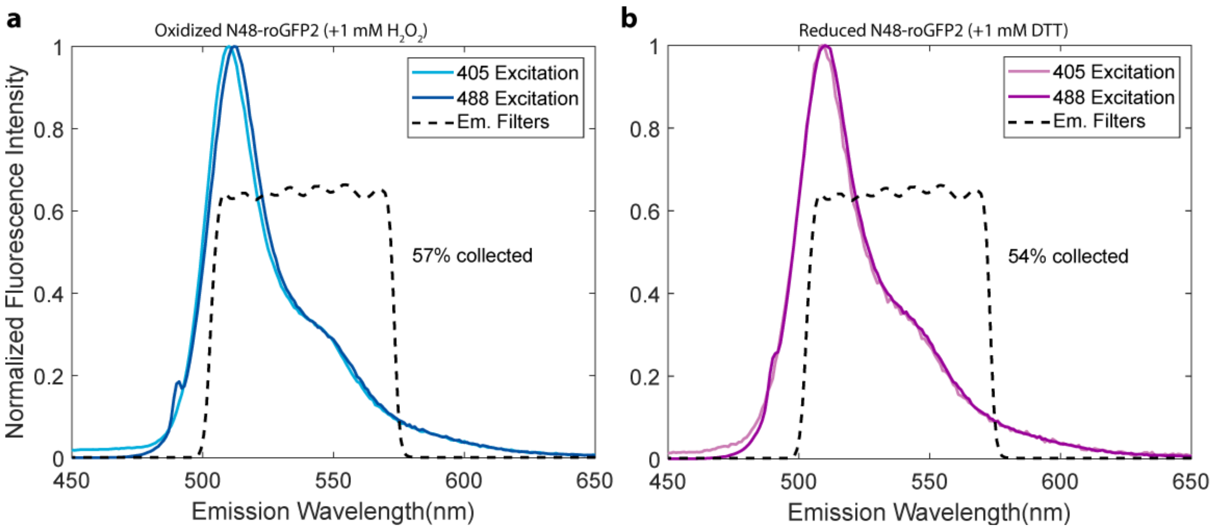


Figure S10. Emission spectra from 1µM N48-roGFP2 in Tris buffer when pushed to (a) full oxidation or (b) full reduction. The peak position can shift by a few nm, but the whole emission spectrum is collected by our 500-570 nm emission filter set regardless of redox state. Filter set: Chroma HQ500 longpass, Semrock EdgeBasic 488 longpass, 570 shortpass, 785 shortpass, 808 notch.

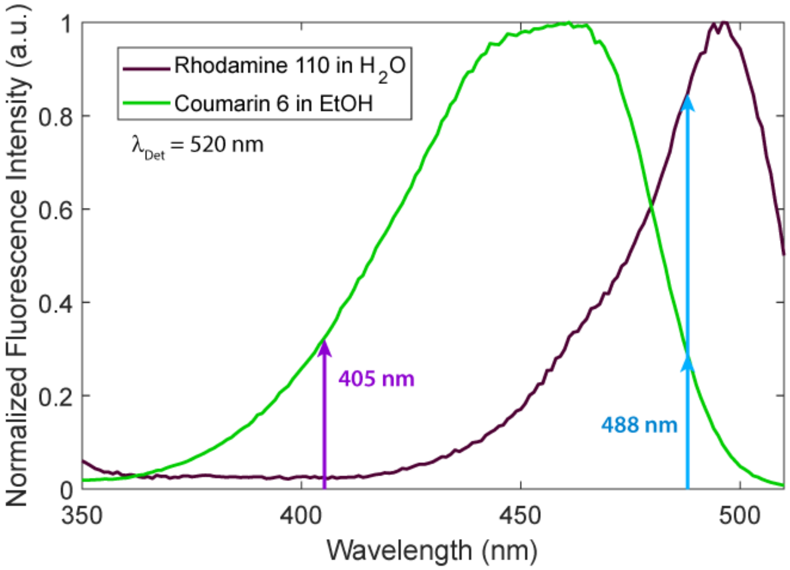


Figure S11. Fluorescence excitation spectra of Rhodamine 110 and Coumarin 6 used as standards for normalizing measured roGFP2 ratios to low and high values, respectively. See Note S3 for ratio normalization.

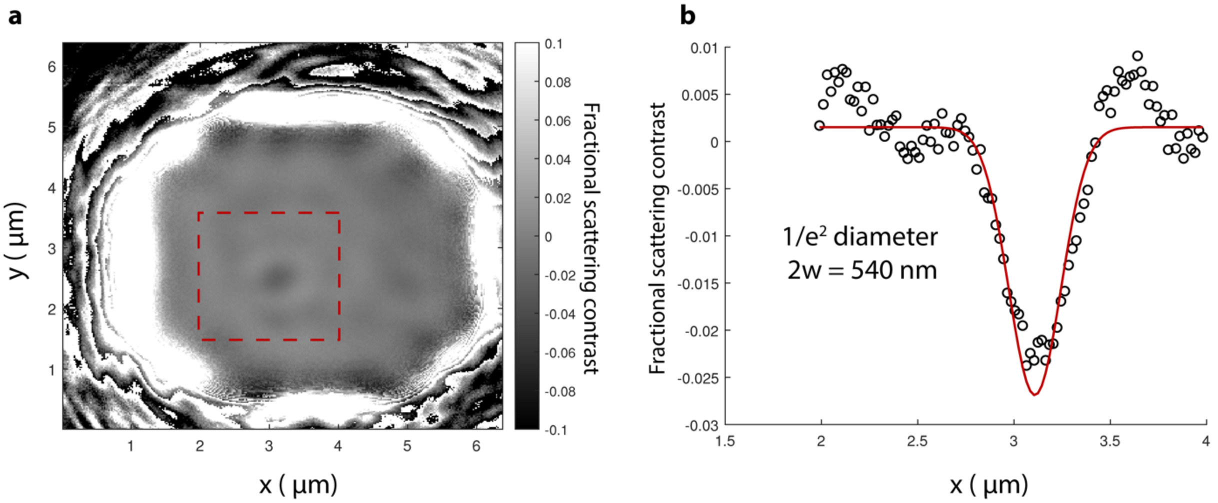


Figure S12. (a) Image of the interferometric scattering PSF from a 40 nm Au bead on a glass coverslip. Image taken with FLIR Blackfly S U3-19S4M CMOS camera. Signed fractional scatter contrast was calculated with the average of 120 signal frames and 600 background frames. Dotted red line indicates the ROI used for 2D Gaussian fit to PSF. (b) Cross-section of PSF and 2D Gaussian fit. This PSF is comparable in size to the 500×500 nm^2 area of one beam position in the Knight's tour scan pattern.

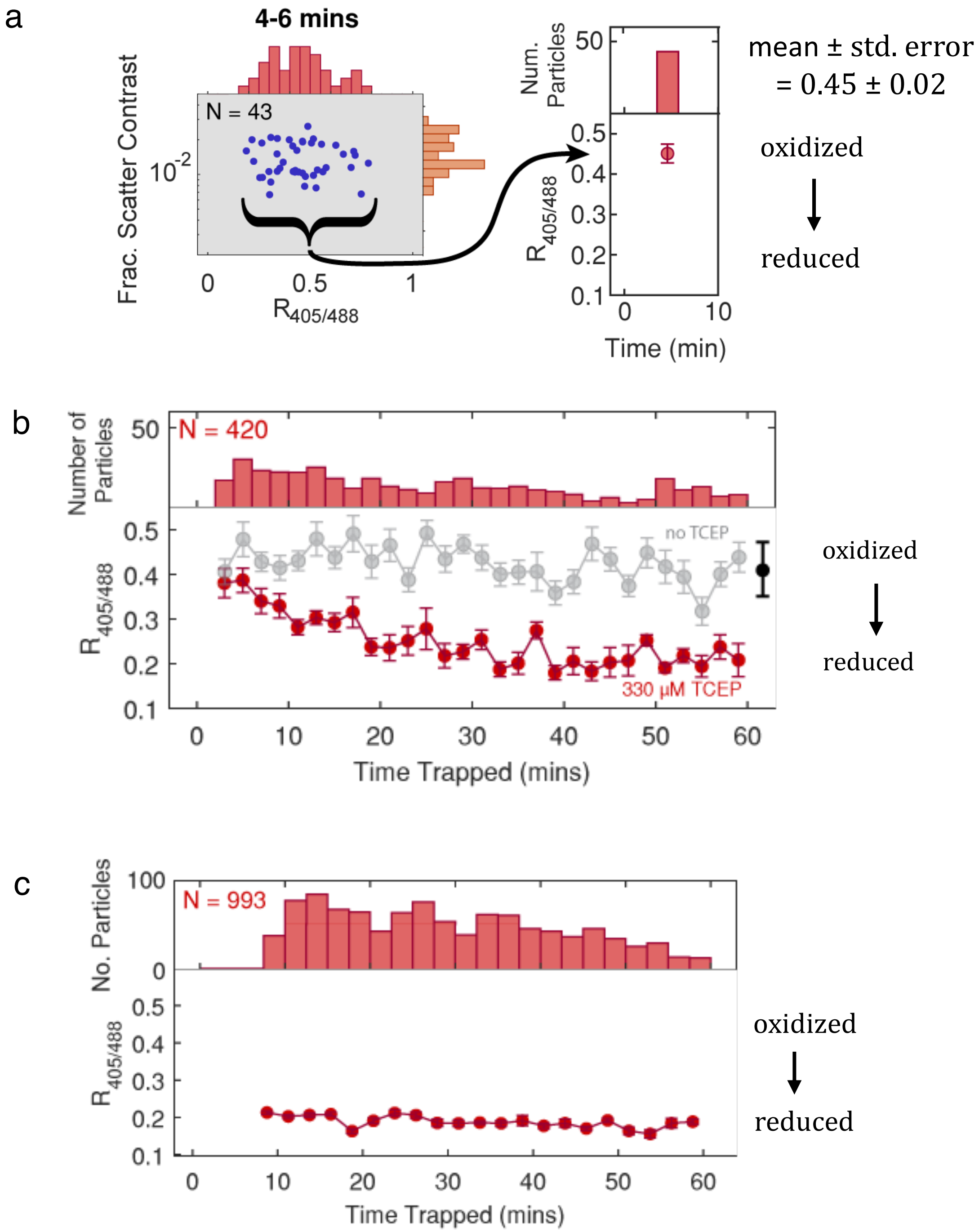


Figure S13. Kinetics of small-molecule reductant permeating into carboxysomes with internal roGFP2. Reductant (330 uM TCEP) was added at t=0 to air-oxidized carboxysomes. Carboxysomes loading onto the trap took 2-3 minutes. (a) Individuals were trapped for a short period of time and binned into 2-minute intervals. (b) Internal reduction of wild-type *E.c.* carboxysomes with and without TCEP over 60 minutes. (c) Internal reduction of ΔCsoS4 mutant carboxysomes. The deleted pentamers leave large gaps in the shell, leading to higher permeability.

LoopName	PV	Setpoint	Out%	Mode	Units	Casc
Agit	200	200	23.0	Auto	RPM	None
Temp	30.2	30.0	4.3	Auto	DegC	None
pH	6.52	6.80	100.0	Auto	pH	Source
DO	-0.0	0.0	-100.0	Off	%DO	None
Air (1)	100.0	100.0	100.0	Manual	%	None
O2 (2)	0.0	0.0	0.0	Manual	%	None
N2 (3)	0.0	0.0	0.0	Manual	%	None
CO2 (4)	0.0	0.0	0.0	Manual	%	None
GasFlo	4.0	4.0	20.0	Auto	SLPM	None

Figure S14. Growing 20L of *H. neapolitanus* in a chemostat. (A) The chemostat setup. (B) Chemostat setup showing the input and output media collection bottles. (C) Continuous optical density (OD) graph. (D) Summary page of the control software. (E) Cells ready for harvesting. (F) Squat, don't bend your back, when lifting heavy things! (G) Pretty purple culture that is ready to spin down. (H) Agonizingly small pellets for each 1L bottle. (I) Tubes of 3g of cell pellets from 7L each that are ready for freezing at -20°C.

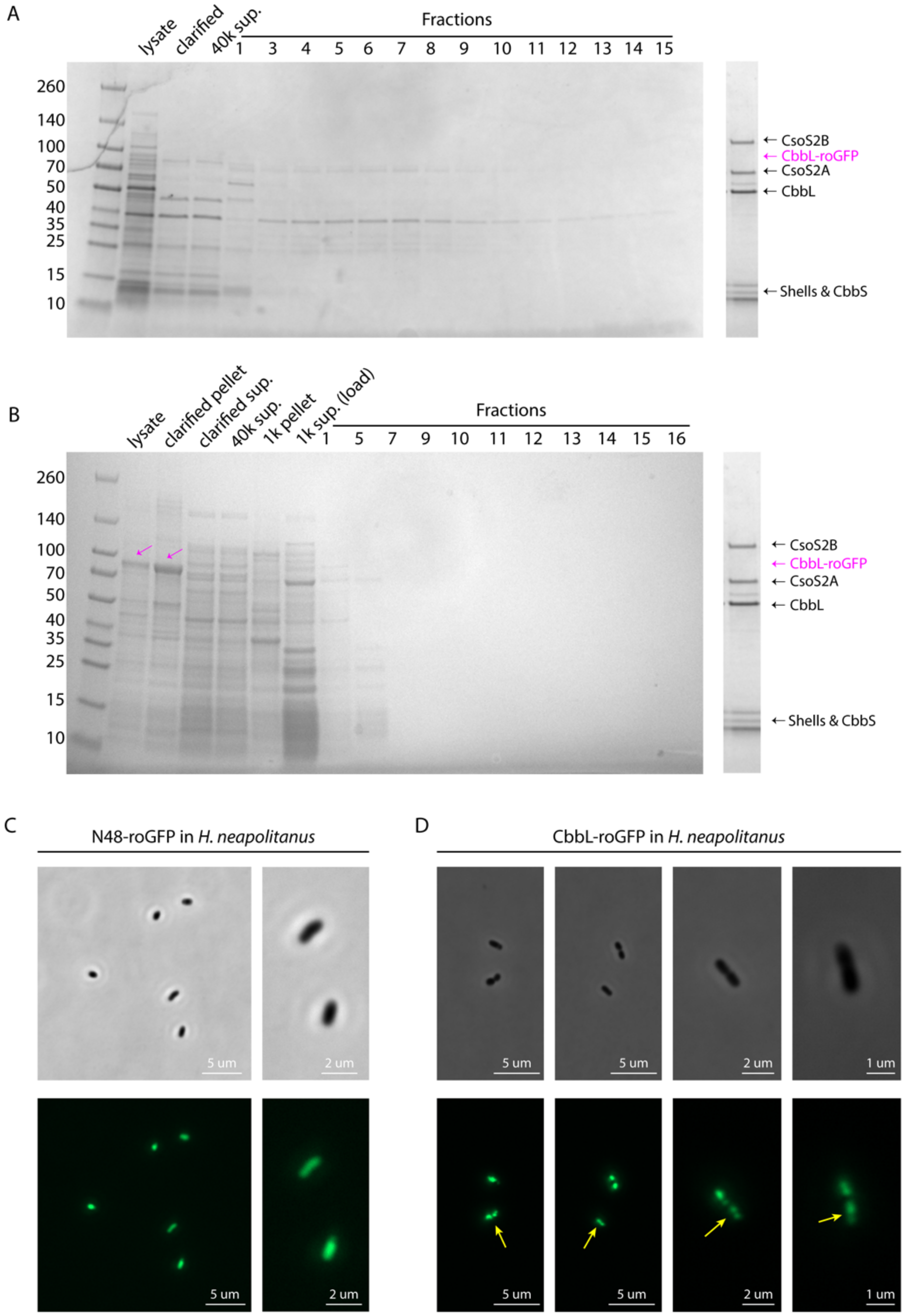


Figure S15. CbbL-roGFP carboxysome purifications were unsuccessful, yet microscopy in *H. neapolitanus* shows phenotypic variances with N48-roGFP. (A) and (B) show unsuccessful carboxysome purifications. The far-right lane shows an example of a successful carboxysome purification from *E. coli* not expressing CbbL-roGFP. (A) Carboxysomes with CbbL-roGFP purified from two *H. neapolitanus* pellets of ~0.4 g total. (B) Carboxysomes with CbbL-roGFP purified from two HnCB10 *E. coli* pellets of ~1 g total. (C) *H. neapolitanus* expressing N48-roGFP. roGFP signal appears more diffuse throughout the cell. (D) *H. neapolitanus* expressing CbbL-roGFP. roGFP signal appears more punctate, possibly indicating more precise carboxysomal localization and less unencapsulated protein.

References

1. Cohen, A. E.; Moerner, W. E., Controlling Brownian motion of single protein molecules and single fluorophores in aqueous buffer. *Opt. Express* **2008,** *16*, 6941-6956.
2. Squires, A. H.; Lavania, A. A.; Dahlberg, P. D.; Moerner, W. E., Interferometric scattering enables fluorescence-free electrokinetic trapping of single nanoparticles in free solution. *Nano Lett.* **2019,** *19*, 4112-4117.
3. Wang, Q. Enabling multivariate investigation of single-molecule dynamics in solution by counteracting Brownian motion. Stanford University, Stanford, CA, 2014.
4. Han, J. C.; Han, G. Y., A procedure for quantitative determination of tris(2-carboxyethyl)phosphine, an odorless reducing agent more stable and effective than dithiothreitol. *Anal. Biochem.* **1993,** *220*, 5-10.
5. Santarino, I. B.; Oliveira, S. C. B.; Oliveira-Brett, A. M., Protein reducing agents dithiothreitol and tris(2-carboxyethyl)phosphine anodic oxidation. *Electrochem. Comm.* **2012,** *23*, 114-117.
6. Kukura, P., Interferometric scattering microscopy. *Ann. Rev. Phys. Chem.* **2019,** *70*, 301-322.
7. Wang, Q.; Moerner, W. E., An adaptive Anti-Brownian Electrokinetic Trap with real-time information on single-molecule diffusivity and mobility. *ACS Nano* **2011,** *5*, 5792-5799.
8. Watkins, L. P.; Yang, H., Detection of intensity change points in time-resolved single-molecule measurements. *J. Phys. Chem. B* **2005,** *109*, 617-628.
9. Bonacci, W.; Teng, P. K.; Afonso, B.; Niederholtmeyer, H.; Grob, P.; Silver, P. A.; Savage, D. F., Modularity of a carbon-fixing protein organelle. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109*, 478.
10. Huang, F.; Hartwich, T. M. P.; Rivera-Molina, F. E.; Lin, Y.; Duim, W. C.; Long, J. J.; Uchil, P. D.; Myers, J. R.; Baird, M. A.; Mothes, W.; Davidson, M. W.; Toomre, D.; Bewersdorf, J., Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms. *Nat. Methods* **2013,** *10*, 653-658.
11. Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B., Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **2012,** *9* (7), 676-682.
12. Chattoraj, M.; King, B. A.; Bublitz, G. U.; Boxer, S. G., Ultra-fast excited state dynamics in green fluorescent protein: Multiple states and proton transfer. *Proc. Natl. Acad. Sci. U.S.A.* **1996,** *93*, 8362-8367.

A.3 Possible functional linkage between carboxysomal shell proteins and PII proteins: What's known, outstanding questions, and initial experiments; supplemental notes & figures

Supplementary Notes

Supplementary Figures

Possible functional linkage between carboxysomal shell proteins and PII proteins: What’s known, outstanding questions, and initial experiments

A.3.S1 Abstract

Hexameric α-carboxysomal shell proteins (Csos1 family) share significant structural homology to PII signaling proteins. PII signaling proteins help regulate carbon and nitrogen levels, and are ubiquitous throughout prokaryotes, nitrogen-fixing archaea, and plant chloroplasts. They respond to fluctuating ATP (or derivatives like cAMP) levels in cells by binding and often hydrolyzing ATP, which directs interaction with downstream effector proteins. In contrast, carboxysomal shell proteins form an icosahedral structure around enzymes that perform carbon fixation. Many questions about the role of shell proteins remain, such as whether or not they are selectively permeable to metabolites, and if so, how they maintain this selective permeability. Because the structural homology between PII proteins and Csos1 proteins is significant, this leads to the question of whether Csos1 proteins share functional similarities to PII proteins, such as the ability to bind and respond to small molecules or other proteins, or even to perform catalytic activity that may lead to a conformational change. In this white paper, I examine what is known about the structure, function, and relationship between these proteins, refine what the outstanding questions are, and do preliminary experiments to investigate this proposal.

A.3.S2 Introduction & Hypotheses

Peering into the evolution of carboxysome shell proteins may shed light on how they regulate metabolite permeability. Mart Krupovic and Eugene Koonin examined the evolutionary origins of carboxysome shell proteins in their 2017 paper ‘Cellular origin of the viral capsid-like bacterial microcompartments’ (1). As the title implies, pictures of bacterial microcompartments (BMCs) draw immediate comparisons to viruses, yet no one had examined if there was actually an evolutionary connection. In their paper, they discovered that BMC-H (hexagonal capsid monomers) and BMC-P (pentameric capsid monomers) have structural homology to cellular proteins, rather than any detectable homology to viral proteins. BMC-H had “multiple, highly significant” matches to the family of PII signaling proteins, which they note are “among the most ancient, ubiquitous and versatile components of signaling systems in nature”. BMC-P had hits against bacterial proteins with an oligonucleotide / oligosaccharide-binding (OB) fold, which they comment is “one of the most ancient and widespread structural folds found in a wide range of functionally diverse proteins.” It is easy to see the similarities in the structural figures of their paper.

PII proteins form one of the largest families of signaling proteins. They are found in all bacteria, all N_2 fixing archaea, and all chloroplasts of red algae and land plants where they help regulate carbon and nitrogen levels in the cell (2). In general, PII proteins are ATPases that are regulated by 2-oxoglutarate (2-OG) (3). When 2-OG levels drop, the PII hydrolyzes ATP to ADP leading to a conformational change that enables binding to a downstream receptor. Though this mechanism is true for many PII proteins, it is not universal, and other metabolites may be involved (4).

In contrast to hexameric BMC-H proteins, PII proteins form homotrimers. One key difference is the absence in BMC-H proteins of the PII ATP/ADP binding “T-loop” (1). The T-loop is a large, flexible, sometimes unstructured subdomain that faces outward from the homotrimer (2). In PII trimers, ATP and 2-OG bind at the interface between two subunits coordinated by the T-loop.

Is it possible that ATP, ADP, or 2-OG binds BMC shell proteins? There is loose evidence that ADP may bind the β-carboxysome BMC-H shell protein CcmP. CcmP is a trimeric stacked hexamer with open and closed conformations (5). In their crystal structure of CcmP, Cai et al. observed electron density in the space between the two fused BMC-H monomers (5). They tried to model 3PG into this density; it fit the open/closed conformation that they observed, and they note that all other components in the mother liquor were too small to explain the density. In the open/open conformation, they saw the same density, but because other 2-4 carbon molecules were in the buffer, it wasn’t possible to fit unambiguously. They did isothermal titration calorimetry on CcmP + 3PG and RuBP, but results were near the limit of detection and binding affinities were very low, so the data were inconclusive. In a separate study, Larsson et al. crystalized CcmP and observed density near the T-loop area (6). Model building into the density was inconclusive; it looked like it could fit an ADP molecule, but occupancy was low. Some of the residues around the density had dual conformations, suggesting one structure when bound to ATP and another when bound to ADP. The authors tried co-crystallization with ATP, ADP, 3PG, bicarbonate, and RuBP, but all were unsuccessful. They note that they used high concentrations of RuBP (50 mM) so it's unlikely that it binds, though it is possible that the crystal lattice precluded binding.

Attempts have been made to visualize other carboxysomal shell proteins bound to relevant metabolites, but none have succeeded thus far. Klein et al., which reports the crystal structure of the α-carboxysome trimeric BMC-H protein CsoS1D, mentions that they co-crystallized with metabolites (RuBP, bicarbonate) without conclusive results (7). They were unable to visualize the N-terminal 50 amino acids, which may help in regulating transport or binding to molecules. They also mention that they screened for enzymatic activities without any results, which is an interesting tidbit relevant to this white paper (see results section). Similarly, Tsai et al. in their paper on the crystal structure of major α-carboxysome BMC-H protein CsoS1A, say that they were unsuccessful in co-crystallizing with metabolites (3PG, RuBP) (8). However, they were able to visualize sulfate, a bicarbonate analog, in the CsoS1A pore, yet they note that they used a very high concentration of sulfate in the buffer (200 mM). Echoing these results, the crystal structures of CcmK1 and CcmK2, the dominant β-carboxysome BMC-H proteins, showed sulfate ions in pores as well as in the spaces between hexamers (9). One paper suggests that CcmK2 forms a stacked dodecamer (10), and related proteins CcmK3 and CcmK4 also form dodecamers (11). One interesting detail is that pH strongly affected this interaction – more dodecamers were observed at lower pHs, and the authors postulate that maybe these proteins have a shell capping function that modulates permeability.

Because Krupovic and Koonin's study evaluated the evolution of BMC-H folds in general, it is worth taking a look at other BMC shell proteins outside of carboxysomal proteins to understand what is known about metabolite binding. The ethanolamine utilizing (Eut) BMC is a class of BMCs that encapsulates enzymes necessary to metabolize ethanolamine. It is thought that the Eut compartment prevents the leakage of the toxic intermediate acetaldehyde, instead recycling it in the compartment lumen into other compounds. The Eut BMC has a permeability paradox: some luminal enzymes require large cofactors (adenosylcobalamin or vitamin B_{12}), as well as ATP, to do catalysis. Both of these molecules are substantially larger than ethanolamine or acetaldehyde, which are supposedly retained (12).

The major Eut BMC-H shell proteins are EutM, EutS, EutL, and EutK. EutM crystallized with a sulfate ion in the pore, though it is unclear if this is physiologically relevant (13). EutS was crystallized with bound glycerol, though it was found not in the central pore but between BMC-H subunits (13). EutL is a trimer of dimers similar to CcmP and CsoS1D with an open/closed structure. In the EutL structure, ethanolamine was observed bound between two domains of a single subunit, rather than in the pore, and this binding was observed in the closed conformation (12). Isothermal titration calorimetry showed a large ΔH when EutL was bound to ethanolamine, and fit a model that describes two sites per monomer (12). In the paper, Thompson et al., hypothesizes negative allosteric regulation; when ethanolamine levels are high, they bind EutL and close the shell, preventing leakage of toxic acetaldehyde. When ethanolamine levels are low, EutL opens and allows passage and replenishment of large cofactors like cobalamin, while risk of toxic leakage is low. Lastly, EutK did not form hexamers in solution, but its C-terminus has an extra 60-amino acid extension that has a helix-turn-helix motif, which is well known to bind to nucleic acids (RMSD of <1 with other nucleic acid binding motifs) (14).

It's also worth mentioning the Eut pentamer, EutN. It's a bit of a paradox because some studies crystallized it as a hexamer (15), yet Nicole Wheatley in her thesis found it to be a pentamer in its soluble form (16). pH may be responsible: EutN crystallized as a hexamer at pH 5.5, and as a pentamer at pH 7.6-8.6.

So, how does all of this relate to carboxysome permeability? Are shell proteins passive structures, maintaining a permeability barrier from physical characteristics such as charge and pore size? Or are they flexible modulators of permeability, sharing properties of PII proteins such as binding small molecule effectors? It might be possible that metabolites are binding in unexpected places, such as in the spaces between shells or in uncrystallized, unstructured tails. Or, that properties such as pH, which changes from day to night in the carboxysome lumen (17, 18), lead to conformational changes. Is it even possible that carboxysome shell proteins display the most basic feature of a PII protein, ATPase activity, that simply no one has tested for?

Here I present preliminary data that attempts to answer these questions. I purified carboxysome shell proteins, and assayed their ATPase activity. I think it helps to know the answer before you read further: the shell proteins ultimately did not show ATPase activity. The data presented here walk you through some of the preliminary experiments I did, starting with those first ones where there appeared to be activity, and ending with more thorough controls where activity disappeared. I'm presenting this work (buried in the back of this thesis) in case someone has the same hypotheses I had; now they can know that I tried these experiments, and that the results were negative. I decided to show the initial "positive" results to be open about the scientific process that occurred throughout this project, and how careful analysis and controls can change the interpretation of the data.

I don't think the negative results presented here necessarily invalidate the hypotheses presented above. They simply show that these shell proteins did not have ATPase activity under the conditions tested.

A.3.S3 Results

To test for ATPase activity, I used a malachite green assay, which directly measures the amount of phosphate released during ATP dephosphorylation. Malachite green and molybdate form a green complex when bound to free orthophosphate, which can be easily measured on a spectrophotometer or plate reader at 620 nm. An example of what this assay looks like can be seen in Figure 1A: the well is yellow at 0 uM ATP (top left) and green when free phosphate is added as a control (top right). Higher concentrations of ATP show background levels of free phosphate (bottom left). Looking by eye, while BSA activity levels mirrored adding no protein at all, as would be expected since BSA is not an enzyme, shell proteins such as CsoS1A showed above-background levels of phosphate the higher the ATP concentration.

In a first-pass test to see if shell proteins showed ATPase activity, I ran four different shell proteins in the assay: SUMO-CsoS4B, SUMO-CsoS1D, SUMO-CsoS1B, and SUMO-CsoS1A. The shell proteins had been purified with a cleavable SUMO tag, and I did not cleave it off prior to this assay since it was just an initial test. Three out of four shells showed some level of activity, with SUMO-CsoS1D showing the most activity (Figure 1B). SUMO-CsoS1B and SUMO-CsoS4B showed activity barely above the BSA background, and SUMO-CsoS1A showed almost no activity. SUMO-CsoS1A has no tryptophan residues, so I thought that maybe it showed no activity because I did not have a correct concentration, and had simply not added enough. I switched to using strep-SUMO-CsoS1A, which has a W residue in the strep tag and could be more accurately measured. The strep tagged protein showed much higher activity, comparable to SUMO-CsoS1D (Figure 1C). To test if ADP acts as an inhibitor of activity, as it does in PII proteins, I added ADP into the assay. I observed a marked difference between the +ADP and -ADP samples: addition of ADP appeared to be blocking ATPase activity in strep-SUMO-CsoS1A (Figure 1D). To test for single vs. multiple turnover, I ran the assay for 1 hour with 50 uM ATP, taking timepoints every 10 minutes. Activity decreased by approximately half every 20 minutes, suggesting single turnover activity, though probably even lower concentrations of ATP should have been used to be sub-saturating (Figure 1E). To test the contribution of the SUMO tag to activity, I purified strep-CsoS1A and strep-CsoS1B. In contrast to previous results, the newly purified proteins had negligible ATPase activity, suggesting that the SUMO tag had somehow been contributing to background phosphate release (even though SUMO is not an ATPase) (Figure 1F). This was an indication that perhaps the tags on my proteins were affecting the outcomes of the results, or that different purifications were affecting the results by potentially bringing along cellular ATPases.

To control for this, I re-evaluated my purification of SUMO-CsoS1D (Figure 2A). I had stored elutions 2 and 3 separately; it is clear that elution 2 is quite dirty, while elution 3 is much cleaner (or just simply diluted out). I ran part of this purification on FPLC size exclusion, separating out most of the higher and lower bands (Figure 2B). I ran all three samples (elution 2, elution 3, and the size exclusion purified sample) in the ATPase assay. The cleaner the protein, the lower the activity, suggesting that I had been measuring activity from a hitchhiker cellular ATPase (Figure 2C). Another clue that I had been measuring false positive activity was that the specific activity of the shells, such as strep-SUMO-S1A, was far below that of a “typical” ATPase, such as Cin8 kinesin from Saccharomyces cerevisiae. Graphing the two together, the difference in activity is obvious (Figure 2D).

With that, I closed the book on shells being ATPases. Though in a twist ending, I ran purified CsoS2 in the assay just for fun. To my surprise, it showed activity above background (Figure 2E). Oh S2, you’ve done it again…

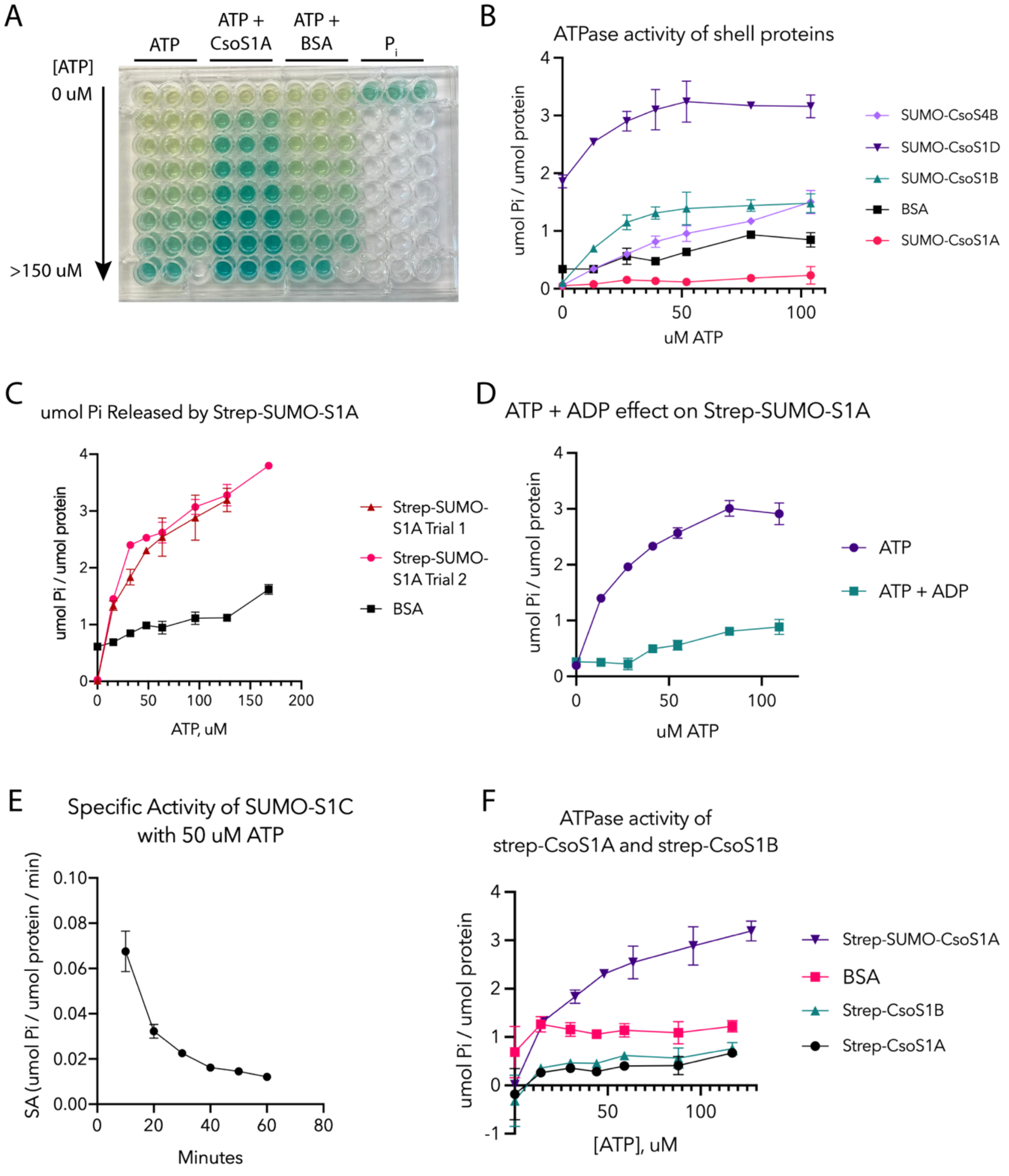


Figure S1. Early results suggest shell ATPase activity. (A) Example of the plate setup and colorimetric results of the assay. (B) ATPase activity of the indicated shell proteins, with BSA as a control. (C) ATPase activity of two independent trials of strep-SUMO-CsoS1A. (D) ATPase activity of strep-SUMO-CsoS1A with and without added 50 uM ADP. (E) Specific activity of SUMO-S1C with 50 uM ATP. (F) ATPase activity of CsoS1A and B without the SUMO tag.

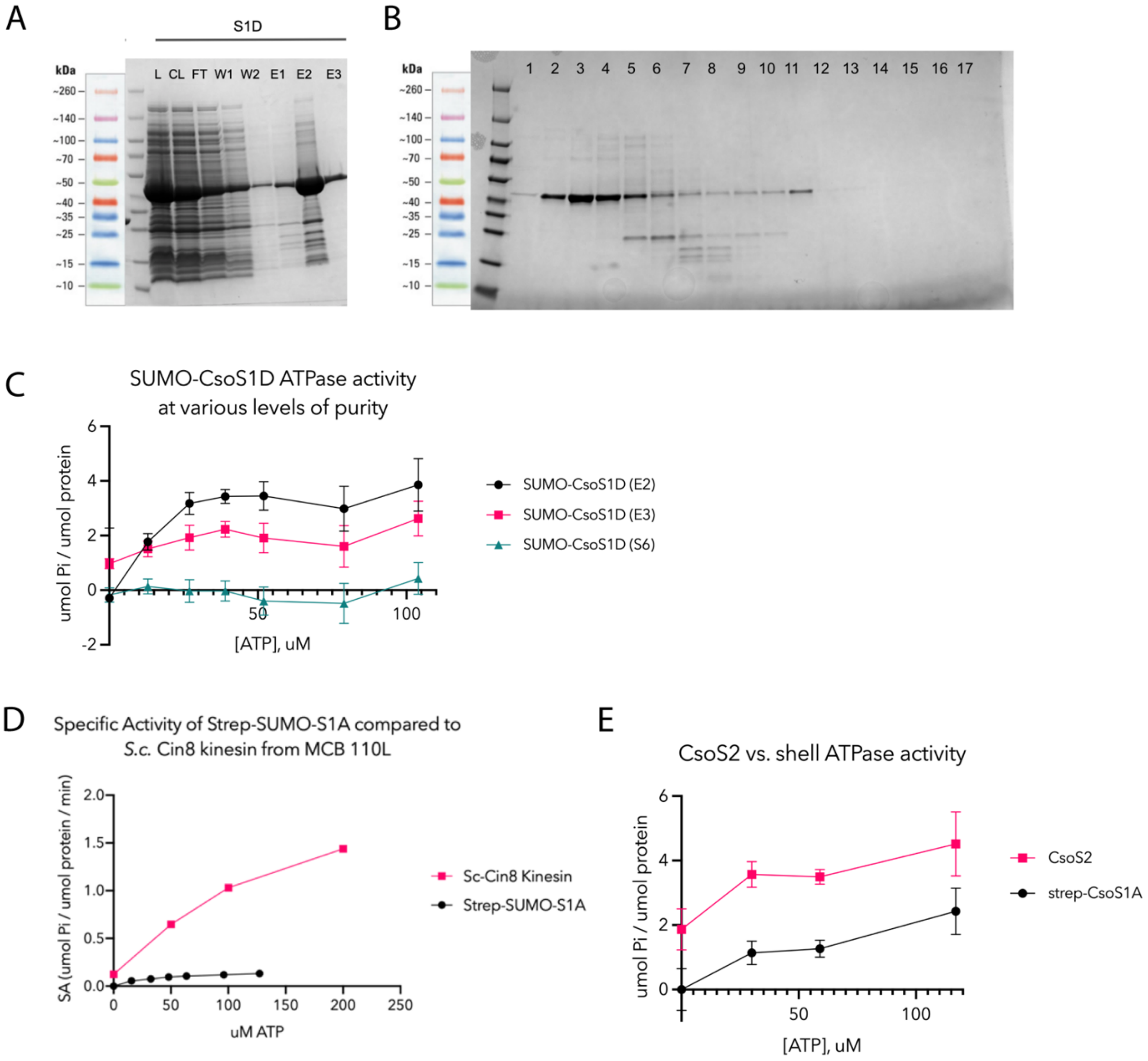


Figure S2. Pure proteins show no ATPase activity. (A) Nickel resin purification of SUMO-CsoS1D. L, lysate, CL, clarified lysate, FT, flow through, W, wash, E, elution. (B) Fractions from the FPLC size exclusion cleanup of SUMO-CsoS1D. (C) ATPase activity of SUMO-CsoS1D from elution 2 (E2), elution 3 (E3), and size exclusion (S6). (D) Specific activity of Strep-SUMO-CsoS1A compared to Cin8 kinesin from *S. cerevisiae*, calculated in a separate assay. (E) ATPase activity of CsoS2 and strep-CsoS1A.

A.3.S4 Conclusion

Though the results in these tests were negative, they don't necessarily invalidate the hypotheses stated above. A first control would be to test if SUMO-shell constructs, or strep-shell constructs, are forming hexamers, since oligomerization may be important for activity. The concentration of protein used in these assays was pretty low, and probably sometimes inaccurate due to the lack of W on some proteins, so maybe more protein could have led to more interesting results. Certainly protein purity is a big variable that needs to be controlled for; cellular ATPases may be responsible for a lot of the measured activity. Having a control purification from *E. coli* not expressing shell proteins may be a good way to measure background. It may be that ATPase activity only occurs in the presence of other untested metabolites, such as 2-OG or RuBP. ITC is another nice parallel approach, where an array of metabolites could be tested for binding. So though these experiments ended here for me, there may still be interesting paths forward that would address these questions.

A.3.S5 Materials & Methods

Purifications

SUMO-shell

All SUMO-shell constructs have an N-terminal his tag. 2L of each SUMO-shell were grown in the standard way, with induction at log phase and overnight growth at 18C. 2L were combined into a single pellet and frozen at -20C. Pellets were thawed in 20 ml of lysis buffer (50 mM Tris, 300 mM NaCl, 20 mM imidazole, + lysozyme, + benzonase, + PMSF, pH 7.5). Cells were lysed on an Avestin emulsiflex homogenizer. Cells were spun at 15000 rpm for 1 hour to clarify. Clarified lysate was loaded onto a Ni-Sepharose fast-flow resin equilibrated with binding buffer. Resin was washed with 2x 25 ml wash buffer (TBS 50/300 + 60 mM imidazole). Protein was eluted with elution buffer (TBS 50/300 + 300 mM imidazole) in 3 fractions: 2 ml for E1, 15 ml for E2, and 15 ml for E3. All proteins purified except for CsoS4A. Fractions E2 and E3 were concentrated separately with a 10K cutoff spin concentrator to 3 ml. It was observed that putting the concentrated protein on ice turned it “milky”, while everything stayed in solution at room temperature. Concentrated proteins was desalted into 50 mM Tris, 150 mM NaCl, pH 7.5. Because the E2 fractions were turning “milky”, salt was added up to 300 mM, which returned it to a clear solution. Some precipitate was spun out at 5 minutes at 4k xG. Protein was frozen in 10% glycerol at -80C. SUMO-CsoS1D was later cleaned up on an AKTA FPLC S6 10/300 size exclusion column.

Strep-shell

Strep-shell was lysed and clarified the same way as above in 50 mM HEPES, 150 mM NaCl lysis buffer, then purified on a Strep-Tactin resin on a gravity column. Clarified lysate was loaded onto the column and allowed to flow through, followed by a wash step and elution with elution buffer (50 mM HEPES, 150 mM NaCl, 2.5 mM D-Desthiobiotin) before adding 10% glycerol, flash freezing in liquid N_2, and storing at -80°C.

ATPase assays

All ATPase assays were performed with the ‘Malachite Green ATPase Assay’ kit from Sigma, #MAK307. All measurements were performed on an M1000 Tecan plate reader. In general, the buffer was 150 mM NaCl, 50 mM Tris, 10 mM $MgCl_2$, unless otherwise noted. Each sample was 80 ul, and all assays were done in 96-well plates in triplicate. A defined volume of shell protein was added to start the reaction, and the reaction was allowed to proceed for 10 minutes. Meanwhile, the working reagent was prepared according to the kit instructions: 1% of Reagent B was added to Reagent A (so, 10 ul B into 1000 of A), and 20 ul of reagent was added to each well. Color maturation occurred over 30 minutes, and then was measured at A620.

Specifics for each assay:

Note: I was optimizing the assay as I was testing different constructs, so some conditions vary between assays. All comparisons should be evaluated with these assay differences in mind. Concentration difference should in principle not matter, since all measurements are per umol of protein.

For Figure 1B, 20 uM of protein was used.

For Figure 1C, the buffer had 50 mM HEPES instead of Tris, and 6.4 uM protein was used. The reaction proceeded for 25 minutes instead of 10.

For Figure 1D, the buffer had 50 mM HEPES instead of Tris and 50 mM HCO_3-. 6.4 uM protein was used. 50 uM ADP was used. The reaction proceeded for 20 minutes instead of 10.

For Figure 1E, 6.4 uM protein was used.

For Figure 1F, the buffer had 50 mM HEPES instead of Tris, and 8 uM protein was used. The reaction proceeded for 16 minutes instead of 10. The strep-SUMO-CsoS1A trace was added in from a separate previous experiment.

For Figure 2C, 1 uM protein was used.

For Figure 2E, the buffer had 50 mM HEPES instead of Tris, and 1 uM protein was used.

References

1. Krupovic, M., and Koonin, E. V. (2017) Cellular origin of the viral capsid-like bacterial microcompartments. *Biol. Direct*. **12**, 25
2. Forchhammer, K., and Lüddecke, J. (2016) Sensory properties of the PII signalling protein family. *FEBS J.* **283**, 425–437
3. Radchenko, M. V., Thornton, J., and Merrick, M. (2013) P(II) signal transduction proteins are ATPases whose activity is regulated by 2-oxoglutarate. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 12948–12953
4. Zeth, K., Fokina, O., and Forchhammer, K. (2014) Structural basis and target-specific modulation of ADP sensing by the Synechococcus elongatus PII signaling protein. *J. Biol. Chem.* **289**, 8960–8972
5. Cai, F., Sutter, M., Cameron, J. C., Stanley, D. N., Kinney, J. N., and Kerfeld, C. A. (2013) The structure of CcmP, a tandem bacterial microcompartment domain protein from the \beta-carboxysome, forms a subcompartment within a microcompartment. *J. Biol. Chem.* **288**, 16055–16063
6. Larsson, A. M., Hasse, D., Valegå, K., and Andersson, I. (2017) Crystal structures of \beta-carboxysome shell protein CcmP: ligand binding correlates with the closed or open central pore. *J. Exp. Bot.* **68**, 3857–3867
7. Klein, M. G., Zwart, P., Bagby, S. C., Cai, F., Chisholm, S. W., Heinhorst, S., Cannon, G. C., and Kerfeld, C. A. (2009) Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. *J. Mol. Biol.* **392**, 319–333
8. Tsai, Y., Sawaya, M. R., Cannon, G. C., Cai, F., Williams, E. B., Heinhorst, S., Kerfeld, C. A., and Yeates, T. O. (2007) Structural analysis of CsoS1A and the protein shell of the Halothiobacillus neapolitanus carboxysome. *PLoS Biol.* **5**, e144
9. Tanaka, S., Sawaya, M. R., Phillips, M., and Yeates, T. O. (2009) Insights from multiple structures of the shell proteins from the beta-carboxysome. *Protein Sci.* **18**, 108–120
10. Samborska, B., and Kimber, M. S. (2012) A dodecameric CcmK2 structure suggests \beta-carboxysomal shell facets have a double-layered organization. *Structure*. **20**, 1353–1362
11. Sommer, M., Sutter, M., Gupta, S., Kirst, H., Turmo, A., Lechno-Yossef, S., Burton, R. L., Saechao, C., Sloan, N. B., Cheng, X., Chan, L.-J. G., Petzold, C. J., Fuentes-Cabrera, M., Ralston, C. Y., and Kerfeld, C. A. (2019) Heterohexamers formed by ccmk3 and ccmk4 increase the complexity of beta carboxysome shells. *Plant Physiol.* **179**, 156–167
12. Thompson, M. C., Cascio, D., Leibly, D. J., and Yeates, T. O. (2015) An allosteric model for control of pore opening by substrate binding in the EutL microcompartment shell protein. *Protein Sci.* **24**, 956–975
13. Pitts, A. C., Tuck, L. R., Faulds-Pain, A., Lewis, R. J., and Marles-Wright, J. (2012) Structural insight into the Clostridium difficile ethanolamine utilisation microcompartment. *PLoS One*. **7**, e48360
14. Tanaka, S., Sawaya, M. R., and Yeates, T. O. (2010) Structure and mechanisms of a protein-based organelle in Escherichia coli. *Science*. **327**, 81–84

15. Tanaka, S., Kerfeld, C. A., Sawaya, M. R., Cai, F., Heinhorst, S., Cannon, G. C., and Yeates, T. O. (2008) Atomic-level models of the bacterial carboxysome shell. *Science*. **319**, 1083–1086
16. Wheatley, N. M. (2014) *Elucidation of genes of unknown function in alpha carboxysome operons: acRAF, BMVs, and carbon PII regulatory proteins*. Ph.D. thesis, University of California, Los Angeles, Ann Arbor, United States
17. Mangan, N. M., Flamholz, A., Hood, R. D., Milo, R., and Savage, D. F. (2016) pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5354-62
18. Long, B. M., Förster, B., Pulsford, S. B., Price, G. D., and Badger, M. R. (2021) Rubisco proton production can drive the elevation of CO2 within condensates and carboxysomes. *Proc. Natl. Acad. Sci. U. S. A.* 10.1073/pnas.2014406118