

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Structural and functional
studies of β -carboxysomal proteins:
CcmM and Rubisco activase

Huping Wang
aus
Shanghai, China

2021

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. F. Ulrich Hartl betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 23.07.2021

Huping Wang

| | |
|------------------------------|----------------------------------|
| Dissertation eingereicht am: | <u>04.08.2021</u> |
| 1. Gutachter: | <u>Prof. Dr. F. Ulrich Hartl</u> |
| 2. Gutachter: | <u>Prof. Dr. Roland Beckmann</u> |
| Mündliche Prüfung am | <u>03.09.2021</u> |

Acknowledgements

I would like to thank Dr. Manajit Hayer-Hartl and Prof. Dr. F. Ulrich Hartl for supervising the projects and giving me the opportunity to work in their international and interdisciplinary department at the Max-Planck Institute of Biochemistry. They constantly provided me with valuable advice and critical suggestions. Their way of approaching complex projects has shaped my view for the future on how to perform science.

I am particularly thankful to my close collaborators, Dr. Andreas Bracher, Dr. Xiao Yan, Mirkko Flecken and Dr. Kun Zang. Dr. Andreas Bracher supervised structural analysis. He is also a very nice friend who is always helpful on a day to day basis. I am grateful to Dr. Xiao Yan for his contributions to the Rubisco-M35 project; Mirkko Flecken for his contributions to the *NosRca* project and Dr. Kun Zang for his contributions to the Rubisco-M58 project. Working together with them was a pleasure. The discussions with them always enlightened me to think more about the projects.

I would also like to thank the technical staff of the department, Emmanuel Burghardt, Darija Pompino, Nadine Wischnewski, Romy Lange, Silvia Gärtner and Albert Ries for organizational matters, and all former and present colleagues for creating a helpful and open-minded working atmosphere.

I especially thank Angelika Gordon of the MPIB international office, and my colleagues, Dr. Rongqin Li, Dr. Peng Xu, Dr. Kun Zang, Darija Pompino, Victoria Trinkaus and Dr. Andreas Bracher for their support during periods of difficult time in a foreign country.

I also appreciate the competent technical support of the staff of the MPIB cryo-EM facility, Dr. Daniel Bollschweiler, Dr. Tillman Schäfer and Dr. Mike Strauss and Dr. Jesuraj Rajan Prabu for his maintenance and development of the computational infrastructure.

I especially thank Dr. Manajit Hayer-Hartl, Dr. Andreas Bracher and the members of my PhD committee Prof. F. Ulrich Hartl, Prof. Roland Beckmann, Prof. Karl-Peter Hopfner, Prof. Julian Stingele, Dr. Bettina Bölter and Prof. Ivana Ivanović-Burmazović for the critical evaluation of this thesis.

Finally, I am grateful to my parents Ming Wang and Ying Hu and my fiancé Yu Bao, who have supported and encouraged me over the years.

Table of Contents

| | | |
|-------|--|----|
| 1 | Summary | 1 |
| 2 | Introduction | 4 |
| 2.1 | Photosynthesis | 4 |
| 2.2 | Rubisco..... | 7 |
| 2.2.1 | The different forms of Rubisco | 8 |
| 2.2.2 | Structural basis for Rubisco catalysis..... | 11 |
| 2.2.3 | Carboxylation and oxygenation (Photorespiration) | 13 |
| 2.2.4 | Rubisco inhibition | 16 |
| 2.3 | CO ₂ concentrating mechanisms | 19 |
| 2.3.1 | Carboxysome..... | 20 |
| 2.3.2 | Pyrenoid | 26 |
| 2.3.3 | C4 and crassulacean acid metabolism photosynthesis | 27 |
| 2.4 | Liquid-like phase separation | 28 |
| 2.5 | AAA+ ATPases..... | 29 |
| 2.6 | Rubisco activase (Rca)..... | 33 |
| 2.6.1 | Rca of red-type Rubisco | 34 |
| 2.6.2 | Rca of green-type Rubisco form IA | 35 |
| 2.6.3 | Rca of green-type Rubisco form IB | 39 |
| 3 | Aim of this study | 42 |
| 4 | Publications | 43 |

| | | |
|-----|---|-----|
| 4.1 | Rubisco condensate formation by CcmM in β -carboxysome biogenesis | 43 |
| 4.2 | Dual functions of a Rubisco activase in metabolic repair and recruitment to carboxysomes | 67 |
| 4.3 | Scaffolding protein CcmM directs multiprotein phase separation in β -carboxysome biogenesis (in revision) | 106 |
| 4.4 | Phase separation of Rubisco by folded SSUL modules of CcmM in β -carboxysome biogenesis (accepted) | 184 |
| 4.5 | Structure and conformational cycle of a bacteriophage-encoded chaperonin..... | 224 |
| 5 | Discussion | 252 |
| 5.1 | The requirement for RbcS-containing Rubisco holoenzyme in CCM microcompartment biogenesis..... | 253 |
| 5.2 | Phase separation in the CCM microcompartment..... | 255 |
| 5.3 | The function of the SSUL modules in β -carboxysome biogenesis..... | 255 |
| 5.4 | The central role of M58 and M35 in β -carboxysome biogenesis | 256 |
| 5.5 | Form IB Rubisco reactivation | 259 |
| 6 | References | 262 |

1 Summary

Photosynthesis is the fundamental process in plants, algae and cyanobacteria by which most of the bio-mass on Earth is produced. It uses the energy from light to fix atmospheric carbon dioxide (CO₂) into organic compounds. The enzyme catalyzing the carbon fixation reaction is ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco). In plants and cyanobacteria, Rubisco belongs to form IB and is a ~530 kDa complex, consisting of eight large (RbcL) and eight small (RbcS) subunits. Despite its pivotal role, Rubisco is a rather inefficient enzyme, with low reaction turnover rate and poor CO₂/O₂ substrate selectivity. Furthermore, Rubisco is an enzyme which can be inhibited easily. Thus, mechanisms exist in nature to increase the efficiency of carbon fixation by Rubisco and to repair the inhibited Rubisco.

To improve carbon fixation by Rubisco and avoid the unproductive reaction with oxygen, photosynthetic organisms have evolved different types of CO₂ concentration mechanisms (CCM). In cyanobacteria, CCM consists of a microcompartment named carboxysome, which contains the enzymes Rubisco and carbonic anhydrase (CA). The formation of β -carboxysome requires the scaffold protein CcmM. CcmM has two isoforms and both isoforms are essential for β -carboxysome biogenesis. In the cyanobacterium *Synechococcus elongatus* PCC 7942 (*Se7942*), the long isoform, called M58, contains an N-terminal catalytically inactive γ -carbonic anhydrase like (γ -CAL) domain followed by a C-terminal domain containing three repeats of Rubisco small-subunit-like (SSUL) modules, joined by flexible linkers. M58 is trimeric via its γ -CAL domains. The short isoform, M35, consists only of the three SSUL modules. The functional CA in β -carboxysome is the β -CA, CcaA. The interaction of the scaffolding protein CcmM with Rubisco and CcaA is believed

to be the first step in the formation of the pre-carboxysome during β -carboxysome biogenesis.

To reactivate inhibited Rubisco, many photosynthetic organisms contain the ATPase associated with various cellular activities (AAA+) chaperone, Rubisco activase (Rca). Similar to many AAA+ proteins, the Rca hexamer functions by structurally remodeling the inhibited Rubisco in an ATP dependent manner. In β -carboxysomes, the Rca protein contains an additional C-terminal SSUL module.

One aspect of this thesis focuses on investigating how the scaffolding protein CcmM mediates multi-protein condensate formation during pre-carboxysome formation. We found that M35 and M58 interact with Rubisco via the SSUL modules, linking Rubisco molecules by multivalent interactions which resulted in the formation of a protein condensate. Trimeric M58, with a total of 9 SSUL modules has a higher affinity for Rubisco than M35, which contains only three SSUL modules. We solved the M35-Rubisco complex at 2.8 Å resolution by cryo-electron microscopy (cryo-EM). The structure reveals that the SSUL module binds asymmetrically in a pocket between the RbcL dimers and the adjacent RbcS subunits, making salt-bridges with both the RbcL and RbcS subunits. This interaction ensures that only the Rubisco holoenzyme is recruited into the carboxysome. Interestingly, the cyanobacterial Rca also interacts with Rubisco via its C-terminal SSUL modules, mediating recruitment into the pre-carboxysome.

Our subsequent analysis of the role of M58 in pre-carboxysome biogenesis revealed a remarkable complexity of interactions. We found that the N-terminal γ -CAL domains of M58 interact with the C-terminal tails of CcaA. Furthermore, the SSUL modules of M58 not only interact with Rubisco but also participate in intermolecular interactions with the γ -CAL domains, which induces liquid like phase separation (LLPS) of M58 on its own. Thus, M58

simultaneously recruits both Rubisco and CcaA into the pre-carboxysome core. Moreover, the γ -CAL trimers engage in head-to-head interactions. Based on this versatility of multivalent interactions, M58 functions as the central scaffolding protein in condensate formation during β -carboxysome biogenesis.

Another focus of this thesis is on the mechanism of form IB Rubisco reactivation by Rca. A cryo-EM structure of the *Nostoc* sp. PCC 7120 Rubisco in complex with its cognate Rca hexamer was solved at 2.9 Å resolution. In the structure, Rca is positioned over the Rubisco catalytic site under repair. Importantly, structural remodeling occurs precisely around the catalytic site. The N-terminal tail of RbcL binds into the central pore of the Rca hexamer, and simultaneously the C-terminal tail of the adjacent RbcL subunit is displaced from its position over the active site pocket. The structural remodeling of both the N- and C-terminal tails by ATP hydrolysis, opens the catalytic site pocket, allowing the release of the inhibitory sugar molecule. This mechanism of Rubisco remodeling is also conserved in plants.

In summary, this thesis provides insight into the mechanisms underlying β -carboxysome assembly and form IB Rubisco reactivation. Our findings may help in engineering crop plants with improved photosynthetic performance to meet the food demand of the growing population.

2 Introduction

2.1 Photosynthesis

Photosynthesis is a process used by plants, algae and cyanobacteria to synthesize organic compounds directly from carbon dioxide and an electron donor, using solar energy. In oxygenic photosynthesis, water is the electron donor and the products are a six-carbon sugar molecule and oxygen. This fundamental process produces essentially all food source for life on Earth and maintains the oxygen level in our present ecosphere (Fig. 1).

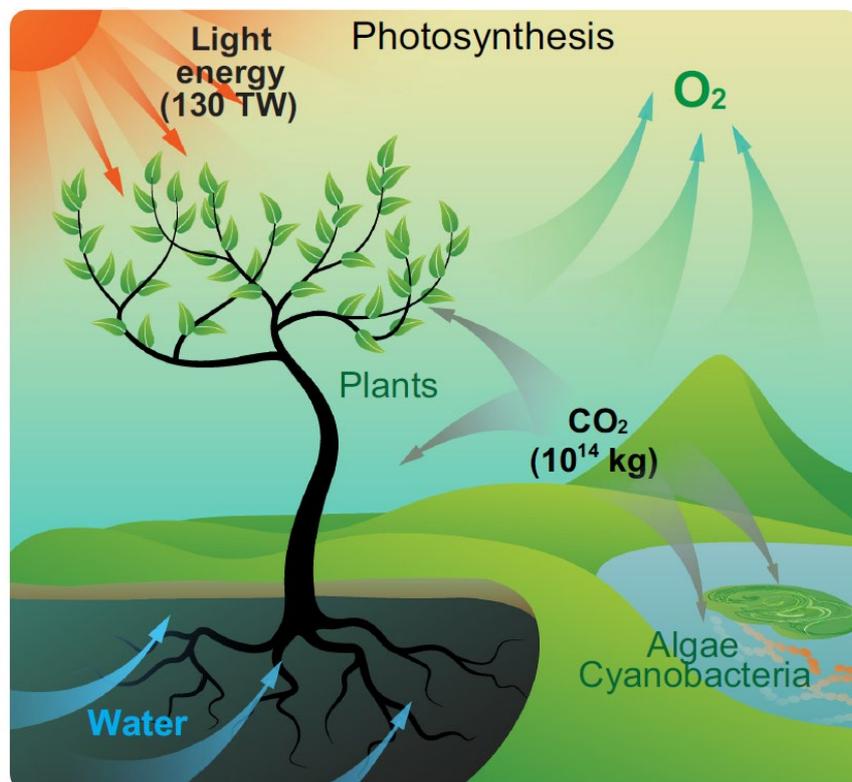


Figure 1: Schematic of oxygenic photosynthesis. Plants, algae and cyanobacteria use the energy from light to fix CO₂ from the atmosphere and release oxygen. Oxygen is a product of the light dependent reaction of the splitting of water. Approximately 130 terawatts of light energy are captured globally by photosynthesis and $\sim 10^{14}$ kg of atmospheric CO₂ are converted into biomass per year. Figure adapted from (Hayer-Hartl and Hartl, 2020).

Photosynthesis occurs in two stages, light-dependent reactions and light-independent reactions. During the light-dependent reactions, solar energy is converted into chemical energy in the form of NADPH and ATP. The oxygenic electron transport reactions of plant

and cyanobacteria take place in thylakoid membranes via two photosystems, PSI and PSII (Fig. 2). Photosystems are protein complexes with pigments (chlorophyll) as cofactors that are able to absorb photons and transfer electrons. The reaction center of PSI absorbs light at a wavelength maximum of 700 nm and PSII at 680 nm, accordingly are named P700 and P680. The first step in the oxygenic electron transport light reaction is light absorption in PSII; electrons in its reaction-center are excited (P680⁺) and trapped by the primary electron acceptors. P680⁺ is a strong oxidant and gains electrons from water via the H₂O splitting reaction, which eventually produces dioxygen. The electrons are transferred in several steps and then reach and excite PSI. The electrons derived from P700 are eventually used to produce NADPH. As a result of the electron transfer chain, a chemiosmotic potential across the thylakoid membrane is generated, which is used by ATP synthase to generate ATP (Duysens et al., 1961; Hill and Bendall, 1960; JAGENDORF, 1967; McCarty, 1966; Mitchell, 1961). The overall equation for the oxygenic light-dependent reaction in green plants and cyanobacteria can be written as follows:

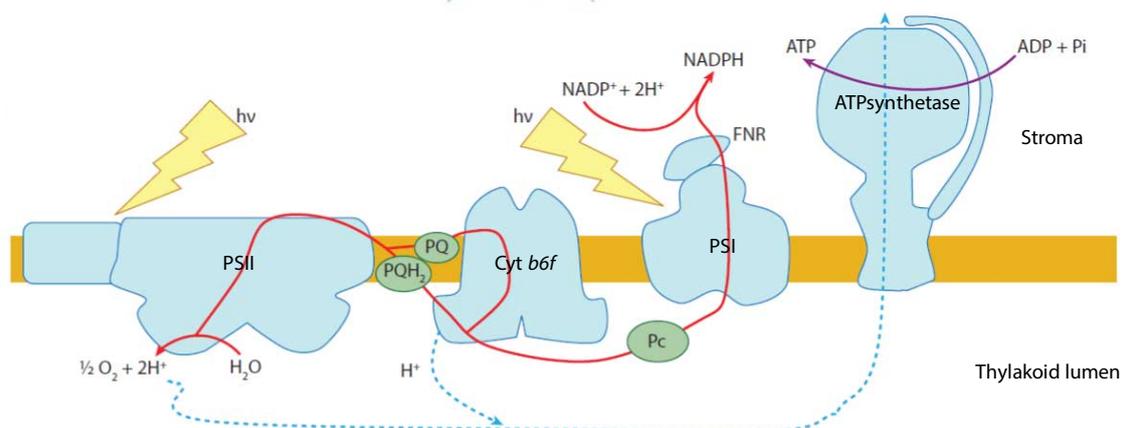
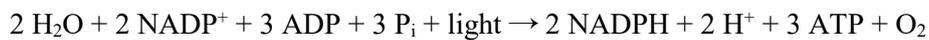


Figure 2: Schematic of the pathway for oxygenic photosynthetic linear electron flow and proton transport through the major protein complexes in the thylakoid membranes. Electrons are extracted from water on the luminal side of the membrane. Electron transfer is driven by the photosynthetic reaction center from PSII and PSI. The intersystem electron carriers consist of a pool of plastoquinone molecules (PQ and PQH₂) soluble within the lipid bilayer,

a transmembrane protein complex (Cyt b_6f), and a small, soluble copper-containing protein, plastocyanin (Pc), in the thylakoid lumen. Protons translocated across the membrane during the linear electron flow are used by the transmembrane ATP synthase to drive ATP synthesis. Figure adapted from (Eberhard et al., 2008).

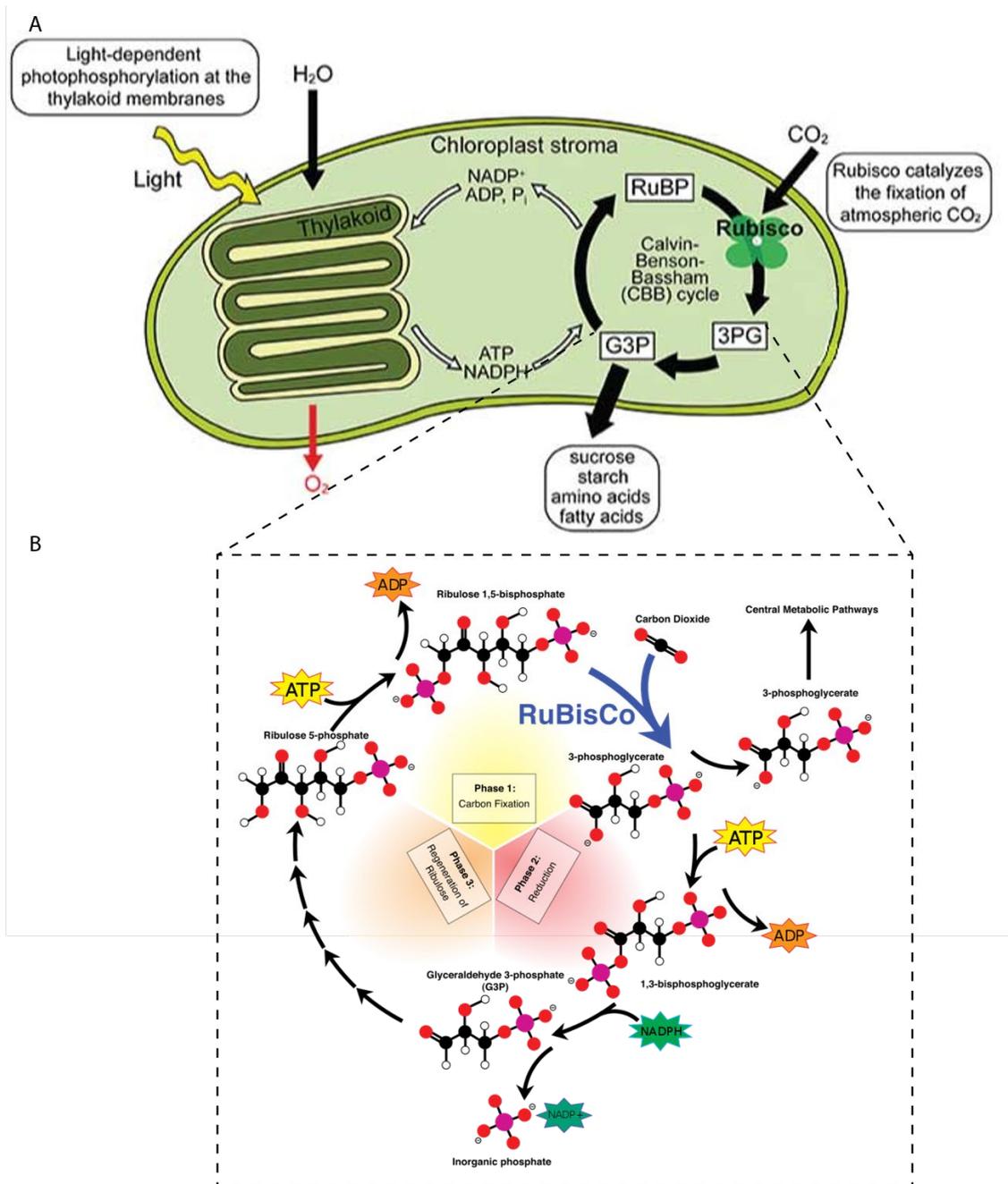
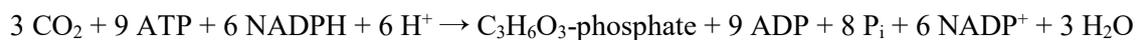


Figure 3: The Calvin-Benson-Bassham (CBB) cycle. (A) Schematic representation of the chloroplast showing the intricate connection of the light and dark reactions. (B) The CBB cycle can be divided into three different phases. In phase 1, Rubisco catalyzes the carboxylation of RuBP generating two molecules of 3PGA. In phase 2, the resulting 3PGA is reduced to form G3P, consuming ATP and NADPH produced in the light-dependent reaction. In phase 3, RuBP is regenerated. The regeneration requires ATP and involves a series of reactions. The intermediates of sugars and enzymes involved are not shown here. Figure modified from (Bhat et al., 2017b) and Wikipedia.

The light-independent reactions are also known as the dark reactions, carbon fixation or the Calvin-Benson-Bassham (CBB) cycle (Fig. 3). Here, the enzyme Rubisco catalyzes a reaction with CO₂ from the atmosphere and a five-carbon sugar, ribulose 1,5-bisphosphate (RuBP), to yield two molecules of a three-carbon compound, 3-phosphoglycerate (3PGA). 3PGA is further processed in the presence of ATP and NADPH, the products of the light-dependent reactions, to glyceraldehyde 3-phosphate (G3P). Most of the G3P molecules produced are used to regenerate RuBP. The excess product that is not needed to regenerate RuBP is then used to produce the final carbohydrate products, such as the building material cellulose, the precursors for lipid and amino acid biosynthesis, or as a fuel in cellular respiration (Miziorko and Lorimer, 1983). The overall net equation of the light-independent reaction is as follows:



2.2 Rubisco

The key enzyme of carbon fixation in the CBB cycle is Rubisco. Rubisco is an ancient enzyme that evolved ~3.5 billion years ago when there was little oxygen, but a large concentration of CO₂ in the atmosphere (Wilson and Hayer-Hartl, 2018). Despite its crucial role in photosynthesis, Rubisco is a rather inefficient enzyme (Zhou and Whitney, 2019) with a low enzymatic turn-over rate and a high sensitivity for many physiological inhibitors. What makes the situation worse is Rubisco's promiscuity in substrate selection, as both CO₂ and O₂ are accepted as catalytic substrates, while only the reaction with CO₂ is productive (more details in section 2.2.3) (Hagemann and Bauwe, 2016). Rubisco is probably the most abundant enzyme on Earth, presumably to compensate for its inefficiency (Ellis, 1979). It has

been intensively studied because Rubisco is clearly a rate limiting factor in carbon fixation due to its many shortcomings. Optimizing Rubisco performance to increase photosynthetic yield has been a long-standing goal in the field to meet the rising food demand of the growing human population (Mann, 1999; Mueller-Cajar and Whitney, 2008; Sharwood, 2017).

2.2.1 The different forms of Rubisco

There are four known Rubisco families found in nature, form I, II, III and IV (Tabita et al., 2008). Form IV is called Rubisco-like because these proteins do not catalyze the carboxylation or oxygenation of RuBP (Tabita et al., 2007), and thus will not be discussed further in this thesis. All Rubisco enzymes contain large catalytic subunits (RbcL) of ~50-55 kDa composed of an N-terminal $\alpha+\beta$ domain, a C-terminal $\beta\alpha_8$ triosephosphate isomerase (TIM)-barrel domain and a flexible C-terminal tail (Fig. 4A). Two RbcL subunits always form an anti-parallel dimer, with the N-terminal $\alpha+\beta$ domain of one RbcL interacting with the C-terminal TIM-barrel domain of the other RbcL, forming one catalytic pocket. As a result, each anti-parallel RbcL dimer contains two catalytic sites. The RbcL anti-parallel dimers are the catalytic unit in all forms of Rubisco.

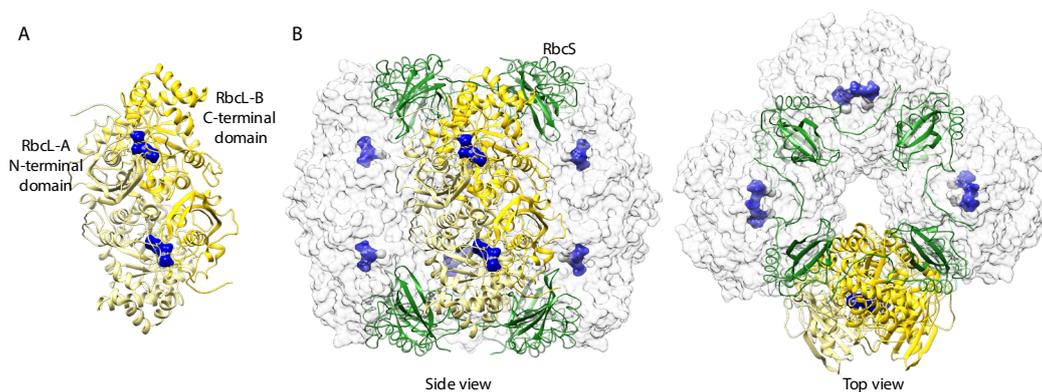


Figure 4: Structure of the anti-parallel RbcL dimer and form I Rubisco. (A) Catalytic unit, anti-parallel RbcL dimer. RbcL subunits are shown in ribbon representation and colored in gold (RbcL-B) and khaki (RbcL-A). Bound sugars are shown in surface representation and colored in blue. One catalytic site is formed by the N-terminal $\alpha+\beta$ domain of RbcL-A and the C-terminal TIM-barrel domain of RbcL-B. (B) Form I Rubisco in side view and top view.

The RbcL₈ core is shown in surface representation with RbcS (green) and one antiparallel RbcL dimer in ribbon representation as shown in (A). Bound sugar molecules are colored in blue. This figure was made with PDB coordinates 1RBL (Newman et al., 1993).

Form I Rubisco, the most prevalent type, is found in plants, algae and photosynthetic bacteria. It forms a hexadecameric complex of ~530 kDa consisting of four RbcL₂ antiparallel dimers packed into a RbcL₈ cylinder. Eight accessory small subunits (RbcS) are capping both ends of the cylinder, with four on each side (Fig.4B) (Andersson and Backlund, 2008; Hayer-Hartl and Hartl, 2020; van Lun et al., 2011). In form I Rubisco, the presence of RbcS is required for catalysis, even though RbcS does not directly participate in forming the catalytic center (Spreitzer, 2003).

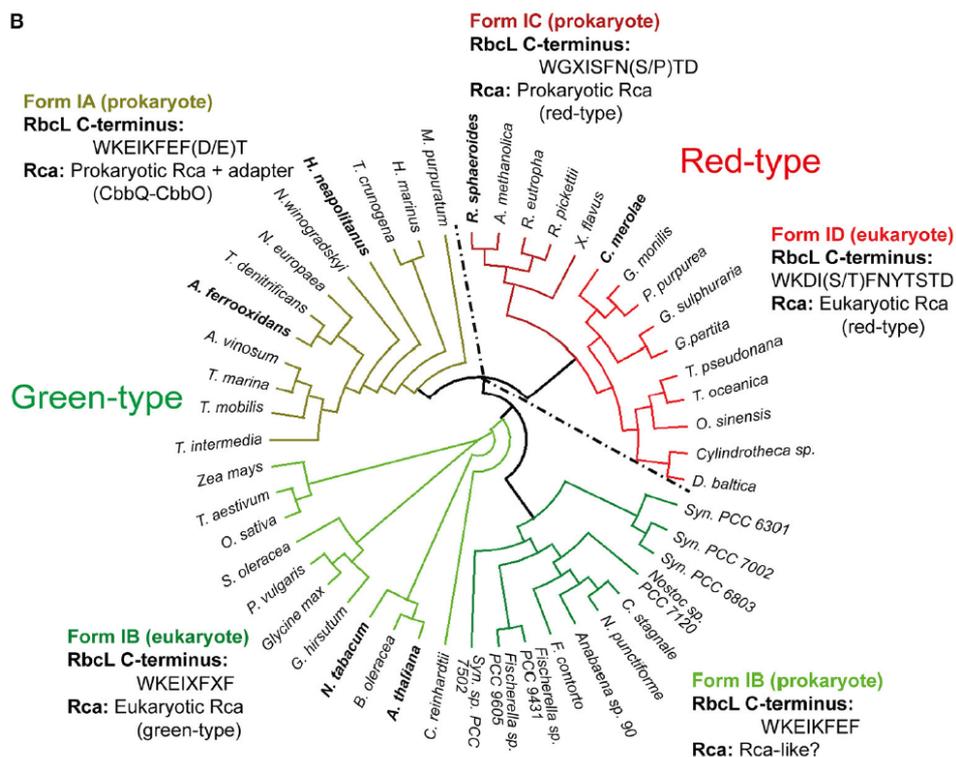


Figure 5: Phylogenetic tree of selected Rubisco RbcL sequences. The green-type enzymes encompass form IA and IB, and the red-type enzymes form IC and ID. The RbcL C-terminal sequences and their associated Rubisco activases (Rca) are indicated. X represents variable residues. Rca isoforms from species indicated in bold have been characterized biochemically and/or structurally. Form IA (prokaryote): *M. purpuratum*, *Marichromatium purpuratum*; *H. marinus*, *Hydrogenovibrio marinus*; *T. crunogena*, *Thiomicrospira crunogena*; *H. neapolitanus*, *Halothiobacillus neapolitanus*; *N. winogradskyi*, *Nitrobacter winogradskyi*; *N. europaea*, *Nitrosomonas europaea*; *T. denitrificans*, *Thiobacillus denitrificans*; *A.*

ferrooxidans, *Acidithiobacillus ferrooxidans*; *A. vinosum*, *Allochromatium vinosum*; *T. marina*, *Thiocapsa marina*; *T. mobilis*, *Thioflaviccoccus mobilis*; *T. intermedia*, *Thiomonas intermedia*. Form IB (eukaryote): *Z. mays*, *Zea mays*; *T. aestivum*, *Triticum aestivum*; *O. sativa*, *Oryza sativa*; *S. oleracea*, *Spinacia oleracea*; *P. vulgaris*, *Phaseolus vulgaris*; *G. hirsutum*, *Gossypium hirsutum*; *N. tabacum*, *Nicotiana tabacum*; *B. oleracea*, *Brassica oleracea*; *A. thaliana*, *Arabidopsis thaliana*. Form IB (prokaryote): *C. reinhardtii*, *Chlamydomonas reinhardtii*; Syn. PCC7502, *Synechococcus* sp. PCC 7502; *F. contorta*, *Fortiea contorta*; *N. punctiforme*, *Nostoc punctiforme*; *C. stagnale*, *Cylindrospermum stagnale*; Syn. PCC6803, *Synechocystis* PCC6803; Syn. PCC7002, *Synechococcus* PCC7002; Syn. PCC6301, *Synechococcus* PCC6301. Form ID (eukaryote): *D. baltica*, *Durinskia baltica*; *O. sinensis*, *Odontella sinensis*; *T. oceanica*, *Thalassiosira oceanica*; *T. pseudonana*, *Thalassiosira pseudonana*; *G. partita*, *Galdieria partita*; *G. sulphuraria*, *Galdieria sulphuraria*; *P. purpurea*, *Porphyra purpurea*; *G. monilis*, *Griffithsia monilis*; *C. merolae*, *Cyanidioschyzon merolae*. Form IC (prokaryote): *X. flavus*, *Xanthobacter flavus*; *R. pickettii*, *Ralstonia pickettii*; *R. eutropha*, *Ralstonia eutropha*; *A. methanolica*, *Acidomonas methanolica*; *R. sphaeroides*, *Rhodobacter sphaeroides*. Figure adapted from (Bhat et al., 2017b).

Form I Rubisco is further classified into green-type (in proteobacteria, cyanobacteria, green algae and plants) and red-type (in photosynthetic bacteria and non-green algae) (Delwiche and Palmer, 1996; Tabita, 1999). Based on molecular phylogenetic analysis of RbcL sequences, green-type Rubisco is further sub-classified into form IA and IB and red-type Rubisco into form IC and ID (Andersson and Backlund, 2008; Tabita et al., 2008) (Fig.5). The subtypes of Rubiscos differ in their assembly pathway (will not be discussed in this thesis) and reactivation mechanism (more details in section 2.6).

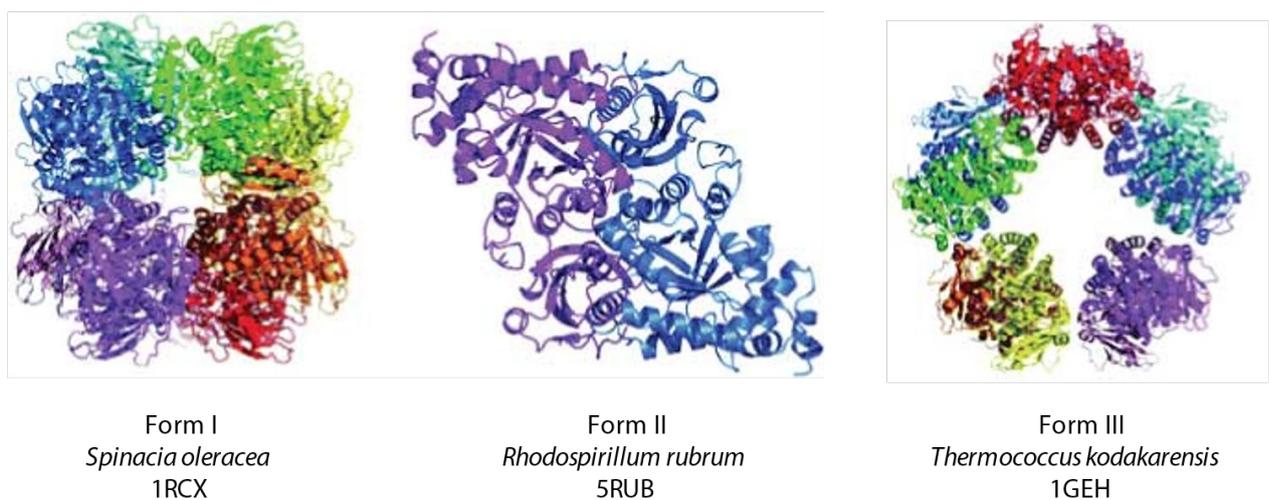


Figure 6: Representative structures of different forms of Rubisco. All forms are comprised of dimers of catalytic RbcL subunits. Form I Rubisco is comprised of four dimers with small

subunits (RbcS) decorating the top and bottom of the RbcL₈ octameric core. Most form II enzymes studied are dimers of RbcL. Form III Rubisco forms tetramers or pentamers of dimers, shown here is the pentameric form. Figure adapted from (Tabita et al., 2008).

Most form II Rubisco enzymes consist of a single RbcL anti-parallel dimer, but can also exist as oligomers of RbcL₂ units, for example a hexameric form II Rubisco has been reported from the bacterium *Rhodopseudomonas palustris* (Satagopan et al., 2014). The most well studied form II Rubisco is from the bacterium *Rhodospirillum rubrum*, which shares only 25-30% sequence identity with RbcL of form I Rubisco (Schneider et al., 1986). Form III Rubisco is only found in certain archaea and fixes carbon in a variation of the CBB cycle (Frolov et al., 2019). Form III Rubisco enzymes have been shown to be comprised of RbcL in arrangements of either one, four or five dimers (Kitano et al., 2001; Maeda et al., 1999). Despite the low homology in the RbcL sequences and oligomerization states among the Rubisco types, the structure of the catalytic unit, the RbcL anti-parallel dimer, is highly conserved (Fig. 6).

2.2.2 Structural basis for Rubisco catalysis

For catalyzing carboxylation as well as oxygenation, the catalytic site of Rubisco must first be activated. Specifically, the highly conserved catalytic site lysine residue K201 (tobacco Rubisco numbering) must be carboxylated at its ϵ amino-group by a non-substrate CO₂ molecule, forming a carbamoyl moiety, and a divalent cation, usually Mg²⁺, must bind to adjacent acidic residues (Lorimer et al., 1976; Lorimer and Mizioro, 1980; NG, 1963). Only when the Rubisco-CO₂-Mg²⁺ (ECM) ternary complex has formed can carboxylation or oxygenation of RuBP proceed (Fig. 7).

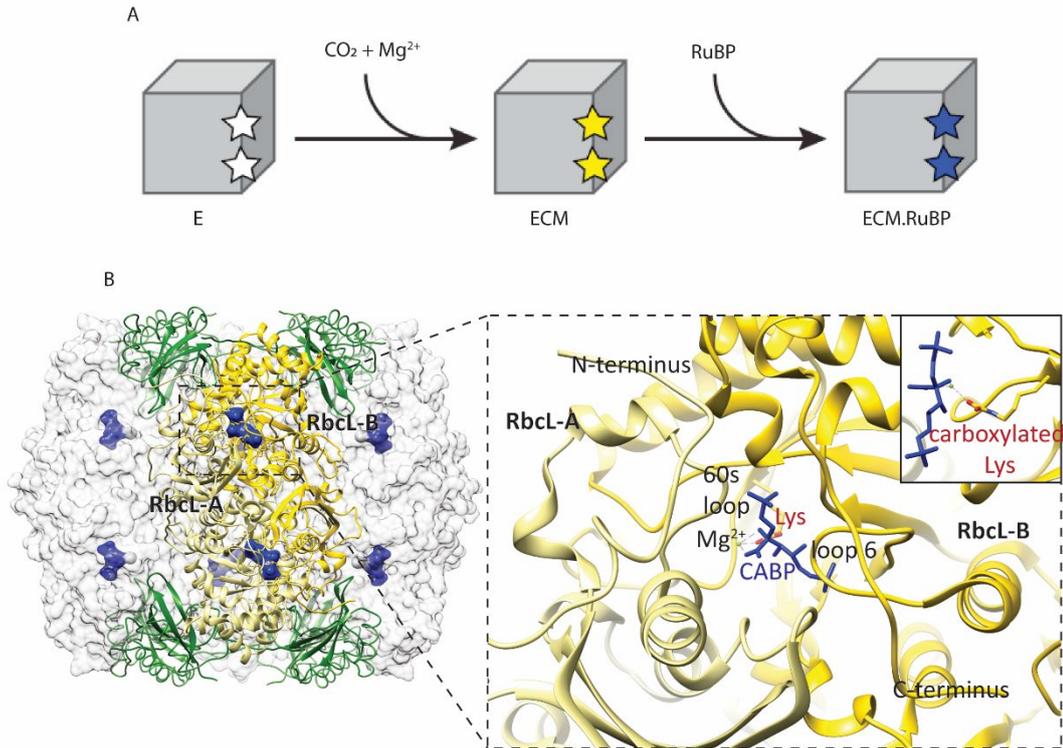


Figure 7: Activation of Rubisco for catalysis. (A) Schematic of the activation of two catalytic sites in Rubisco holoenzyme. Abbreviations: E, the apo form of the Rubisco; ECM, the Rubisco-CO₂-Mg²⁺ ternary complex; ECM.RuBP, Rubisco-CO₂-Mg²⁺-RuBP complex. (B) The critical structural features for catalysis. A Rubisco holoenzyme (left) and a zoomed-in view focusing on one catalytic site in the closed conformation (right). The insertion is a further zoomed-in view focusing on the carboxylated lysine residue. The bound inhibitory sugar CABP (2-carboxyarabinitol-1,5-bisphosphate) is colored in blue. The structural elements important for catalysis are highlighted and labeled, including Mg²⁺, carboxylated lysine, 60s loop of RbcL-A, loop 6 of RbcL-B, N-terminus of RbcL-A (pinning down the 60s loop) and C-terminus of RbcL-B (pinning down the loop 6). Figure made from PDB: 6Z1F (Flecken et al., 2020).

Furthermore, a solvent-free environment is required for Rubisco catalysis. After the substrate RuBP and CO₂/O₂ have bound to the ECM complex, the catalytic pocket needs to be closed properly. A comparison of the Rubisco structures solved in the open and close states trapped by the strong inhibitor 2-carboxyarabinitol-1,5-bisphosphate (CABP) (Karkehabadi et al., 2003; Taylor and Andersson, 1997) reveals that catalytic site closure is achieved by a local conformational change mainly in two loops, one from each RbcL subunit, the so-called 60s loop and loop 6. Mutations in the 60s loop and loop 6 lead to impaired

carbon fixation activity (Karkehabadi et al., 2007). The closed conformation of the 60s loop and loop 6 is further stabilized by the N- and C-terminal tails of both RbcL subunits stretching across the respective loops. (Fig. 7).

2.2.3 Carboxylation and oxygenation (Photorespiration)

Carboxylation is the crucial reaction catalyzed by Rubisco to produce organic compounds from inorganic carbon. In this reaction, Rubisco catalysis involves five steps to produce two molecules of 3PGA as follows: 1) The substrate RuBP is deprotonated at C3 to an unstable enediolate intermediate. 2) Electrophilic addition of substrate CO₂ at C2 generates a branched six-carbon intermediate, CKABP. 3) A hydration takes place at C3, followed by 4) a cleavage of the C2-C3 bond, generating one molecule of 3PGA and one molecule of carbanion intermediate. 5) This intermediate undergoes a stereospecific protonation to generate the second 3PGA molecule (Fig. 8).

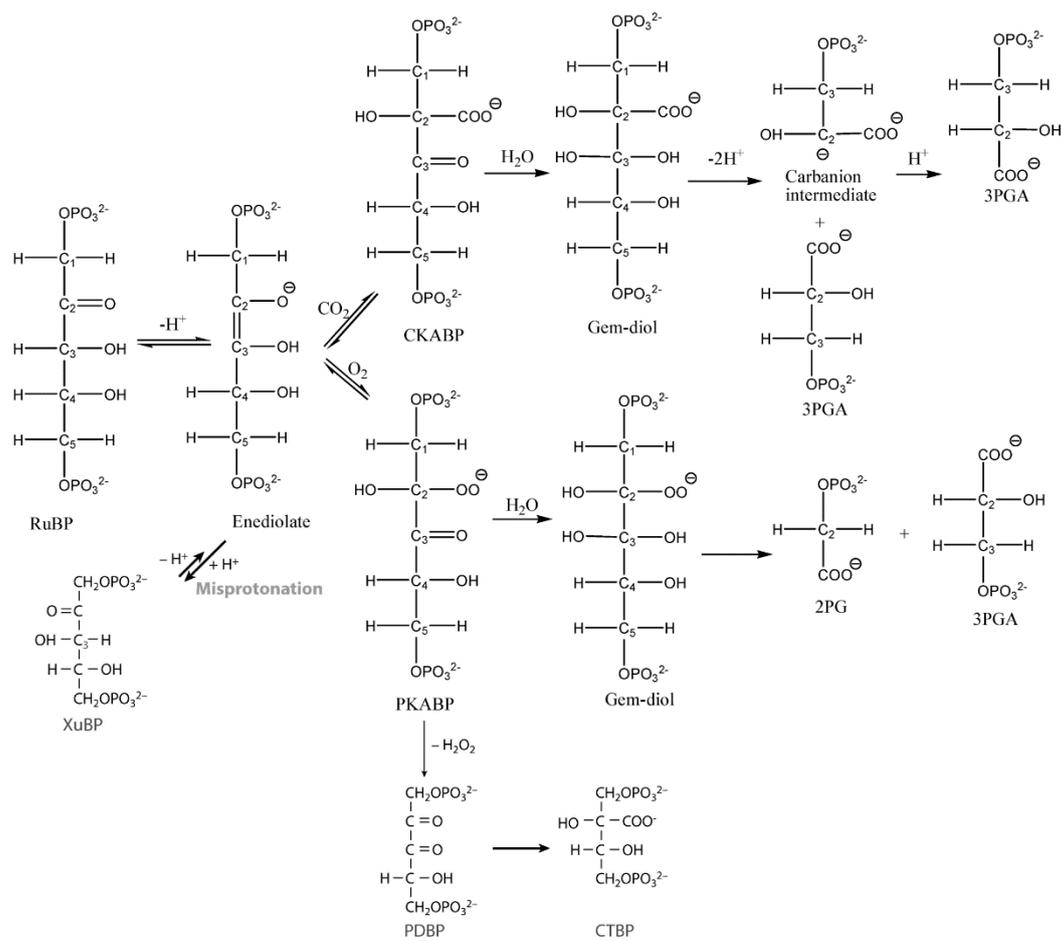


Figure 8: Sequence of the reactions catalyzed by Rubisco and the resulting products. Figure reproduced from (Bracher et al., 2017) and (Kannappan and Gready, 2008). Abbreviations: CTBP, 2-carboxytetritol-1,4-bisphosphate; PDBP, D-glycero-2,3-pentodiulose-1,5-bisphosphate; 2PG, 2-phosphoglycolate; 3PGA, 3-phosphoglycerate; Rubisco, ribulose-1,5-bisphosphate carboxylase/oxygenase; RuBP, ribulose-1,5-bisphosphate; XuBP, D-xylulose-1,5-bisphosphate; CKABP, 2-carboxy-3-keto-D-arabinitol-1,5-bisphosphate; PKABP, 2-peroxo-3-keto-D-arabinitol-1,5-bisphosphate. The inhibitors XuBP and PDBP will be discussed in section 2.2.4.

Due to the structural similarity between O_2 and CO_2 , Rubisco can accept both of them as substrates. The reaction with O_2 is known as oxygenation (Bowes et al., 1971; Lorimer et al., 1973). Oxygenation involves four steps to produce one molecule of 2-phosphoglycolate (2PG) and one molecule of 3PGA as follows: 1) The first step of deprotonation of RuBP is the same as in carboxylation, which results in enolization of RuBP. 2) Binding of O_2 results in the production of a five-carbon intermediate, PKABP. 3) and 4) Hydration followed by a

cleavage of the C2-C3 bond produces the final products, one molecule of 2PG and one molecule of 3PGA (Fig. 8).

The salvage pathway to convert 2PG to G3P is called photorespiration or C_2 photosynthesis (Fig. 9). The reason for requiring such pathway is that 2PG cannot be used in the CBB cycle directly, and moreover, 2PG is an inhibitor of the CBB cycle and causes cellular toxicity in plants (Bauwe, 2010). The photorespiratory pathway in plants is comprised of eight enzymes in four different compartments: chloroplast, cytosol, peroxisome and mitochondrion (Anderson, 1971; Bauwe, 2010; Kelly and Latzko, 1976). In brief, two molecules of 2PG are converted into one molecule of 3PGA, and one molecule of CO_2 , hence resulting in net carbon loss. This pathway consumes ATP and produces ammonia, resulting in energy waste. Although the salvage pathway is essential in higher plants, this pathway was thought to be superfluous in cyanobacteria perhaps due to the evolution of a more efficient CO_2 concentration mechanism (CCM) to inhibit photorespiration (Eisenhut et al., 2008). However, a recent study has reported the presence of three metabolic pathways to convert 2PG to 3PGA in cyanobacteria (Eisenhut et al., 2008). A triple mutant of all three pathways resulted in a high- CO_2 -requiring phenotype, suggesting that the process of photorespiration is essential also in cyanobacteria.

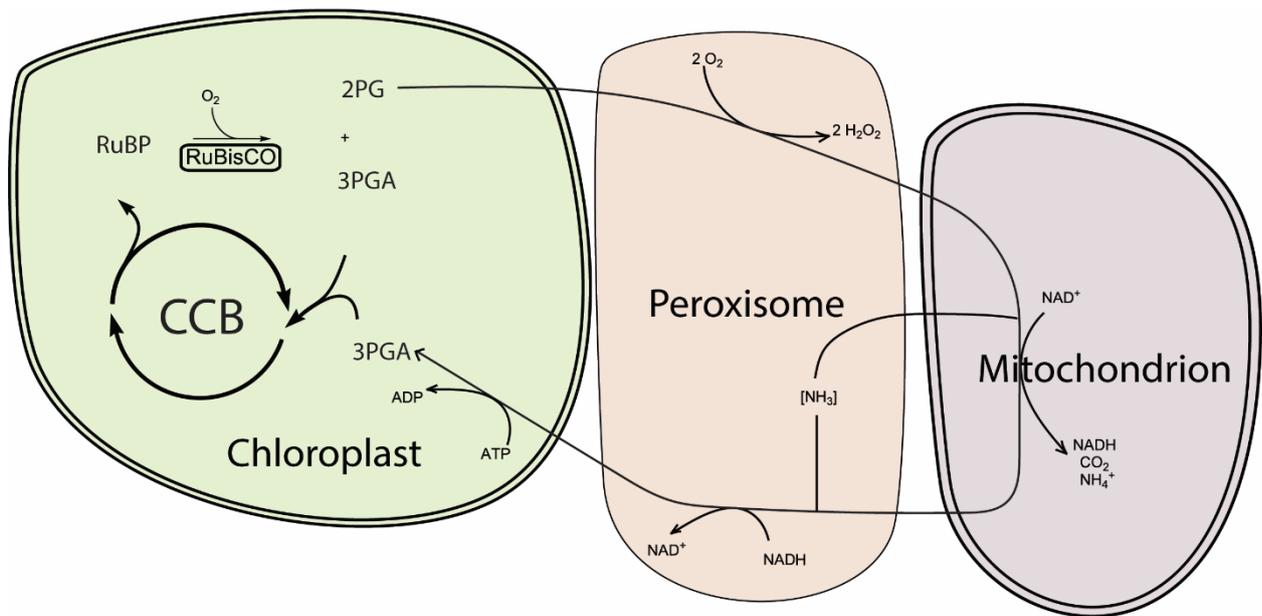


Figure 9: A simplified schematic of plant C₂ photosynthesis cycle. Figure adapted from Wikipedia.

Although ultimately, photorespiration produces G3P, the process appears wasteful and requires energy expenditure. Whether photorespiration has a biological function is under intense discussion. It has been suggested that photorespiration is important in plant pathogen defense, nitrate assimilation and preventing photoinhibition (Bloom et al., 2010; Kozaki and Takeba, 1996; Osmond et al., 1997; Osmond and Grace, 1995; Taler et al., 2004; Wu et al., 1991). It is interesting to note in this context that certain plants, green algae and cyanobacteria have evolved CO₂ concentrating mechanisms (CCM) to increase the local concentration of CO₂ around Rubisco to avoid the oxygenation reaction (more details in section 2.3). Optimizing substrate specificity towards CO₂ of Rubisco is a long-term goal in rational Rubisco engineering (Bainbridge et al., 1995). Certain red-type Rubiscos have considerably improved CO₂ versus O₂ specificity in comparison to Rubisco of green plants (Badger and Bek, 2008).

2.2.4 Rubisco inhibition

Rubisco is prone to inhibition by different mechanisms. In the normal multistep catalytic reaction of Rubisco, the cleavage of the C2-C3 bond of the product gem-diol triggers catalytic site opening (Fig. 8) (Parry et al., 2008). Inhibition occurs when sugar phosphates resembling the transition state of catalysis bind to the catalytic site tightly and arrest the loop 6 and the 60s loop in the closed conformation (Parry et al., 2008). This kind of Rubisco inhibition may also function as a regulatory mechanism in response to fluctuations in the environment during photosynthesis (Geiger and Servaites, 1994).

2-carboxy-D-arabinitol-1-phosphate (CA1P) produced in dark/low irradiation conditions is a night inhibitor of Rubisco found in the chloroplasts of certain plants (Berry et al., 1987; d Moore and Seemann, 1992; Gutteridge et al., 1986; Moore et al., 1995; Parry et al., 1999). At night the light-dependent reactions are arrested. To couple the light-independent reactions with the light-dependent reactions of photosynthesis (Fig. 3), CA1P is synthesized to inhibit Rubisco, and the plant switches to respiration to produce energy. In light conditions, CA1P is removed from the Rubisco catalytic site by the AAA+ chaperone, Rubisco activase (Rca, more details in section 2.6), and is inactivated by dephosphorylation via a specific CA1P phosphatase (Salvucci and Anderson, 1987; Seemann et al., 1985). The CA1P phosphatase is activated by reducing agents such as NADPH (Holbrook et al., 1989; Kingston-Smith et al., 1992), the product of the light-dependent reactions of photosynthesis. Thus, CA1P is only effective at night. Based on the chemical structure of CA1P, the compound CABP was rationally designed as a high affinity inhibitor of Rubisco, and is widely used in experiments to quantify Rubisco active sites in leaf extract or to analyze the activity of Rca (Berry et al., 1987; Robinson and Portis, 1988; Seemann et al., 1985; Sharkey et al., 1991).

In addition, the complex multistep catalytic reactions of Rubisco can lead to side reactions, such as misprotonation, rearrangement, or elimination of H₂O₂ from a

peroxyketone intermediate of the oxygenation reaction, that result in the production of so-called misfire products (Fig. 8) (Parry et al., 2008). Several of these are potent Rubisco inhibitors, including D-xylulose-1,5-bisphosphate (XuBP), D-glycero-2,3-pentodiulose-1,5-bisphosphate (PDBP) and 2-carboxytetritol-1,4-bisphosphate (CTBP) (Harpel et al., 1995; Kane et al., 1998; Keys et al., 1995). The oxygenation of RuBP produces one molecule of PDBP, on average in every 260 turnovers (Kane et al., 1998; Kim and Portis, 2004; Pearce and Andrews, 2003). PDBP can further transform into other Rubisco inhibitors such as CTBP. Enhancing oxygenation produces more PDBP, and Rubisco enzymes with poor CO₂/O₂ selectivity or operating in an environment with high O₂ concentration or elevated temperature tend to increase the oxygenation reaction, and thus produce higher amounts of PDBP (Kim and Portis, 2004).

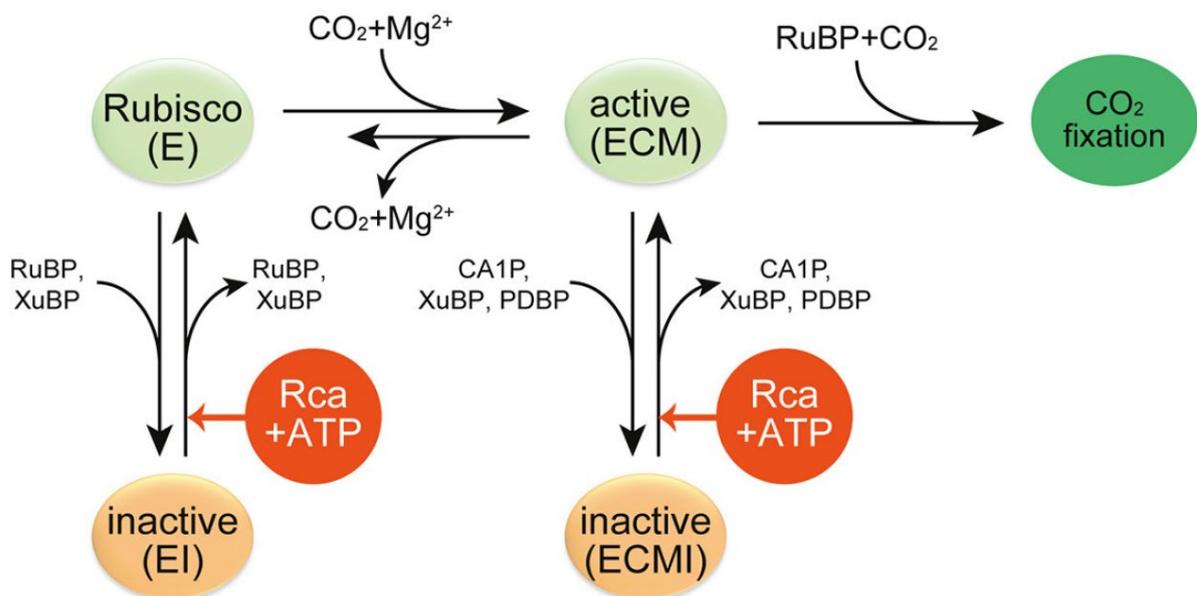


Figure 10: Rubisco regulation by inhibitors and Rca. E, the non-carbamylated enzyme; ECM, the carbamylated and Mg²⁺ ion-bound enzyme; EI, the sugar phosphate inhibited E form; ECMI, the inhibited ECM form; Rca, Rubisco activase. Figure adapted from (Bhat et al., 2017b).

Plant Rubisco is sensitive to binding of different inhibitors in both the uncarbamylated form (E) or carbamylated form (ECM) of the enzyme. When sugar phosphate binds to

uncarbamylated Rubisco, catalysis cannot occur. In this case, even the binding of the natural substrate RuBP leads to Rubisco inhibition (Fig. 10). CA1P and PDBP mainly bind to the activated Rubisco enzyme, ECM (Parry et al., 2008). XuBP can bind to both E and ECM forms (Edmondson et al., 1990; Zhu and Jensen, 1991). Interestingly, no matter which inhibitor is bound, Rca is able to repair the inhibited Rubisco in an ATP-dependent manner (more details in section 2.6).

In vitro, plant Rubisco is inhibited progressively either due to the accumulation of misfire products or the spontaneous decarboxylation of the active site lysine, followed by RuBP binding. This observation is known as Rubisco ‘fallover’ (Edmondson et al., 1990; Robinson and Portis, 1989a). It should be noted that the form I Rubisco enzymes from cyanobacteria, and form II Rubisco are generally not very sensitive to inhibition (Pearce, 2006; Pearce and Andrews, 2003). As mentioned above, Rubisco inhibition also regulates enzyme activity in response to environmental conditions. However, efforts to increase food production and crop yields aim to avoid this regulatory inhibition of Rubisco.

2.3 CO₂ concentrating mechanisms

Many photosynthetic organisms have evolved so-called CO₂ concentrating mechanisms (CCM) to generate CO₂-rich environments that increase the efficiency of carbon fixation by Rubisco, avoiding the side-reaction with oxygen. Different CCM strategies have evolved to benefit organisms living in low CO₂ environments. Photosynthetic bacteria have evolved a microcompartment with a proteinaceous shell called carboxysome. Eukaryotic algae contain a distinct carbon fixation microcompartment, the pyrenoid, within the chloroplast. Certain clades of higher plants, such as C₄- and CAM-plants, spatially or temporally separate carbon uptake and fixation using a specialized leaf anatomy. In the following sections, these distinct CCM strategies are discussed in detail.

2.3.1 Carboxysome

Carboxysomes are found in all cyanobacteria and many chemoautotrophic proteobacteria. They are relatively large structures, with diameters ranging from 100 to 500 nm, consisting of an icosahedral proteinaceous shell (Espie and Kimber, 2011; Price et al., 1998; Shively and English, 1991; Yeates et al., 2008) (Fig. 11). The proteins involved in carbon fixation such as Rubisco, Rubisco activase (Rca) and carbonic anhydrase (CA), are encapsulated inside the shell. Carboxysomes are thought to have evolved as a consequence of the increase in oxygen concentration in the ancient atmosphere to improve photosynthetic performance (Badger and Price, 2003). There are two lineages of carboxysomes, α and β , which have presumably evolved independently. α -Carboxysomes encapsulate form IA Rubisco while β -carboxysomes encapsulate form IB Rubisco (Espie and Kimber, 2011; Kerfeld and Melnicki, 2016; Rae et al., 2013). Apart from having different forms of Rubisco, α - and β -carboxysomes differ in their gene organization and their protein components. The components of the α -carboxysome are usually organized into one operon, with the names of the expressed proteins starting with the prefix Cso, whereas the genes for the β -carboxysome components are generally more dispersed and the names of expressed proteins begin with Ccm (Kinney et al., 2011) (Fig. 11). Remarkably, both α - and β -carboxysome have a similar structure and function, likely a consequence of convergent evolution (Oltrogge et al., 2020; Ryan et al., 2019; Wang et al., 2019).

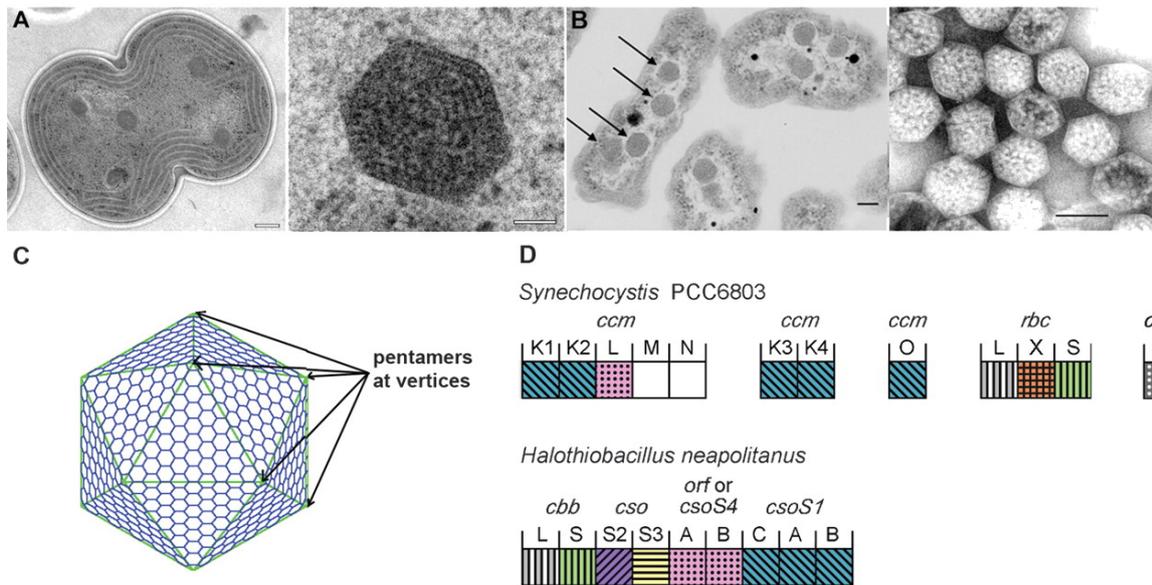


Figure 11: Carboxysome architecture and operon organization. (A) Transmission electron micrographs of β -carboxysomes from *Synechococcus elongatus* PCC6803. A section through a dividing cyanobacterial cell (left, scale bar, 200 nm) and an enlargement of a single carboxysome (scale bar, 50 nm). (B) Transmission electron micrographs of α -carboxysomes from *Halothiobacillus neapolitanus*. A section through cells with their carboxysomes highlighted by arrows (left) and purified α -carboxysomes (right). Scale bar indicates 100 nm. (C) A diagram showing the construction of a large icosahedron from many smaller hexagons and 12 pentagons at the vertices. (D) Genomic arrangement of carboxysome-associated proteins. Homologous proteins from the two model organisms are shaded and colored similarly. The hexameric shell proteins are colored in blue. The pentameric shell proteins are colored in pink. The *rbc* and *cbb* genes code for RbcL and RbcS. Figure adapted from (Tanaka et al., 2008a).

In order to function efficiently, carboxysomes must be provided with raw materials.

The CCM of cyanobacteria and chemoautotrophic proteobacteria starts with uptake of inorganic carbon (C_i) at the outer membrane (Fig. 12). C_i uptake systems such as NDH-1₃ and NDH-1₄ acquire CO_2 selectively (Desmarais et al., 2019; Klughammer et al., 1999; Maeda et al., 2002; Mangiapia et al., 2017; Prommeenate et al., 2004; Shibata et al., 2001), while BicA, SbtA and BCT1 transport dissolved HCO_3^- across the membrane (Omata et al., 1999; Price et al., 2004; Shibata et al., 2001). Independent of the C_i uptake system used, all systems ultimately function as HCO_3^- pumps, increasing the cytosolic HCO_3^- concentration up to 30 times compared to its equilibrium concentration in water (Mangan et al., 2016; Price,

2011). Cytosolic HCO_3^- diffuses freely into the carboxysome lumen via pores in the proteinaceous shell, where the enzyme carbonic anhydrase (CA) is located. In α -carboxysomes the CA function is provided by the protein CsoS3 (CsoSCA) (Price, 2011; So et al., 2004), while in β -carboxysomes it is CcaA and/or CcmM (Fukuzawa et al., 1992; Peña et al., 2010; So et al., 2002; Yu et al., 1992). The CA catalyzes the conversion of HCO_3^- into CO_2 , the substrate of Rubisco. CO_2 is highly concentrated inside the carboxysome due to the selective permeability of the proteinaceous shell that only allows the free diffusion of HCO_3^- but not CO_2 (Kinney et al., 2011; Yeates et al., 2011). Generally, the high catalytic rate of cyanobacterial Rubisco, compared to the plant enzyme, comes at a price of lower CO_2/O_2 selectivity. However, this trade-off is compensated by the high CO_2 concentration in the carboxysome (Badger and Bek, 2008; Davidi et al., 2020; Flamholz et al., 2019).

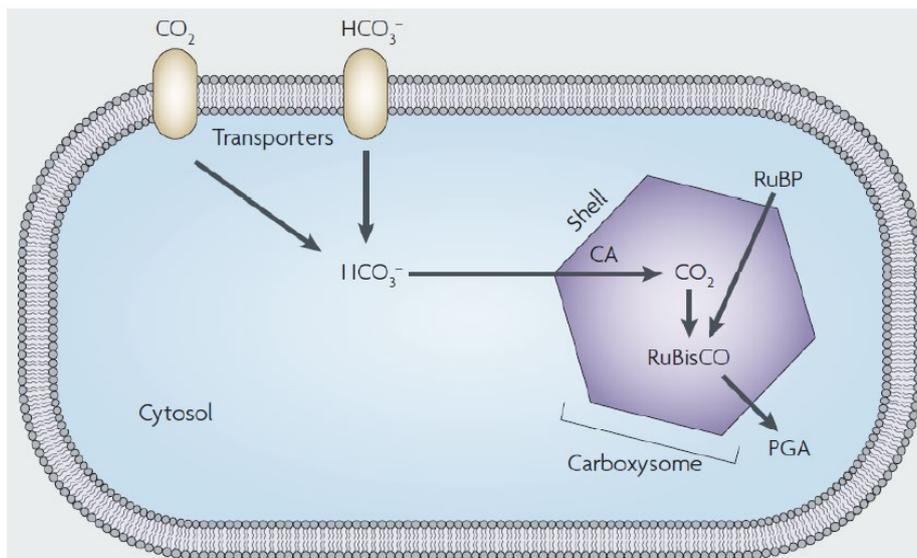


Figure 12: Carboxysomes as bacterial CCM. In the first step of the CCM, bicarbonate is concentrated inside the cell by transporters (Ci uptake systems) located in the cell membrane. The carboxysome functions in the second part of the CCM, and enhances CO_2 fixation by co-localizing Rubisco and CA inside. Metabolites, including RuBP and PGA, traverse the carboxysome shell freely. Figure adapted from (Yeates et al., 2008).

The carboxysome shell consists of multiple proteins containing the bacterial microcompartment (BMC) domain and ideally has icosahedral symmetry (Kinney et al.,

2011) (Fig. 13A). According to the current model, in both α - and β -carboxysomes, each face of the icosahedron is built of hexameric units, which form a two-dimensional layer. The vertices are made of pentameric units. Several crystal structures of the shell components are available (Fig. 13B). CsoS4A (OrfA) and CcmL have a pentameric structure and are thought to be the vertex units in α - or β -carboxysome shells, respectively (Tanaka et al., 2008b). CsoS1A, CsoS1C, CcmK1, CcmK2 and CcmK4 are known as hexamers (Kerfeld et al., 2005; Tanaka et al., 2009; Tsai et al., 2007; Tsai et al., 2009). A hypothetical model for the selective permeability of the carboxysome shell is based on a pore-mediated transport mechanism. The crystal structures of some shell protein oligomers, such as CcmK2 and CcmK4, reveal a highly positively charged pore in the center with a diameter of $\sim 4\text{-}7$ Å, which might attract negatively charged metabolites such as HCO_3^- , RuBP and 3PG, while it disfavors any attractive force for non-charged molecule, such as CO_2 (Kinney et al., 2011). A superposition of all the shell protein structures solved so far shows a similar structure of their BMC domains (Fig. 13C). Some shell proteins such as CsoS1D and CcmO have been predicted to have tandem repeats of this BMC domain (Klein et al., 2009). The crystal structure of CsoS1D reveals a pseudo hexamer which consists of three CsoS1D tandem repeat protomers (Klein et al., 2009). Surprisingly, in two independent crystal structures, two CsoS1D pseudo hexamers interact at roughly a $\sim 60^\circ$ rotational angle. One pseudo hexamer has a wide, open pore (~ 14 Å in diameter) and the other has a closed pore (Klein et al., 2009). Thus, an airlock model for metabolite flux regulation has been proposed (Kinney et al., 2011). Despite the available structural information, the current understanding of the function and organization of shell components is still rather preliminary.

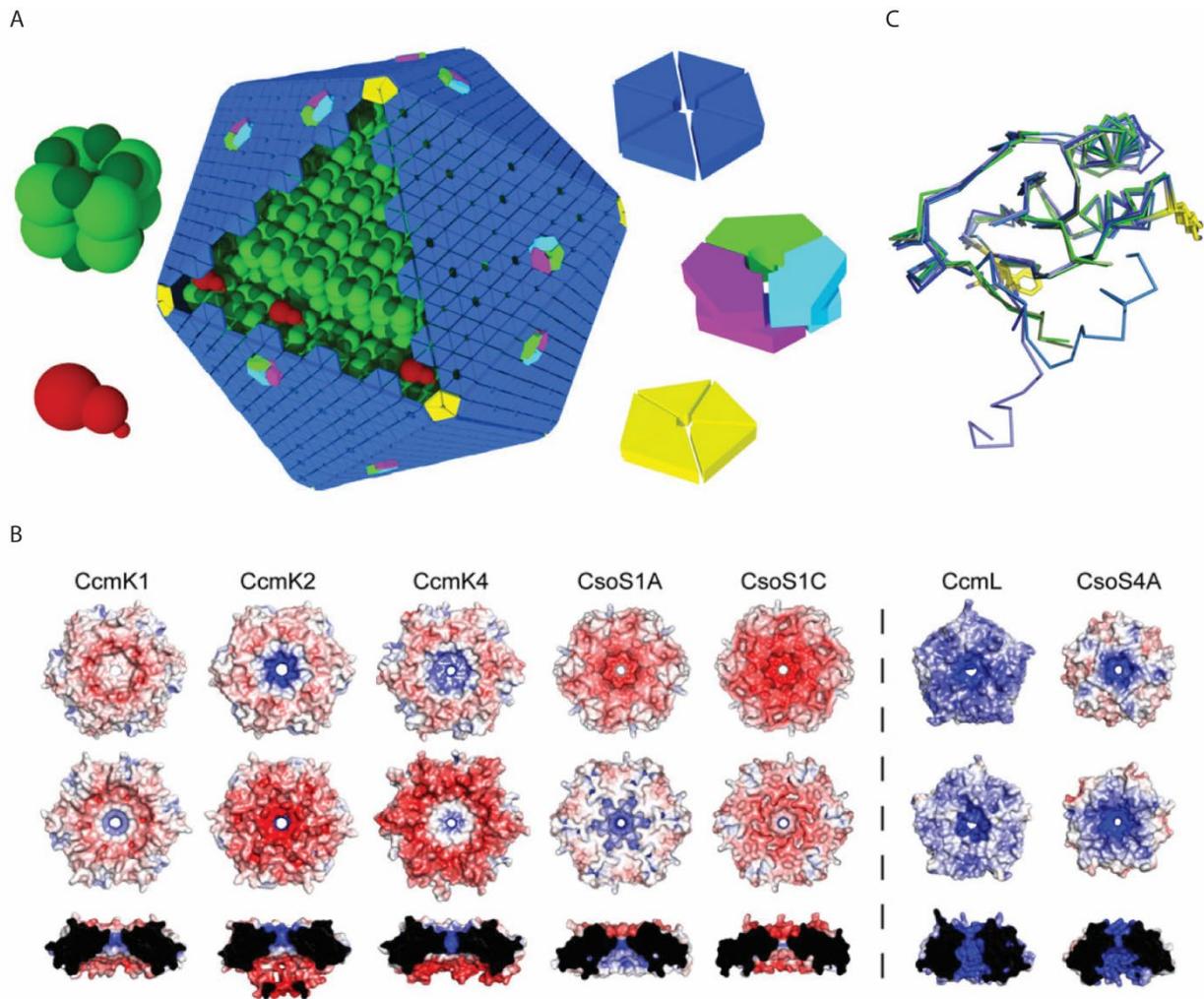


Figure 13: Carboxysome shell proteins. (A) Schematic model of the carboxysome containing RbcS (dark green) and RbcL (green) Rubisco subunits and CA (red). The shell is composed of hexamers (blue), pseudo-hexamers (light blue, magenta, and light green), and pentamers (yellow). (B) Electrostatic comparison of structurally characterized single BMC domain hexameric shell proteins [PDB:3BN4 (CcmK1), 2A1B (CcmK2), 2A10 (CcmK4), 2G13 (CsoS1A), 3H8Y (CsoS1C)] and pentameric shell proteins [PDB:2QW7 (CcmL), 2RCF (CsoS4A)]. Convex (top), concave (middle), and pore cross-section (bottom) views are shown for each structure. Red denotes negative electrostatic potential; blue denotes positive electrostatic potential. (C) Superpositioned BMC domain carboxysome shell monomers from the β - (blue shades) and α - (green shades) carboxysomes. One pore residue (Arg from CcmK4, Lys from CcmK1 and CcmK2, Phe from CsoS1A and CsoS1C) and the conserved Lys found at the edge of the hexamer are shown in yellow sticks. Figure adapted from (Kinney et al., 2011).

The current models of carboxysome assembly differ between α - and β -carboxysomes. α -Carboxysomes co-assemble shell and the lumen proteins, while β -carboxysomes assemble from the inside out (starting with Rubisco aggregation into a condensate) (Kerfeld and

Melnicki, 2016). Interestingly, no matter which part forms first, in both carboxysomes, a scaffold protein, which links individual Rubisco complexes, is essential for carboxysome biogenesis. The scaffold proteins have been identified as CsoS2 and CcmM for α - and β -carboxysomes, respectively. Both CsoS2 and CcmM exist in short and long isoforms. The isoforms are produced by ribosomal frameshifting in CsoS2 (short isoform: CsoS2A; long isoform: CsoS2B) and by an internal ribosome entry site in CcmM (short isoform: M35; long isoform: M58) (Chaijarasphong et al., 2016; Long et al., 2010). In both cases, the short isoform contains 3-5 copies of a putative Rubisco binding motif/domain connected by flexible linker sequences. These Rubisco binding units may be subject to redox-regulation, since most repeat segments contain 1-2 cysteines. The function of additional domains in the long isoforms is not fully understood.

In addition to Rubisco, the scaffold protein (CsoS2/CcmM) and CA, other proteins have also been found in carboxysomes, but at relatively low abundance. Homologs of Rubisco activase (Rca) exist in both α -carboxysomes (more details in section 2.6.2) and β -carboxysome (more details in section 2.6.3), but do not appear to be essential in carboxysome-containing bacteria (Lechno-Yossef et al., 2020; Sutter et al., 2015b). Besides, some Rubisco assembly factors/chaperones have been identified. However, the factors/chaperones vary among species and whether they function inside the carboxysome or are just carried over is unclear (Huang et al., 2019; Wheatley et al., 2014). In β -carboxysomes, the protein CcmN is thought to recruit carboxysome shell components to the CcmM-Rubisco condensate (Kinney et al., 2012). An interaction between CcmN and CcmM was reported (Kinney et al., 2012).

Some proteins that do not locate to the carboxysome are also essential to maintain its function. Recently, an additional operon including several uncharacterized proteins has been

found to be responsible for the function of the α -carboxysome CCM (Desmarais et al., 2019). In the case of β -carboxysomes, McdA and McdB have been known to arrange β -carboxysomes linearly along a central longitudinal axis throughout growth and to equally distribute carboxysomes to daughter cells during division (Savage et al., 2010). Cells with McdA/B knockouts show carboxysome aggregation and possess a longer doubling time (MacCready et al., 2021; Savage et al., 2010).

Because of their high potential in improving photosynthetic performance, transferring artificial, engineered carboxysomes into crop plants is thought to be a promising strategy for improving yields. A deep understanding of carboxysomes and the related CCM mechanisms is however required for rational engineering.

2.3.2 Pyrenoid

The pyrenoid is a non-membrane-bound microcompartment which is found in the chloroplast of nearly all algae and a group of non-vascular plants where it functions as a CCM (Badger et al., 1998; Raven, 2010; Toyokawa et al., 2020). The pyrenoid contains densely packed Rubisco, typically surrounded by a sheath of starch. The Rubisco matrix is usually traversed by specialized membranes, extended from the photosynthetic thylakoid membranes, containing the enzyme carbonic anhydrase and thus supply a high concentration of CO₂ locally (Badger and Price, 1994; Moroney and Mason, 1991). Early biochemical analysis of pyrenoids isolated from green and brown algae revealed that ~90% of the matrix is composed of Rubisco. Alongside Rubisco, Rca and EPYC1(essential pyrenoid component 1) were identified (Meyer and Griffiths, 2013). Rca will be discussed below, in section 2.6. EPYC1 serves as the scaffold protein in pyrenoids, similar to CsoS2 and CcmM in α - or β -carboxysomes, and is essential for normal pyrenoid function. A recent study revealed that the Rubisco matrix in the *Chlamydomonas* pyrenoid has liquid-like properties (Freeman

Rosenzweig et al., 2017). *In vitro* studies also showed that the interaction between EPYC1 and cognate Rubisco is multivalent and induces phase separation (more details see section 2.4) (Wunder et al., 2018).

2.3.3 C₄ and crassulacean acid metabolism photosynthesis

Plants which live in hot or dry conditions close their stomata to prevent water loss via their leaves. The closing of the stomata hinders gas exchange, reducing CO₂ concentration and increasing O₂ concentration in photosynthetic cells, as photosynthesis proceeds. The rising O₂/CO₂ ratio causes an increase of photorespiration. To compensate, these plants have evolved a CCM that is based on a spatial or temporal separation of CO₂ uptake and Rubisco-mediated carbon fixation.

In C₄ photosynthetic plants, Rubisco is enriched in bundle sheath cells, which are surrounded by a layer of mesophyll cells. The CO₂ uptake happens in the mesophyll cells, where the enzyme phosphoenolpyruvate (PEP) carboxylase catalyzes the addition of CO₂ to the three-carbon molecule PEP. This generates the four-carbon organic acid, oxaloacetate, which can be chemically reduced to malate. The products are translocated into bundle sheath cells, where CO₂ is released by decarboxylation and consumed by Rubisco catalysis. Thus, PEP carboxylase functions as a CO₂ pump (Morita et al., 1997), increasing the local CO₂ concentration in bundle sheath cells.

Crassulacean acid metabolism (CAM) plants separate CO₂ uptake and carbon fixation temporally. They take up CO₂ at night when the stomata are open. Fixed carbon is stored as malate, which is only decarboxylated to release CO₂ during the day when the stomata are closed (Dodd et al., 2002). CAM plants thereby increase the CO₂ concentration during photosynthesis.

2.4 Liquid-like phase separation

A fundamental question for all biochemical processes concerns their spatial regulation. One strategy that cells have evolved is to create compartments or organelles. A compartment has two important properties: it must have a boundary that separates it from its surroundings, and the components within it must be able to diffuse freely so as to facilitate the chemical reactions that take place inside (Hyman et al., 2014). Many compartments possess membrane boundaries, such as the classical organelles mitochondria and lysosomes. The membranes separate the lumen and cytosol, thus maintaining distinct chemical environments. The reactions take place in the lumen and transporters in the membranes facilitate metabolite exchange with the cytosol. Some compartments use other materials as structural boundaries such as the protein shell of carboxysomes. However, many stable compartments, such as centrosomes, nucleolus and stress granules, do not have structural boundaries. These are thought to phase-separate from their surroundings, forming a protein- or nucleic acid-rich fluid phase without crystallization.

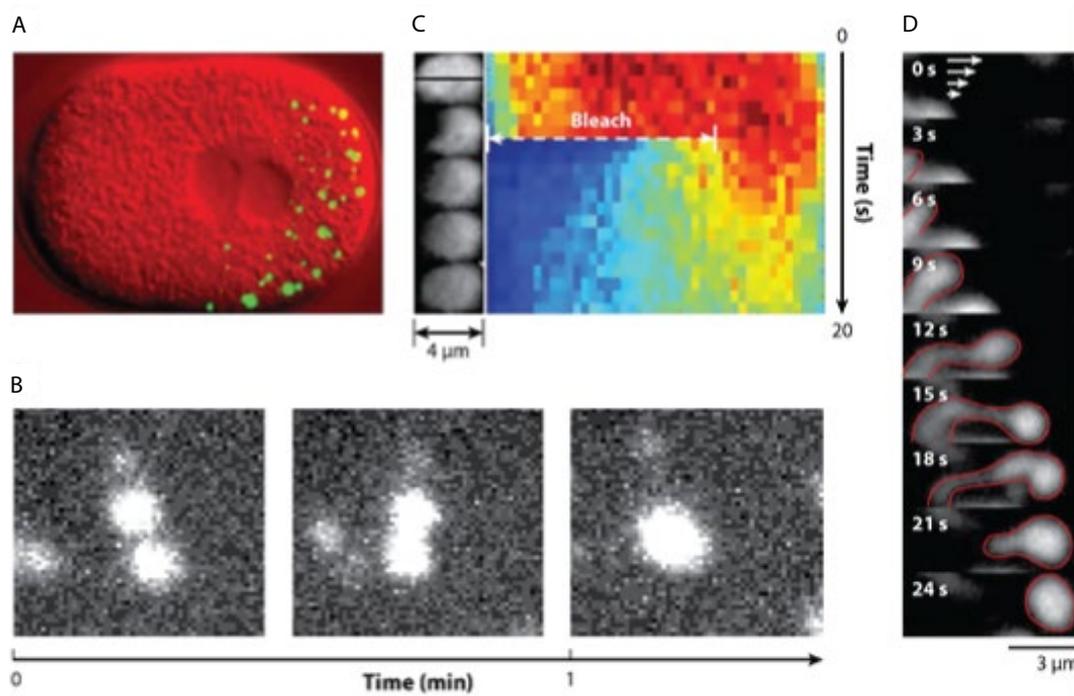


Figure 14: P granules exhibit characteristics of liquid droplets. (A) P granules (green; GFP tagged) in the cytoplasm of a single cell–stage *Caenorhabditis elegans* embryo. (B) Two P granules (white) fuse and relax their shape within about one minute. (C) Fluorescence distribution before and after photobleaching of a large GFP-tagged P granule (left). Kymograph of linear intensity profiles along the anterior-posterior axes (right). Red color indicates high intensity and blue corresponds to background intensity. Fluorescence recovery occurs in about 5 s. (D) P granule (red outline) deformed by sheared flow with a direction indicated by the white arrows. Figure adapted from (Brangwynne et al., 2009) and (Hyman et al., 2014).

Many of these non-membrane-bound compartments have liquid-like properties, such as cohesion, viscosity, dynamicity and surface tension driven spherical shape (Hyman et al., 2014). They are generated by a phenomenon called liquid-like phase separation (LLPS). The first clear example of a liquid-like compartment was the P granule from *Caenorhabditis elegans* embryos. P granules were identified by electron microscopy (Wolf et al., 1983) and fluorescence microscopy (Strome and Wood, 1983) and were shown to fuse, exchange components rapidly with the cytoplasm, be easily deformed by flows and have a viscosity similar to runny honey (Brangwynne et al., 2009) (Fig. 14). LLPS may be also involved in pathology. For example, the RNA-binding protein TDP-43 related to amyotrophic lateral sclerosis (ALS) undergoes LLPS *in vitro* and *in vivo* (Conicella et al., 2016; Maharana et al., 2018). A recent study suggested that the Rubisco matrix in the pyrenoid behaves like a liquid as well (Freeman Rosenzweig et al., 2017). LLPS is an emerging field, it may serve as a fundamental process in many biochemistry events but more evidence is required to reveal its role in detail.

2.5 AAA+ ATPases

The ATPases associated with various cellular activities (AAA+) family form a large family of proteins, many of which are essential, and are found in all three kingdoms of life. AAA+ proteins represent a subset of the Walker (P-loop) ATPases (Erdmann et al., 1991; Kunau et

al., 1993), which hydrolyze ATP to power mechanical work for various functions, including proteolysis, protein folding, membrane trafficking, cytoskeletal regulation, organelle biogenesis, DNA replication, and intracellular motility (Dean and Annilo, 2005; Erdmann et al., 1991; Lee et al., 2003; Simonetta et al., 2009; Singleton and Wigley, 2002).

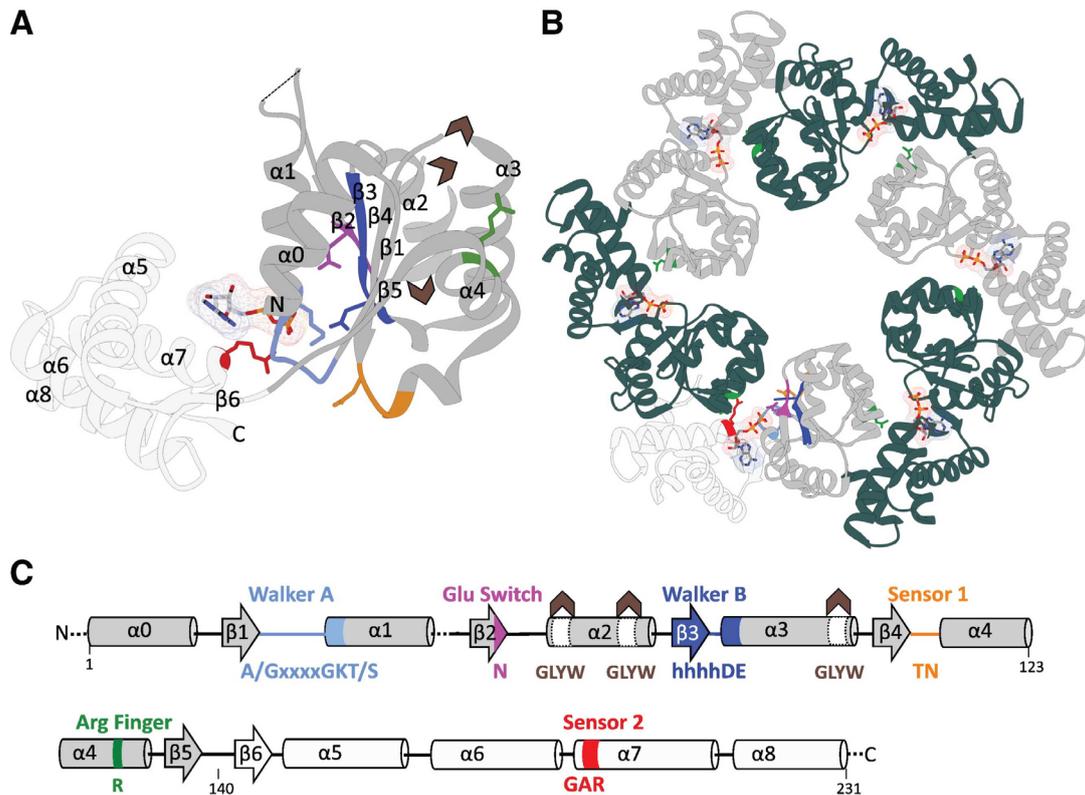


Figure 15: Structure and oligomeric arrangement of the AAA+ domain. The AAA+ ATPase module (A) and its hexameric assembly (B) as seen in many crystal structures are depicted using the example of Lon from *Thermococcus onnurineus* bound to ADP (PDB ID: 3K1J). Only the core ATPase domains of Lon are shown. In (A) the NDB is colored in dark grey and the C-terminal domain in light grey. All structural elements are highlighted according to the color code in the secondary structure annotation (C). Consensus sequences for each motif are given. Brown arrows indicate the position of pore loop insertions present in different AAA+ subgroups. Visible secondary structural elements are numbered in the order of appearance from N- to C-terminus. Protomers of the hexameric assembly in (B) are colored alternating in dark green and grey. One protomer is colored as shown in (A). Arginine fingers are colored in green and the nucleotide is shown as stick and surface representation in all protomers. Figure adapted from (Wendler et al., 2012).

AAA+ family members have in common one or more conserved AAA+ module of ~230 amino acids, comprised of a N-terminal nucleotide binding domain (NDB) with a α/β Rossman fold and a C-terminal α -helical domain. Most of AAA+ proteins function as

oligomers, often hexamers, with a donut shape, and a pore in center. The ATP binding pocket is located at the interface between two subunits (Fig. 15). The central pore of certain AAA+ subfamilies/clades is active in protein remodeling/unfolding and can engage a flexible loop or tail of the substrate and exert mechanical force (Bhat et al., 2017a; Dong et al., 2019; Flecken et al., 2020; Twomey et al., 2019).

The NDB is structurally characterized by a central 5 stranded parallel β -sheet in a Rossman fold with a 5-1-4-3-2 β -strand order (Fig. 15). Characteristic sequence motifs have been shown to play critical roles in ATP binding and hydrolysis. The Walker A (A/GxxxxGKT/S, x denotes any amino acid) and Walker B (hhhhDE, h denotes a hydrophobic amino acid) motifs, which coordinate the β and γ phosphate groups of ATP and a water molecule-activating Mg^{2+} , are located at the tips of β strands 1 and 3. A conserved glutamate in Walker B switches most ATPases from an active to an inactive conformation upon ATP binding (glutamate switch), which is reversed by substrate binding (Zhang and Wigley, 2008). This so called “glutamate switch” couples substrate binding to ATP hydrolysis. The tip of β strand 4 typically contains a motif, sensor 1, carrying polar residues that coordinate the attacking water molecule to hydrolyze ATP in concert with the Walker B residue. Another important motif for ATP hydrolysis is the “R-finger” with one or more arginine residues, usually located behind $\alpha 4$. During ATP hydrolysis, R-fingers are close to the γ phosphate of the bound nucleotide in the neighboring subunit (Wendler et al., 2012). In addition to the motifs crucial for nucleotide binding and hydrolysis mentioned above, some loop insertions in the $\alpha 2$ and/or $\alpha 3$ helices facing towards the central pore have been shown to play critical roles in substrate interaction. They are called “pore loops”, often having a conserved aromatic-hydrophobic-glycine motif (Dong et al., 2019; Fei et al., 2020; Twomey et al., 2019).

The core topology of the C-terminal domain is a 4-helix-bundle. This domain also harbors a well-conserved arginine residue in the sensor 2 motif, located at the tip of $\alpha 7$ and contacting the bound nucleotide (Ammelburg et al., 2006; Botos et al., 2004). It should be noted that although the AAA+ core is structurally highly conserved, the residues in the R-finger, sensor 1 or sensor 2 are not conserved in all clades.

The AAA+ ATPase proteins can be grouped into two major classes, based on how many AAA+ modules each protomer contains. Class I proteins contain two modules and class II proteins contain one module (Erzberger and Berger, 2006; Hanson and Whiteheart, 2005). Interestingly, in some class I proteins, the ATPase activity of one module is dispensable and diminished (Davies et al., 2008; Lenzen et al., 1998). In another classification based on the topology of the AAA+ domain, some subfamilies or clades of AAA+ ATPases have been identified, such as the classic AAA+ ATPases (e.g., NSF, Cdc48, proteasome, katanin, Vps4 and Rca), pre-sensor 1 insert clades (e.g., σ^{54} activator, Lon A, and dynein), and the HEC clade (including helicases and clamp loaders) (Wendler et al., 2012).

Most of the initially identified AAA+ proteins such as Cdc48, ClpX and Rca belong to the “classic” AAA+ ATPase clade. This clade is structurally distinct from others by a unique insertion of a small α -helix element between $\beta 2$ and $\alpha 2$ in the NBD (Erzberger and Berger, 2006; Iyer et al., 2004), the lack of the sensor 2 and the presence of two arginine residues in the R-fingers. Besides, most proteins from this clade contain an additional N-terminal domain relevant to their functional specialization (Roll-Mecak and Vale, 2008). Proteins from this clade are mostly involved in structural unfolding or remodelling. Some unfoldases belonging to this clade have been studied intensively, such as the 26S proteasome, Valosin-containing protein (VCP, alternatively named p97 or Cdc48) and ClpX (Dong et al.,

2019; Fei et al., 2020; Twomey et al., 2019). The conserved aromatic residue located in pore loop1 has been shown to interact with a flexible loop element of the substrate directly (Dong et al., 2019; Twomey et al., 2019). Sequential ATP hydrolysis in the oligomer triggers the rearrangements of pore loops in a hand-over-hand fashion, threading the substrate through its central pore and thereby unfolding it (Dong et al., 2019; Twomey et al., 2019).

2.6 Rubisco activase (Rca)

Rubisco activase (Rca), a specialized AAA+ remodeler facilitating the release of inhibitory sugar phosphates from blocked Rubisco, was first identified in plants (Salvucci et al., 1985; Somerville et al., 1982). Functional homologs were subsequently found in many photosynthetic organisms from chemoautotrophic bacteria to eukaryotic algae, containing either green-type or red-type Rubiscos (Loganathan et al., 2016; Mueller-Cajar et al., 2011; Sutter et al., 2015a; Tsai et al., 2015). The mutant plants carrying Rca mutations had a high-CO₂-requiring phenotype and accumulate RuBP because of a substantially decreased activation status of Rubisco (Portis Jr and Salvucci, 2002; Somerville et al., 1982). Rca is a heat labile enzyme (Robinson and Portis, 1989b; Salvucci et al., 2001). A heat stable form of Rca might improve plant performance in high temperature environments. The crucial role of Rca in maintaining Rubisco activity makes it another hot target in plant engineering.

The diverse Rca proteins have been suggested to have evolved independently in different lineages of Rubisco-harboring organisms (Wachter et al., 2013) (Fig. 5). Hence, although all Rca proteins function in releasing tight binding Rubisco inhibitors, they differ in domain structure, oligomerization, regulation and binding partners. Rca proteins have co-evolved with their cognate Rubiscos, and as a result, even in plants, some sequence specificities exist for cognate Rubisco recognition. Despite these specific differences, all Rca

proteins share the core subunit architecture belonging to the AAA+ family (Mueller-Cajar et al., 2011; Neuwald et al., 1999; Sutter et al., 2015b).

2.6.1 Rca of red-type Rubisco

Red-type Rca is encoded by the *cbbX* gene in both prokaryotes containing form IC Rubisco and eukaryotic red algae containing form ID Rubisco.

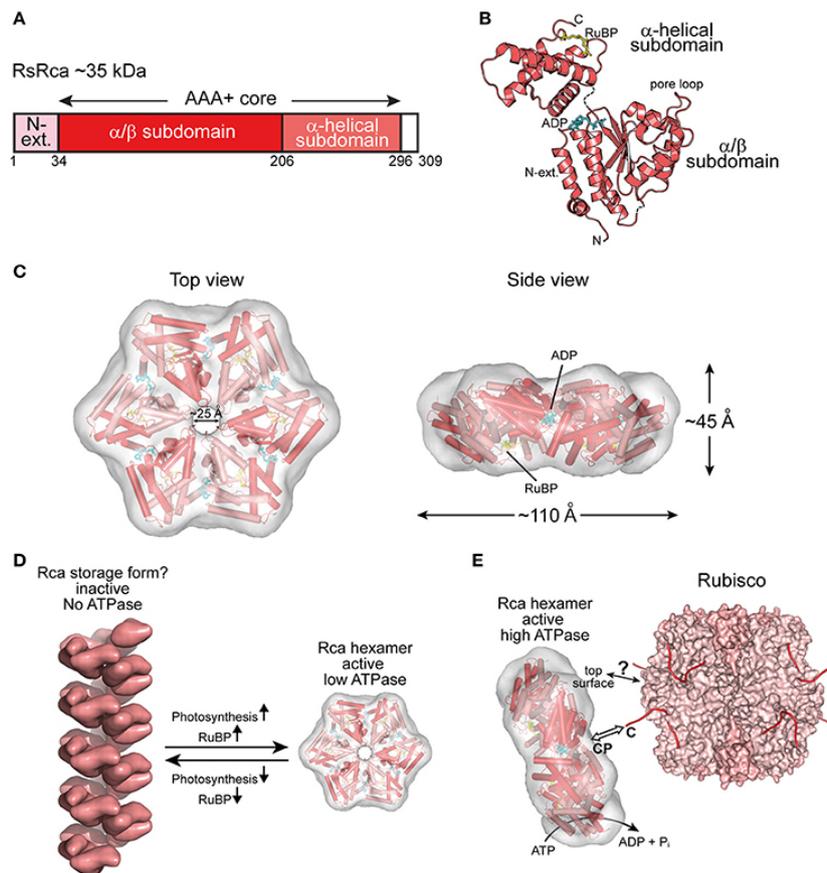


Figure 16: The prokaryotic Rca of red-type form IC Rubisco. (A) Schematic representation of the domain structure of Rca from *R. sphaeroides* (*RsRca*). (B) Crystal structure of the monomer (PDB: 3SYL) shown in ribbon representation. The α/β and α -helical subdomains of the AAA+ core are indicated, as well as the N-terminal extension (N-ext.) of *RsRca*. The positions of the canonical pore loop, ADP (cyan) and the allosteric regulator, RuBP (yellow), are also indicated. (C) Top and side views of the *RsRca* hexameric model superposed on the electron microscopic reconstruction, with alternating subunits shown in two shades of red (EMDB EMD-1932; PDB 3ZUH). (D) Model of the putative storage form of prokaryotic Rca from red-type form IC and its conversion to active hexamer. (E) Model of the mechanism of prokaryotic Rca from red-type form IC Rubisco. Central pore (CP). Rca is displayed as in (C). Rubisco (PDB: 4F0K) is shown in surface representation with the RbcL and RbcS subunits in different shades of pink. The RbcL C-termini are drawn as lines in red. Figure adapted from (Bhat et al., 2017b).

The most well-studied prokaryotic red-type Rca is from the proteobacterium *Rhodobacter sphaeroides* (*RsRca*) (Bhat et al., 2017a; Gibson and Tabita, 1997; Mueller-Cajar et al., 2011) (Fig. 16). *RsRca* only forms the functionally active hexamer in the presence of both ATP and RuBP (Mueller-Cajar et al., 2011). The crystal structure suggests that RuBP binds to the C-terminal α -helical domain, which is located at the bottom of the hexamer (*trans* side) opposite the Rubisco interface. One pore loop bears the conserved Tyr-Ile-Gly sequence that is required for couples the CbbX ATPase and activase functions (Mueller-Cajar et al., 2011). The central pore of the hexamer is narrow with a diameter of about 25 Å. In the presence of ATP and the absence of RuBP, *RsRca* forms a spiral-shaped high molecular weight assembly without ATPase activity, which may represent a storage form during conditions when RuBP is limited (Mueller-Cajar et al., 2011). A recent study using cryo-EM, hydrogen/deuterium (H/D) exchange and chemical crosslinking mass spectrometry (MS) suggested that the central pore of Rca engages the C-terminal extension of RbcL (Bhat et al., 2017b).

Interestingly, two *cbbX* genes, one nuclear-encoded and one plastid-encoded, have been identified in the red alga *Cyanidioschyzon merolae* (*Cm*), containing form ID Rubisco (Hovde et al., 2015). A recent study suggested that the functional *CmRca* is a hetero-hexamer (Loganathan et al., 2016). Although *CmRca* does not require RuBP for hexamer formation, RuBP acts as an allosteric regulator and stimulates ATPase activity. *CmRca* can reactivate both *CmRubisco* and *RsRubisco* (Loganathan et al., 2016). C-terminal truncation of RbcL abolishes reactivation, suggesting a conserved mechanism with the bacteria red-type Rca (Loganathan et al., 2016; Mueller-Cajar et al., 2011).

2.6.2 Rca of green-type Rubisco form IA

In photoautotrophic or chemoautotrophic proteobacteria with form IA Rubisco contain the AAA+ protein CbbQ, which was identified as a functional Rca homolog (Hayashi et al., 1997; Tsai et al., 2015). What makes CbbQ different from other Rca proteins is the requirement of a substrate adaptor, CbbO, that binds Rubisco through a von Willebrand factor A (VWA) domain (Ammelburg et al., 2006; Snider and Houry, 2006) (Fig. 17). CbbO is comprised of around 770 residues, containing a C-terminal VWA domain with several metal-ion-dependent adhesion site (MIDAS) motifs (Whittaker and Hynes, 2002), which usually interact with a surface-exposed carboxylate group. The basal ATPase rate of CbbQ is low, but can be gradually simulated by binding to CbbO and the inhibited Rubisco substrate (Tsai et al., 2015; Tsai et al., 2020).

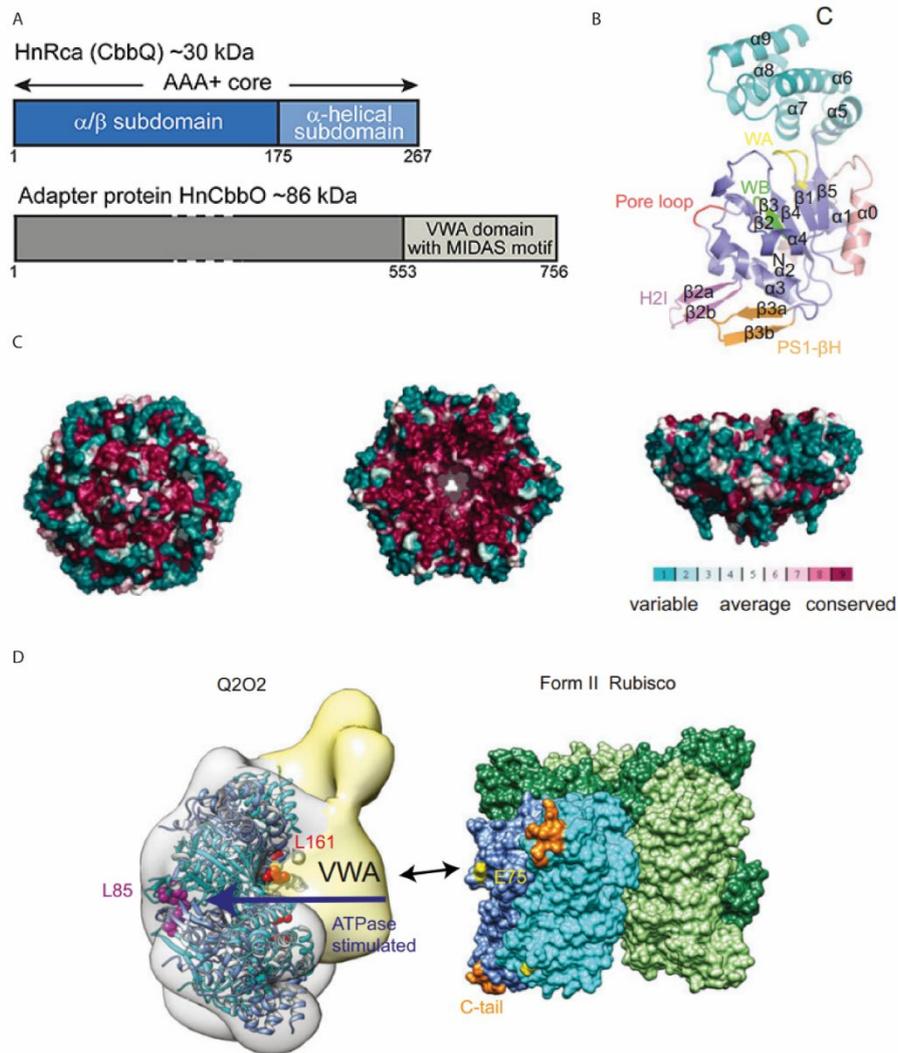


Figure 17: The prokaryotic Rca of green-type form IA Rubisco and form II Rubisco. (A) Schematic representation of the domain structure of Rca from *H. neapolitanus* and its adapter protein (CbbO). (B) Crystal structure of the *AfCbbQ2* monomer (PDB: 6L1Q). H2I, helix 2 insert; PS1-βH, presensor 1-β hairpin; WA, Walker A motif; WB, Walker B motif. (C) A sequence alignment of 100 CbbQ2 sequences was used to map surface conservation onto the *AfCbbQ2* hexamer. (D) Model of CbbQO function on form II Rubisco. Rubisco is a surface representation of the hexameric *Rhodospseudomonas palustris* form II Rubisco (PDB: 4lf1). Crystal structure of *AfCbbQ2* hexamer is docked into the negative staining EM map of *AfCbbQ2O2* complex (EMDB: 0789). Figure reproduced from (Bhat et al., 2017b) and (Tsai et al., 2020).

In many chemoautotrophic proteobacteria, both form IA and form II Rubisco coexist. Accordingly, two open-reading frames (ORFs) of CbbQ and CbbO have been found, one near to each Rubisco gene (Badger and Bek, 2008). Each functional hetero-oligomer contains one CbbQ hexamer and one CbbO molecule (Tsai et al., 2015). Interestingly, the CbbQ-CbbO

complex is only able to reactivate the Rubisco expressed from the neighboring gene (CbbQ1O1 represents the complex working for form IA Rubisco and CbbQ2O2 for form II Rubisco). Although, according to *in vitro* studies, form II Rubisco is not sensitive to inhibition, CbbQ2O2 does increase the carbon fixation rate of form II Rubisco (Tsai et al., 2015). However, the requirement for CbbQ2O2 *in vivo* remains to be demonstrated.

Recently the crystal structures of *Halothiobacillus neapolitanus* CbbQ1 (*HnCbbQ1*), which is located inside the α -carboxysome and *Acidithiobacillus ferrooxidans* CbbQ2 (*AfCbbQ2*), acting on form II Rubisco, were solved (Sutter et al., 2015a; Tsai et al., 2020). The C-terminal domain of the AAA+ core in both structures has five helices instead of four, which is the standard in other AAA+ proteins. Due to this additional helix, the CbbQ hexamer has a hexagonal shape. The concave side of the hexamer is sequence conserved. A reconstruction from negative-stain EM suggested that the *AfCbbO2* VWA domain interacts with *AfCbbQ2* and is located at the center of the hexamer, occluding the pore (Tsai et al., 2020). Mutations in the pore loop-1 of CbbQ, which has no aromatic residues, do not show strong impairment in reactivation. The *AfCbbQ2* crystal structure revealed a helix 2 insertion (H2I), occluding the pore to less than 10 Å. Thus, it is unclear whether the central pore of CbbQ interacts with substrate directly.

Site-directed mutagenesis suggests conserved interaction sites of Rubisco with cognate Rca. D82 and E75 in form IA and form II RbcL, respectively, are involved in the interaction with the MIDAS motifs of CbbO (Tsai et al., 2015). E75 corresponds to residue 89 in plant RbcL, which mediates Rca specificity for solanaceous versus non-solanaceous Rubisco. In addition, RbcL subunits associated with CbbQ and CbbO have a conserved H(K/R) motif at the C-terminus, which is not present in other form I or form II Rubiscos. The deletion of the motif or addition of further residues to this tail abolishes reactivation.

Although mutagenesis provides some information on interactions within the CbbQ-CbbO-

Rubisco ternary complex, a high-resolution structure of the complex will be required for a detailed mechanistic understanding.

2.6.3 Rca of green-type Rubisco form IB

Plant Rcas have been studied intensively since their identification in *Arabidopsis thaliana* almost three decades ago (Portis, 2003; Portis Jr and Salvucci, 2002). Crystal structures were published for Rca from *Arabidopsis thaliana* (*AtRca*) and *Nicotiana tabacum* (*NtRca*), and for the C-terminal domain of the AAA+ core from *Larrea tridentata* (Hasse et al., 2015; Henderson et al., 2011; Stotz et al., 2011).

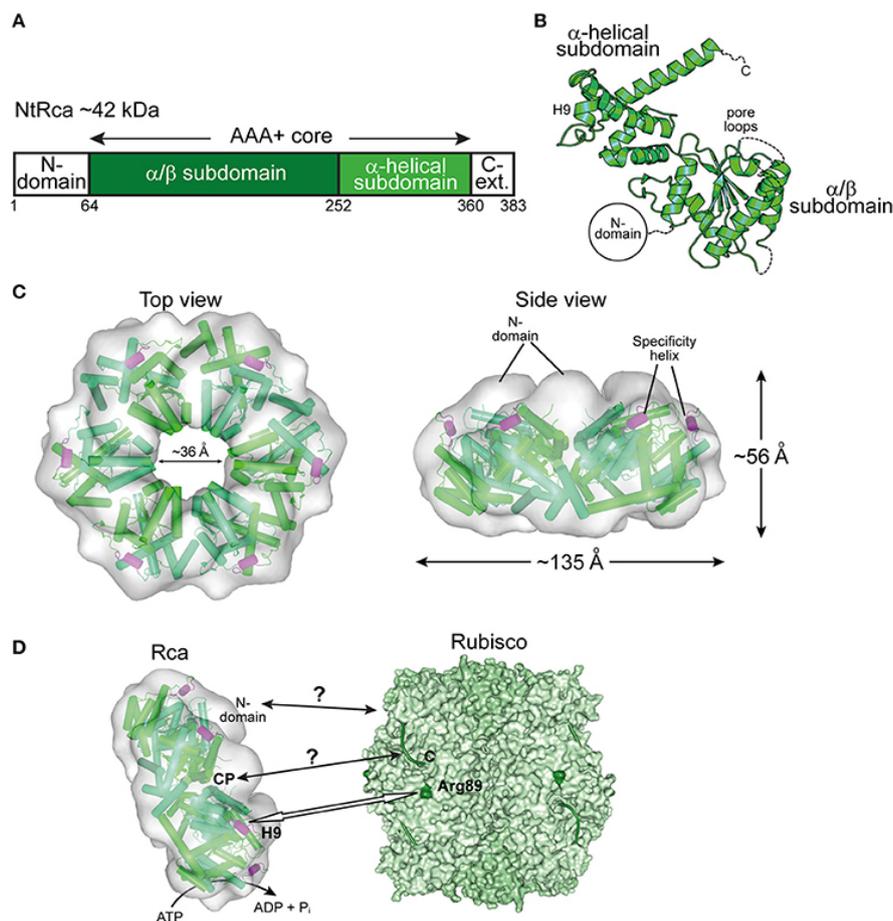


Figure 18: The eukaryotic Rca of the green-type form IB Rubisco. (A) Schematic representation of the domain structure of Rca from *N. tabacum* (*NtRca*). (B) Crystal structure of the monomer (PDB: 3T15) shown in ribbon representation. (C) Top and side views of the *NtRca* hexameric model (PDB: 3ZW6) superposed on the electron microscopic reconstruction (EMDB: 1940). The unfilled electron density at the top of the hexamer probably represents the N-domains. Alternating subunits are shown in two shades of green

and the specificity helix (H9) in purple. (D) Model of the mechanism of eukaryotic Rca from green-type form IB Rubisco. Rubisco (PDB: 1EJ7) is shown in surface representation with the RbcL and RbcS subunits in different shades of green. The RbcL C-termini are shown as green lines. Figure adapted from (Bhat et al., 2017b).

Plant Rca has some characteristic features. A conspicuous difference to other Rca proteins is the presence of long extensions at both chain termini in addition to the AAA+ core (Fig. 18). The N-terminal domain of ~70 residues in plant Rca is required for the interaction with Rubisco (Esau et al., 1996; Stotz et al., 2011; van de Loo and Salvucci, 1996). However, this domain is either missing or is disordered in the available structures. Higher plants express two Rca isoforms, either from separate genes or by alternate splicing (Portis Jr et al., 2008). The α -isoform possesses a long C-terminal extension, containing two cysteine residues, which undergo reversible oxidation (Zhang and Portis, 1999). A disulfide bond is thought to form between the cysteine residues during the night, to inhibit ATP binding and thus Rubisco reactivation when the production of reducing agents is inefficient due to lack of the light-dependent photosynthetic reactions (Carmo-Silva and Salvucci, 2013; Gontero and Salvucci, 2014; Portis, 2003; Portis Jr et al., 2008; Shen and Ogren, 1992; Wang and Portis, 2006; Zhang and Portis, 1999; Zhang et al., 2001). In addition to redox regulation, plant Rca activity is regulated by the ATP/ADP ratio and the Mg^{2+} concentration in chloroplasts. Although plant Rca undergoes dynamic assembly into a range of oligomeric states, the hexamer is thought to be the major functional state. ATP shifts the equilibrium towards functional hexamer, while the presence of ADP destabilizes Rca hexamer (Stotz et al., 2011). The Mg^{2+} concentration in the chloroplast stroma fluctuates in response to the presence of light. At high Mg^{2+} concentration, the catalytic rate of *Nt*Rca has been shown to increase approximately eight-fold (Hazra et al., 2015). These metabolites appear to control Rca activity in response to environmental changes.

In the available crystal structures of Rca from *Arabidopsis thaliana* (*AtRca*) and *Nicotiana tabacum* (*NtRca*), a short helical insertion in the C-terminal domain of the AAA⁺ core, helix 9, was observed. Mutational analysis suggested that helix 9 interacts directly with the Rubisco surface close to RbcL residues 89 and 94. The data were consistent with salt bridges in this interface, however with different orientations between solanaceous and non-solanaceous Rubisco proteins (Li et al., 2005; Portis Jr et al., 2008; Wachter et al., 2013). Specifically, the salt bridge was suggested to form between *AtRca*-K313 and *AtRbcL*-D94 (non-solanaceae) or *NtRca*-D316 and *NtRbcL*-K94 (solanaceae). Helix 9, called specificity helix, together with the N-terminal domain, was proposed to mediate species recognition of Rca to its cognate Rubisco. In both *NtRca* and *AtRca* crystal structures, which exhibit helical subunit arrangements, the central pore has an apparent diameter larger than 30 Å, probably due to the disordered structure of the pore loops (Hasse et al., 2015; Stotz et al., 2011). The *NtRca* hexamer structure solved by negative stain electron microscopy also shows a large central pore with a diameter of 36 Å. The structure reveals additional densities which probably represents the missing N-domains in crystal structure (Stotz et al., 2011). A canonical pore-loop motif containing an aromatic residue is not present in plant Rca sequences. Based on the structures, three conserved loops point towards the central pore, which were named pore-1, pore-2 and pore-3. Mutations in each of these loops, respectively, impair reactivation (Stotz et al., 2011).

A close sequence homolog to the Rca AAA⁺ core has been found in some cyanobacteria containing form IB Rubisco, named Rca-like protein (Li et al., 1999). In contrast to plant Rca, Rca-like proteins have no additional N-terminal domain, instead they have a small domain structurally resembling the small Rubisco subunit (SSU-like, SSUL) at the C-terminus. Whether Rca-like proteins function as Rubisco activases was unclear at the outset of this study.

3 Aim of this study

The human population currently stands at ~7.9 billion, and is expected to grow to ~10 billion by 2050. Recent forecasts suggest that global food production will need to rise by more than 25% to meet the increasing demand (Hauser et al., 2015). Engineering a catalytically more efficient Rubisco is a long-standing goal in the field. There are several promising strategies towards that goal, including implementation of a carboxysome-like CCM into crop plants and engineering a more efficient Rca. However, the limited understanding of carboxysome biogenesis and the Rca reactivation mechanism hinders a rational engineering approach.

The aim of this study was to provide insights into the biogenesis of carboxysomes and the Rca reactivation mechanism. We focused on the β -carboxysome, which encloses form IB Rubisco, the closest relative to plant Rubisco. Specifically, we structurally and functionally characterized the scaffolding protein CcmM from the cyanobacterium *Synechococcus elongatus* PCC7942 and its interaction with two cognate and essential enzymes in β -carboxysomes, Rubisco and the β -carbonic anhydrase, CcaA. Furthermore, we studied the function of the cyanobacterial Rca-homolog from *Nostoc* sp. PCC 7120 (Rca-like) and the Rcas of plants. Mutational analysis suggested that form IB Rubiscos have a conserved reactivation mechanism.

4 Publications

4.1 Rubisco condensate formation by CcmM in β -carboxysome biogenesis

Rubisco condensate formation by CcmM in β -carboxysome biogenesis

H. Wang^{1,4}, X. Yan^{1,4}, H. Aigner^{1,3}, A. Bracher¹, N. D. Nguyen², W. Y. Hee², B. M. Long², G. D. Price², F. U. Hartl¹ & M. Hayer-Hartl^{1*}

Cells use compartmentalization of enzymes as a strategy to regulate metabolic pathways and increase their efficiency¹. The α - and β -carboxysomes of cyanobacteria contain ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco)—a complex of eight large (RbcL) and eight small (RbcS) subunits—and carbonic anhydrase^{2–4}. As HCO_3^- can diffuse through the proteinaceous carboxysome shell but CO_2 cannot⁵, carbonic anhydrase generates high concentrations of CO_2 for carbon fixation by Rubisco⁶. The shell also prevents access to reducing agents, generating an oxidizing environment^{7–9}. The formation of β -carboxysomes involves the aggregation of Rubisco by the protein CcmM¹⁰, which exists in two forms: full-length CcmM (M58 in *Synechococcus elongatus* PCC7942), which contains a carbonic anhydrase-like domain⁸ followed by three Rubisco small subunit-like (SSUL) modules connected by flexible linkers; and M35, which lacks the carbonic anhydrase-like domain¹¹. It has long been speculated that the SSUL modules interact with Rubisco by replacing RbcS^{2–4}. Here we have reconstituted the Rubisco–CcmM complex and solved its structure. Contrary to expectation, the SSUL modules do not replace RbcS, but bind close to the equatorial region of Rubisco between RbcL dimers, linking Rubisco molecules and inducing phase separation into a liquid-like matrix. Disulfide bond formation in SSUL increases the network flexibility and is required for carboxysome function in vivo. Notably, the formation of the liquid-like condensate of Rubisco is mediated by dynamic interactions with the SSUL domains, rather than by low-complexity sequences, which typically mediate liquid–liquid phase separation in eukaryotes^{12,13}. Indeed, within the pyrenoids of eukaryotic algae, the functional homologues of carboxysomes, Rubisco adopts a liquid-like state by interacting with the intrinsically disordered protein EPYC1¹⁴. Understanding carboxysome biogenesis will be important for efforts to engineer CO_2 -concentrating mechanisms in plants^{15–19}.

To understand the function of CcmM in β -carboxysome biogenesis, we solved the crystal structure of the first SSUL domain (SSUL1) of CcmM from *S. elongatus* PCC7942 (Se7942) at 1.65 Å resolution (Fig. 1a, Extended Data Table 1a). Although the two chains of SSUL1 in the asymmetric unit of the crystal structure form a dimer, the protein is monomeric in solution (Extended Data Fig. 1a, b). Whereas the backbone of SSUL1 resembles RbcS (PDB 1RSC; C α root mean square deviation (r.m.s.d.) 1.61–1.62 Å; Fig. 1b), residues in RbcS that make critical contacts with RbcL in the holoenzyme (PDB 1RBL) are not conserved in the SSUL modules (Extended Data Fig. 1c). A notable feature in SSUL1 is a short helical insertion ($\alpha 2$, residues 250–254) after β -strand βA , which replaces the A–B loop in RbcS (Fig. 1b). Besides helix $\alpha 2$, the exposed face of the β -sheet and the loop that follows helix $\alpha 2$ are highly conserved among SSUL sequences (Fig. 1c). The conserved SSUL surface contains two clusters of charged residues (Extended Data Fig. 1d). Another distinctive feature of SSUL1 is a disulfide bond between Cys261 and Cys279 (Fig. 1a). These cysteines are conserved in SSUL2 but not in SSUL3 (Extended Data Fig. 1c), suggesting that the function of the M35 and M58 forms of CcmM might

be redox-regulated. We also solved the crystal structure of reduced SSUL1 (SSUL1_{red}) at 1.2 Å resolution (Fig. 1d, Extended Data Fig. 1e, Extended Data Table 1a). Except for the absence of the disulfide bond, the structures of SSUL1_{red} and oxidized SSUL1 (SSUL1_{ox}) are highly similar (C α r.m.s.d. 0.74–0.94 Å; Extended Data Fig. 1f).

To investigate how Rubisco complexes are linked to form β -carboxysomes, we performed in vitro experiments with purified Rubisco (RbcL₈S₈) and M35 from Se7942 (Extended Data Fig. 2a). Reduced and oxidized M35 (M35_{red} and M35_{ox}, respectively) were chosen specifically to analyse the role of the SSUL domains in this process. Unexpectedly, when the proteins were combined at a molar ratio of Rubisco:M35 of 1:8, the solution became turbid, suggesting the formation of aggregates. By monitoring turbidity, we found that the Rubisco–M35_{red} complex formed with a half-life ($t_{1/2}$) of about 15 s at 25 °C, and an apparent affinity (K_D^{app}) of M35 for Rubisco of approximately 0.2 μM . Formation of the Rubisco–M35_{ox} complex was about four times slower ($t_{1/2} \approx 50$ s) with K_D^{app} increased to approximately 1.0 μM (Fig. 1e, f). Rubisco in complex with M35 is functionally active (Extended Data Fig. 2b). M35 did not interact with RbcL₈ or RbcS alone (Fig. 1e), or with chaperone-bound assembly intermediates^{20–22} (Extended Data Fig. 2c), ensuring that the Rubisco holoenzyme is packaged into carboxysomes.

We next generated variants consisting of one or two SSUL modules (Fig. 1g, Extended Data Fig. 2a) and performed turbidity assays under reducing conditions. No network formation was observed with the single SSUL constructs, M13-1, M13-2 or M13-3, either alone (Fig. 1h) or in combination (Extended Data Fig. 2d). M24-1/2 (SSUL1 and SSUL2) mediated network formation, albeit with reduced efficiency ($t_{1/2} \approx 40$ s) compared to M35_{red}, and M24-2/3 (SSUL2 and SSUL3) failed to induce turbidity (Fig. 1h). To estimate the relative affinities of the SSUL modules for Rubisco, we analysed their ability to displace M35_{red} from Rubisco–M35 complexes. Addition of excess M13-2 (25 μM) resulted in loss of turbidity within seconds ($t_{1/2} \approx 7$ s; Fig. 1i), indicating that the M35–Rubisco interaction is highly dynamic. Dual-colour fluorescence cross-correlation spectroscopy (dcFCCS) using M35 N-terminally labelled with the fluorophore NT-650-NHS (NT650–M35) and Rubisco labelled with Alexa Fluor 532 NHS ester (A532–Rubisco) confirmed that M35 was displaced from Rubisco by excess M13-2 (Extended Data Fig. 2e, f). M13-1 was only slightly less effective in dispersing the Rubisco–M35 network, but dissociation was about three times slower than with M13-2 (Fig. 1i). By contrast, excess M13-3 could not completely dissociate the Rubisco–M35 complex ($t_{1/2} \approx 20$ s). Formation of the Rubisco network is therefore predominantly mediated by SSUL1 and SSUL2.

The interaction of M35_{red} with Rubisco was salt sensitive, consistent with the conserved SSUL surface being highly charged (Extended Data Fig. 1d). Binding was efficient at 50 mM KCl, as used above, but decreased at 100 mM KCl (Extended Data Fig. 3a). Interaction at physiological salt concentration was observed at fourfold higher concentrations of Rubisco and M35 (Extended Data Fig. 3b). The salt dependence of M35_{ox} was more pronounced, with addition of a reducing agent restoring the function of M35 (Extended Data Fig. 3c, d).

¹Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, Martinsried, Germany. ²Australian Research Council Centre of Excellence for Translational Photosynthesis, Research School of Biology, The Australian National University, Canberra, Australian Capital Territory, Australia. ³Present address: KWS SAAT SE, Einbeck, Germany. ⁴These authors contributed equally. H. Wang, X. Yan. *e-mail: mharti@biochem.mpg.de

The interactions of Rubisco with M24-1/2 and M24-2/3 were also salt sensitive (Extended Data Fig. 3e, f). M24-2/3 interacted with Rubisco at very low salt concentrations, indicating that all SSUL modules can participate in network formation.

Liquid–liquid phase separation (LLPS) of proteins, mediated by low-complexity amino acid sequences, underlies the formation of non-membrane-bound compartments in eukaryotes^{12,13,23,24}. The dynamic nature of the Rubisco–M35 network suggested that these proteins undergo LLPS. To test this possibility, we analysed the interaction of NT650–M35_{red} and A532–Rubisco by fluorescence microscopy. The proteins on their own (mixed 1:10 with unlabelled protein) were diffusely distributed (Fig. 2a). Combining Rubisco (0.25 μ M) and M35_{red} (1 μ M) induced demixing into round fluorescent condensates with an average Feret's diameter of about 1.3 μ m (Fig. 2a, Extended Data Fig. 4a). The fusion of these condensates is characteristic of liquid droplets²⁵ (Extended Data Fig. 4b). LLPS of Rubisco with M35_{ox} was considerably less efficient than with M35_{red}, with fewer and smaller droplets forming in a background of diffuse M35 (Extended Data Fig. 4c). Fluorescence recovery after photobleaching (FRAP) of the droplets showed a slower rate of recovery with M35_{red} ($t_{1/2} \approx 130$ s) than with M35_{ox} ($t_{1/2} \approx 70$ s; Fig. 2b, c). Fluorescence recovery with M35_{ox} was incomplete (Fig. 2c), presumably being limited by the rate of diffusion of unbleached M35_{ox} from the bulk into the smaller droplets. The addition of an excess of unlabelled M13-2 dissolved the droplets (Extended Data Fig. 4d). Thus, multivalent M35 condenses Rubisco into a liquid matrix, reminiscent of the behaviour of Rubisco in the pyrenoid of green algae¹⁴. The Rubisco condensate is more dynamic under oxidizing conditions, the presumed state in the carboxysome^{7–9}.

To investigate whether carboxysome function is redox-regulated, we expressed either wild-type or cysteine mutants of CcmM (Fig. 3a) in the carboxysome-less Δ *cmmM* strain of *Se7942*, which requires high CO₂ for growth¹⁰. Expression of plasmid-encoded wild-type CcmM rescued this phenotype (Fig. 3b, Extended Data Fig. 5a). Disruption of the disulfide bond in either SSUL1 or SSUL2 (mutants C279S and C395S, respectively; Fig. 3a), resulted in an approximately 10- to 20-fold increase in CO₂ requirement and an approximately 2- to 2.5-fold increase in doubling time (Fig. 3b, Extended Data Fig. 5a). A further increase in CO₂ requirement and an approximately fourfold increase in doubling time was observed when the disulfide bonds were disrupted in both SSUL1 and SSUL2 (CcmM-4S; Fig. 3b, Extended Data Fig. 5a). The amount of Rubisco in strains expressing the cysteine mutants was similar to that in strains expressing wild-type CcmM, as observed upon fractionation into soluble and carboxysome-enriched (pellet) fractions (Fig. 3c, Extended Data Fig. 5b), indicating that the requirement for extra CO₂ and slower growth are due to dysfunction of the CO₂-concentrating mechanism. Analysis of cells expressing the cysteine mutants by electron microscopy showed abnormal carboxysomes with an increased length-to-width ratio in some cells (Fig. 3d, Extended Data Fig. 5c). Furthermore, the number of carboxysomes per cell was significantly ($P \leq 0.05$) reduced in cells expressing the C395S and CcmM-4S mutants (Extended Data Fig. 5d). Together, these results suggest that disulfide-bond formation in the SSUL modules is essential for carboxysome biogenesis.

To define the structural basis of the M35_{red}–Rubisco interaction, we next performed cryo-electron microscopy (cryo-EM) and single-particle analysis. Reference-free 2D class averages revealed an additional density at the equator of Rubisco (Extended Data Fig. 6a, b). The extra density appeared to be twofold symmetric along the dyad axis between RbcL anti-parallel dimers in the D4 symmetry-averaged map (Extended Data Fig. 6c). However, docking of two SSUL1 molecules into this density showed an overlap of their α 1 helices, suggesting that binding of one module blocks the symmetry-related binding site. This is consistent with the SSUL density values in the D4 map being approximately half as high as those for the RbcL subunits (Extended Data Fig. 6c). Accordingly, we observed density for only one SSUL module per symmetry-related binding site in a focused 3D classification without symmetry (C1 map, Extended Data Fig. 6d, left). To

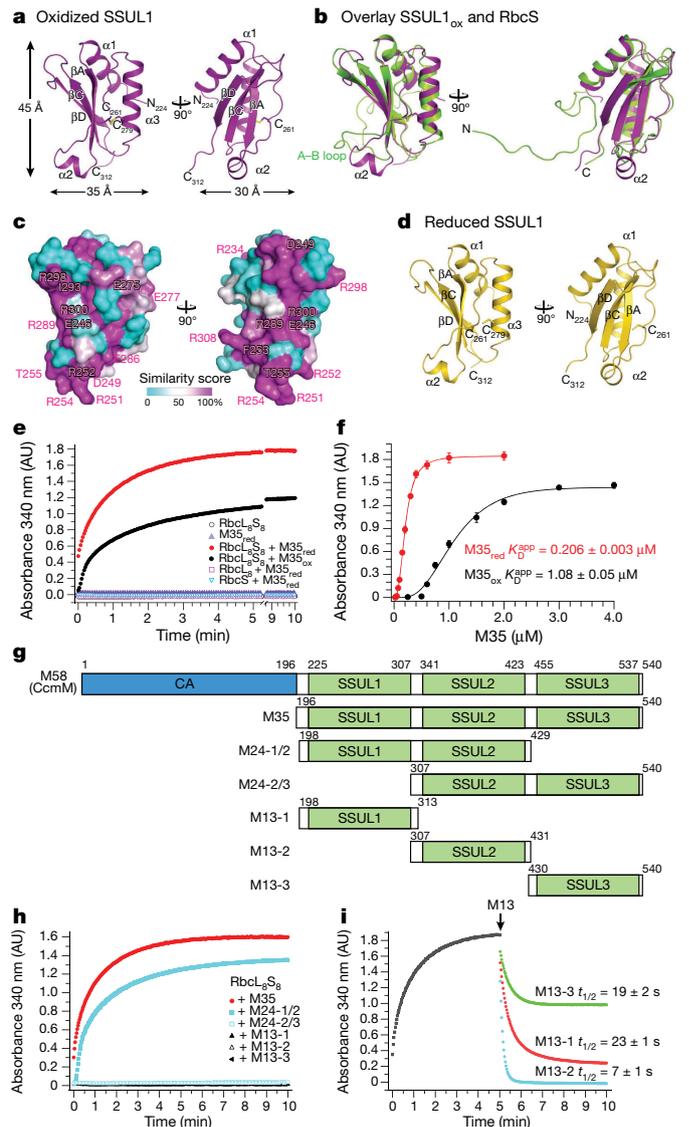


Fig. 1 | Crystal structure of SSUL1 and formation of Rubisco network by M35. **a**, Ribbon representation of thiol-oxidized SSUL1 from CcmM of *Se7942*. Two views rotated by 90° are shown. **b**, Alignment of SSUL1 (magenta) with RbcS (green) from *Se6301* (PDB 1RBL). Note that *Se7942* and *Se6301* have identical sequences for Rubisco and CcmM proteins. **c**, Surface conservation in SSUL1, based on an alignment of eight sequences. Orientation as in **a**. Conserved residues in the interface with Rubisco (see Extended Data Fig. 1c) are indicated. **d**, Ribbon representation of thiol-reduced SSUL1. **e**, Kinetics of Rubisco network formation by turbidity assay ($A_{340\text{ nm}}$) in buffer A at 25°C. Data shown for Rubisco (0.25 μ M), RbcL₈ (0.25 μ M) or RbcS (2 μ M) in the presence of M35_{red} or M35_{ox} (2 μ M). Rubisco and M35 alone served as controls. A representative experiment is shown ($n = 3$ independent experiments). **f**, Apparent binding affinity (K_D^{app}) of M35_{red} and M35_{ox} to Rubisco. Network formation of Rubisco was measured as in **e** at increasing concentrations of M35. Turbidity values reached after 10 min are plotted as mean \pm s.d. ($n = 3$ independent experiments). **g**, Schematics of full-length M58 (CcmM) and M35, as well as M35 truncation constructs. **h**, Kinetics of Rubisco network formation analysed as in **e**. Rubisco (0.25 μ M) was combined with M35 (2 μ M) or the M35 variants M13-1 (6 μ M), M13-2 (6 μ M), M13-3 (6 μ M), M24-1/2 (2 μ M), and M24-2/3 (2 μ M). A representative experiment is shown ($n = 3$ independent experiments). **i**, Dissociation of the Rubisco–M35 network. Rubisco (0.25 μ M) was combined with M35_{red} (1 μ M) and network formation monitored by turbidity assay for 5 min, followed by addition of M13-1, M13-2 or M13-3 (25 μ M each). The apparent $t_{1/2}$ of dissociation is indicated. Data are mean \pm s.d. ($n = 3$ independent experiments). AU, arbitrary units.

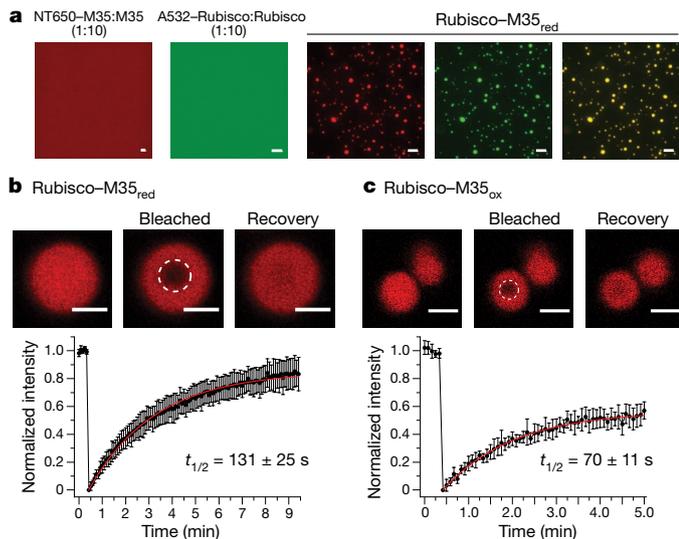


Fig. 2 | Condensation of Rubisco into a network with liquid-like properties. **a**, LLPS of Rubisco–M35. NT650–M35 and A532–Rubisco were mixed 1:10 with the respective unlabelled proteins at total concentrations of M35 and Rubisco of 1 μ M and 0.25 μ M, respectively. The proteins were imaged by fluorescence at 25 $^{\circ}$ C either alone (left two panels) or in combination (right three panels). Droplets contained both M35 and Rubisco, as visualized by merging the fluorescence of NT650–M35 and A532–Rubisco (right). Scale bar, 10 μ m. A representative experiment is shown ($n = 3$ independent experiments). **b**, **c**, FRAP of Rubisco condensates formed by unlabelled Rubisco (0.5 μ M) and M35_{red} (**b**) or M35_{ox} (**c**) (2 μ M, containing 10% of the respective labelled NT650–M35). One representative condensate is shown before and after bleaching. The site of bleaching is marked by a dashed circle. Scale bar, 2 μ m. The change in fluorescence was analysed as a function of time and the $t_{1/2}$ of fluorescence recovery is indicated as mean \pm s.d. ($n = 15$ and $n = 7$ for M35_{red} and M35_{ox}, respectively).

obtain a structural model of the Rubisco–SSUL complex, we extracted units comprising two RbcL, two RbcS and two SSUL for averaging (Extended Data Fig. 6d, main scheme). The resulting density map at 2.77 \AA resolution shows one SSUL module bound asymmetrically in a cleft between RbcL dimers (Fig. 4a, Extended Data Fig. 6e, Extended Data Table 1b). The density of SSUL was well-defined (Extended Data Fig. 7a–d), resulting in a model in which four SSUL modules are bound per Rubisco (Fig. 4a, b).

Two interfaces with Rubisco can be distinguished. The more extensive interface I is formed by contacts of SSUL helix α 2 and the loop that follows, which fill a pocket between RbcL dimers and the β CD loop of RbcS (Fig. 4c, Extended Data Fig. 7c). Specifically, three arginine residues (Arg251, Arg252, Arg254) and one phenylalanine (Phe253) of helix α 2 contact the RbcL subunits that line the groove (RbcL-A, RbcL-B) and RbcS via salt bridges and van der Waals contacts (Fig. 4c). Interface II involves mainly van der Waals interactions between the β C– β D hairpin of SSUL and the RbcL-B subunit (Fig. 4d, Extended Data Fig. 7d). The contact residues in SSUL are highly conserved between SSUL modules (Extended Data Fig. 1c). Mutation of the conserved Arg251 and Arg252 to Asp (interface I) in SSUL1, while preserving the structure of M35, disrupted the function of M35 in Rubisco condensation (Extended Data Fig. 7e–g). Network formation was critically dependent on the interaction of SSUL with RbcS (Extended Data Fig. 7f).

To visualize the distribution of Rubisco in the Rubisco–M35 matrix, we performed cryo-electron tomography (cryo-ET). We observed dense clusters of Rubisco complexes, with a median nearest-neighbour distance of 150 \AA (Extended Data Fig. 8a, b). Accordingly, the flexible linkers between SSUL modules (around 30 residues, spanning up to about 110 \AA), though not resolved, must adopt compact conformations. We also performed cryo-ET on Rubisco–M24-1/2 (at twofold

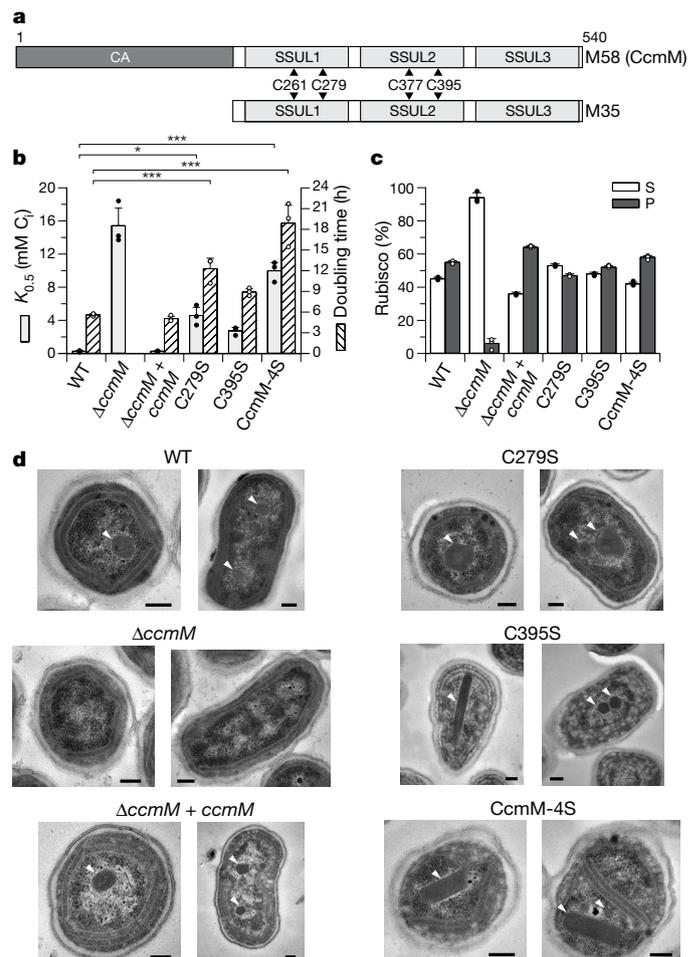


Fig. 3 | Role of disulfide bonds in SSUL in carboxysome function. **a**, Schematic representation of full-length M58 (CcmM) and M35 proteins with the positions of the disulfide bond-forming cysteines in SSUL1 and SSUL2 domains indicated. **b**, Photosynthetic half-saturation constants as $K_{0.5}$ and doubling times for wild-type (WT) *Se7942* or Δ *ccmM* cells expressing wild-type (WT) or cysteine mutant CcmM, grown in air. $*P = 0.01$ – 0.05 ; $***P < 0.001$ (Tukey's multiple comparisons test). P values indicated for wild-type versus mutant cells ($K_{0.5}$: wild-type versus CcmM-4S $P < 0.0001$, wild-type versus C279S $P = 0.0218$; doubling time: wild-type versus CcmM-4S $P < 0.0001$, wild-type versus C279S $P = 0.0003$). Data are mean \pm s.e.m. of three biological repeats. **c**, Rubisco amounts in carboxysome-enriched cell pellets and their corresponding supernatants in wild-type or Δ *ccmM* *Se7942* and cysteine mutants. Data are mean \pm s.e.m. of three biological repeats. **d**, Representative transmission electron micrographs of ultrathin sections of wild-type *Se7942* cells, Δ *ccmM* cells, and Δ *ccmM* cells complemented with wild-type *ccmM* (Δ *ccmM* + *ccmM*) or with the C279S, C395S or CcmM-4S mutant proteins. Arrowheads, carboxysomes; scale bars, 200 nm. Images are representative of carboxysomes from wild-type and mutant cell lines (wild-type $n = 106$, Δ *ccmM* + *ccmM* $n = 78$, C279S $n = 106$, C395S $n = 98$, CcmM-4S $n = 104$; see Extended Data Fig. 5c).

lower protein concentration) and found a similar distribution of particle distances within clusters (Extended Data Fig. 8b, c), ruling out a random distribution.

In contrast to the previous model of β -carboxysome formation, the SSUL modules of CcmM do not displace the RbcS subunit but bind close to the equator of the Rubisco cylinder. The critical interaction of SSUL with the β CD loop of RbcS (Fig. 4e) explains the high specificity of M35 for the holoenzyme. Disulfide bond formation in SSUL attenuates the affinity of M35 for Rubisco, a property that is critical for the function of β -carboxysomes in vivo. The multivalency of M35 and the fluctuating nature of the individual interactions seem to be critical for liquid-like matrix formation, consistent with results obtained with

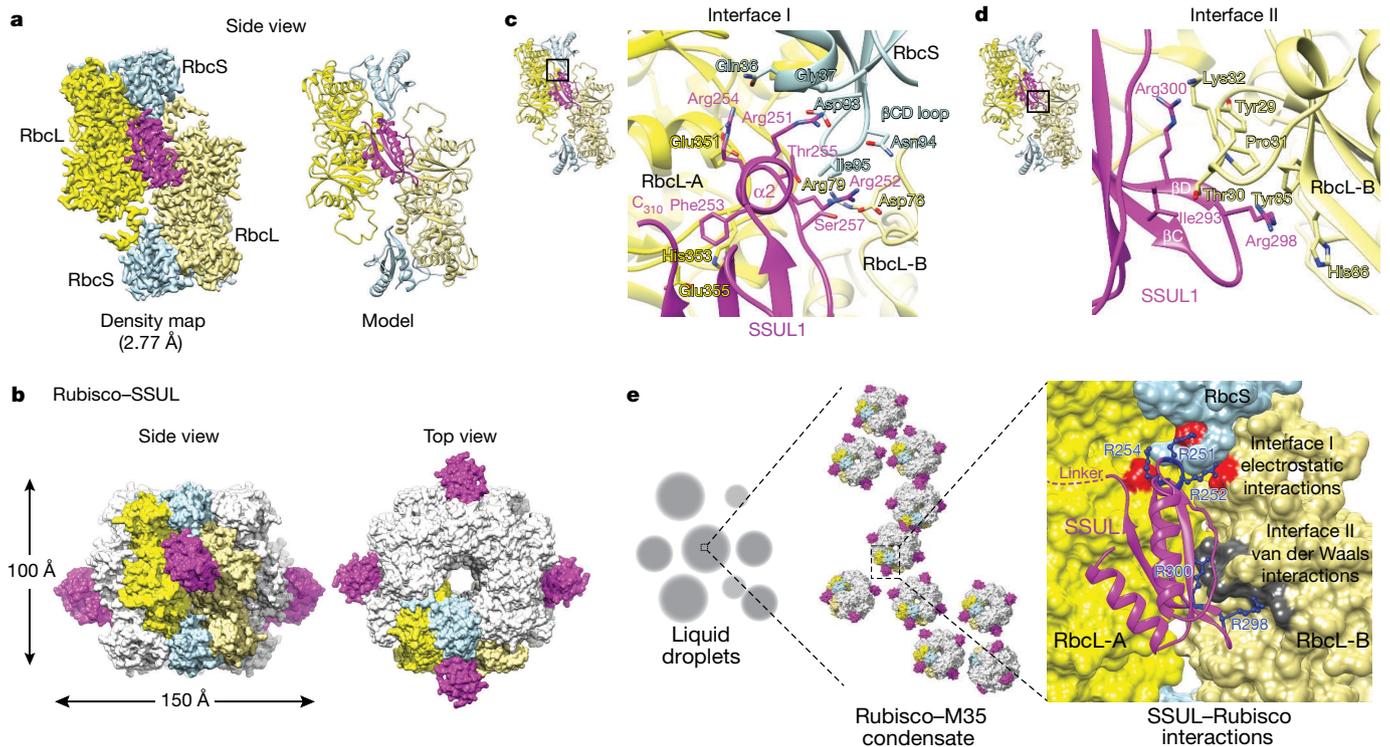


Fig. 4 | Cryo-EM structure of Rubisco-M35 complexes. **a**, Density map of 3D reconstruction from cryo-EM and single-particle analysis of 2RbcL–2RbcS–SSUL units at 2.77 Å resolution (left) (see Extended Data Fig. 6d, e) and the refined structural model in ribbon representation (right). Gold and pale yellow, RbcL subunits in contact with SSUL (from adjacent anti-parallel RbcL dimers); pale blue, RbcS; magenta, SSUL. **b**, Model of the

Rubisco–SSUL complex showing a hypothetical alternating arrangement of four SSUL modules (magenta). Side and top views are shown. One 2RbcL–2RbcS–SSUL unit is coloured as in **a**. **c**, **d**, Structural features of interfaces I (**c**) and II (**d**). Contact residues are indicated (see Extended Data Fig. 7c, d for details). **e**, Model of the liquid-like Rubisco–M35 matrix at different scales, from droplets to clusters to molecular contacts.

modular model proteins^{26,27}. How the CA-like domain of M58 modulates phase separation and how shell formation is coordinated with Rubisco condensation are important questions for future research. Our analysis of the interactions required for the condensation of Rubisco in β -carboxysomes may provide insight into the mechanism of α -carboxysome biogenesis²⁸.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-019-0880-5>.

Received: 2 September 2018; Accepted: 14 December 2018;

Published online 23 January 2019.

- Hinzpeter, F., Gerland, U. & Tostevin, F. Optimal compartmentalization strategies for metabolic microcompartments. *Biophys. J.* **112**, 767–779 (2017).
- Espie, G. S. & Kimber, M. S. Carboxysomes: cyanobacterial RubisCO comes in small packages. *Photosynth. Res.* **109**, 7–20 (2011).
- Rae, B. D., Long, B. M., Badger, M. R. & Price, G. D. 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 (2013).
- Kerfeld, C. A. & Melnicki, M. R. Assembly, function and evolution of cyanobacterial carboxysomes. *Curr. Opin. Plant Biol.* **31**, 66–75 (2016).
- Dou, Z. et al. 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 (2008).
- Price, G. D. & Badger, M. R. Expression of human carbonic anhydrase in the cyanobacterium *Synechococcus* PCC7942 creates a high CO₂-requiring phenotype: evidence for a central role for carboxysomes in the CO₂ concentrating mechanism. *Plant Physiol.* **91**, 505–513 (1989).
- Price, G. D., Coleman, J. R. & Badger, M. R. Association of carbonic anhydrase activity with carboxysomes isolated from the cyanobacterium *Synechococcus* PCC7942. *Plant Physiol.* **100**, 784–793 (1992).
- Peña, K. L., Castel, S. E., de Araujo, C., Espie, G. S. & Kimber, M. S. Structural basis of the oxidative activation of the carboxysomal γ -carbonic anhydrase, CcmM. *Proc. Natl Acad. Sci. USA* **107**, 2455–2460 (2010).

- Chen, A. H., Robinson-Mosher, A., Savage, D. F., Silver, P. A. & Polka, J. K. The bacterial carbon-fixing organelle is formed by shell envelopment of preassembled cargo. *PLoS One* **8**, e76127 (2013).
- Long, B. M., Badger, M. R., Whitney, S. M. & Price, G. D. Analysis of carboxysomes from *Synechococcus* PCC7942 reveals multiple Rubisco complexes with carboxysomal proteins CcmM and CcaA. *J. Biol. Chem.* **282**, 29323–29335 (2007).
- Ludwig, M., Sültemeyer, D. & Price, G. D. Isolation of *ccmKLMN* genes from the marine cyanobacterium, *Synechococcus* sp. PCC7002 (Cyanophyceae), and evidence that CcmM is essential for carboxysome assembly. *J. Phycol.* **36**, 1109–1119 (2000).
- Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid–liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
- Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
- Freeman Rosenzweig, E. S. et al. The eukaryotic CO₂-concentrating organelle is liquid-like and exhibits dynamic reorganization. *Cell* **171**, 148–162.e19 (2017).
- Hanson, M. R., Lin, M. T., Carmo-Silva, A. E. & Parry, M. A. Towards engineering carboxysomes into C3 plants. *Plant J.* **87**, 38–50 (2016).
- Ort, D. R. et al. Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc. Natl Acad. Sci. USA* **112**, 8529–8536 (2015).
- Price, G. D. et al. The cyanobacterial CCM as a source of genes for improving photosynthetic CO₂ fixation in crop species. *J. Exp. Bot.* **64**, 753–768 (2013).
- Zarzycki, J., Axen, S. D., Kinney, J. N. & Kerfeld, C. A. Cyanobacterial-based approaches to improving photosynthesis in plants. *J. Exp. Bot.* **64**, 787–798 (2013).
- Long, B. M. et al. Carboxysome encapsulation of the CO₂-fixing enzyme Rubisco in tobacco chloroplasts. *Nat. Commun.* **9**, 3570 (2018).
- Bracher, A., Starling-Windhof, A., Hartl, F. U. & Hayer-Hartl, M. Crystal structure of a chaperone-bound assembly intermediate of form I Rubisco. *Nat. Struct. Mol. Biol.* **18**, 875–880 (2011).
- Hauser, T. et al. Structure and mechanism of the Rubisco-assembly chaperone Raf1. *Nat. Struct. Mol. Biol.* **22**, 720–728 (2015).
- Aigner, H. et al. Plant RuBisCo assembly in *E. coli* with five chloroplast chaperones including BSD2. *Science* **358**, 1272–1278 (2017).
- Cuevas-Velazquez, C. L. & Dinneny, J. R. Organization out of disorder: liquid–liquid phase separation in plants. *Curr. Opin. Plant Biol.* **45** (Pt A), 68–74 (2018).
- Wang, J. et al. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688–699.e16 (2018).

25. Alberti, S. et al. A user's guide for phase separation assays with purified proteins. *J. Mol. Biol.* **430**, 4806–4820 (2018).
26. Li, P. et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
27. Banani, S. F. et al. Compositional control of phase-separated cellular bodies. *Cell* **166**, 651–663 (2016).
28. Cai, F. et al. Advances in understanding carboxysome assembly in *Prochlorococcus* and *Synechococcus* implicate CsoS2 as a critical component. *Life (Basel)* **5**, 1141–1171 (2015).

Acknowledgements We thank S. Gärtner, R. Lange, N. Wischnewski and L. Rourke for technical assistance; D. Balchin and R. H. Wilson for critically reading the manuscript; M. Strauss, J. Plitzko, F. Beck, S. Albert and Q. Guo (MPIB cryo-EM facility) for input on data collection and processing; the Centre for Advanced Microscopy (ANU) for assistance with electron transmission microscopy of cyanobacteria; and the staff at the MPIB Crystallization and Imaging facilities, and at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. This work was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG) (SFB1035) to M.H.-H. and F.U.H., funding to G.D.P. from the Australian Government through the Australian Research Council Centre of Excellence for Translational Photosynthesis (CE1401000015), and the Minerva foundation of the Max Planck Society (M.H.-H.).

Reviewer information *Nature* thanks J. Dinneny, T. Yeates and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions H.W. and X.Y., together with M.H.-H. and F.U.H., designed the experiments. H.W. performed most of the structural analyses by cryo-EM and tomography, including sample screening, grid preparation, data acquisition and single-particle/tomography reconstructions. X.Y. performed most of the biochemical experiments. H.A. cloned M35 and some of the SSUL modules, performed initial biochemical experiments, and crystallized SSUL1. A.B. solved the crystal structures of SSUL1_{ox} and SSUL1_{red} and supervised the structural analysis by H.W. G.D.P., B.M.L., W.Y.H. and N.D.N. designed the cyanobacterial experiments and generated the pHUE-SSUL1 plasmid, with data collection conducted by N.D.N., W.Y.H. and B.M.L. M.H.-H. conceived the project, participated in data interpretation and wrote the manuscript with contributions from F.U.H. and other authors.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-019-0880-5>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-0880-5>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to M.H.-H.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

Cloning, plasmids and strains. For generation of pHUE-SSUL1, the 267-bp *ccmM*-SSUL1 region was synthesized to include an EcoRI site immediately upstream of base 673 of *ccmM* from *Se7942* (UniProtKB_Q03513) (context gaatccCTCAGTTCC) and a stop codon (taa) plus a HindIII site immediately downstream of base 939 of *ccmM* (context GGTTCGGTTaaaagctt) cloned into EcoRI/HindIII sites of pHUE²⁹. After cleavage of the His₆-tagged product with deubiquitylating enzyme, this yielded a SSUL1 protein of 89 amino acids (amino acids 225–313 of M58) with three additional residues (Ser/Glu/Phe) at the N terminus.

The *ccmM* gene (UniProtKB_A0A0H3JZU5) was amplified from genomic DNA of *Se6301* (*Se6301* is indistinguishable from *Se7942* in the *ccm* operon). Cyanobacteria from a 100-ml culture (OD₆₀₀ ≈ 2) were washed once with fresh medium (BG-11) and then pelleted. Cells were suspended in 20 mM Tris-HCl pH 6.8. We placed 1.2 ml of cell suspension and 300 mg of glass beads (0.5 mm diameter) into a 2-ml tube and crushed them in a bead-beating homogenizer. After a clarifying spin the supernatant was mixed with an equal volume of extraction buffer (20 mM Tris-HCl pH 6.8, 1 M NaCl, 2% SDS). The solution was centrifuged (21,000 r.p.m., 5 min at room temperature (RT)) and the supernatant combined with an equal volume of isopropanol, gently mixed by inversion, and placed on ice for 20 min. After centrifugation (21,000 r.p.m., 5 min, at RT), the pellet was washed with 500 μl 70% (v/v) ethanol and dried (10 min at RT). The dried pellet was dissolved in 50 μl deionized water, centrifuged (21,000 r.p.m., 2 min at RT) and stored at -20°C. From the DNA thus obtained, the full-length *ccmM* sequence was amplified by PCR and cloned into the pHUE vector²⁹. The shorter constructs were prepared by PCR from this plasmid and cloned into the pHUE vector. Plasmids, proteins and strains used or generated in this study are listed in Extended Data Table 2.

SeRbcL₈S₈ preparation. *SeRbcL₈* and *SeRbcS₈* were expressed and purified as previously described^{30,31}. The purified RbcL₈ core complexes and RbcS were mixed at a ratio of 1:16 and incubated at 25°C for 30 min. The RbcL₈S₈ complex was further purified by size-exclusion chromatography (Superdex200, GE Healthcare). Formation of assembled complexes was analysed using 4–12% Tris-glycine gradient gels (Thermo Fisher).

Expression and purification of M35 and variants. M35 and truncation constructs were transformed into *E. coli* BL21 cells. Pre-cultures (5 ml) were grown in LB medium overnight at 37°C and 180 r.p.m. The pre-culture was added to 1 l LB medium containing ampicillin (100 μg ml⁻¹), resulting in a final OD₆₀₀ of 0.05. When the cells reached an OD₆₀₀ of ~0.5 (37°C and 180 r.p.m.), the culture was shifted to 18°C and IPTG (0.25 mM final concentration) and proteins were expressed for ~16 h.

The cell pellets from 3 × 1-l cultures were resuspended in 50 ml buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 2 mM DTT, protease inhibitor, benzamide), and lysis performed using an Emulsiflex homogenizer (Avestin). The cell lysate was cleared by centrifugation (45 min, 40,000 r.p.m.) and the supernatant loaded on a 5-ml Ni-NTA column equilibrated with buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 2 mM DTT). The His₆-ubiquitin fusion protein was eluted with buffer containing 250 mM imidazole. The ubiquitin moiety was cleaved overnight at 4°C using the deubiquitylating enzyme Usp2³², in the presence of 5 mM DTT. After overnight digestion, the buffer was exchanged to 50 mM Tris-HCl pH 8.0, 50 mM NaCl and 2 mM DTT and loaded on a 5-ml Ni-NTA column. The eluted protein was further purified by size-exclusion chromatography (Sephacryl S-100) in buffer A (50 mM Tris-HCl pH 8.0, 50 mM KCl, 10 mM Mg(OAc)₂, 5 mM DTT) containing 10% (v/v) glycerol. Fractions containing pure protein were concentrated, and aliquots were flash-frozen in liquid nitrogen for storage at -80°C.

Oxidized M35 and variants were purified as above in the absence of DTT. Disulfide bond formation occurred through air oxidation. The concentration of free thiol groups in the oxidized M35 was quantified by Ellman's assay with the chemical 5,5'-dithio-bis-(2-nitrobenzoic acid)³³. The purified proteins were >95% pure as judged by SDS-PAGE (Extended Data Fig. 2a).

Turbidity assay. Measurements were performed at 25°C in buffer (50 mM Tris-HCl pH 8.0, 10 mM Mg(OAc)₂) containing different concentrations of KCl and in the absence or presence of 5 mM DTT as indicated in the figure legends. Reactions (100 μl) containing Rubisco (0.25 μM) and different concentrations of M35 constructs were rapidly mixed by vortexing, and absorbance at 340 nm was monitored as a function of time on a Jasco V-560 spectrophotometer.

Rubisco activity assay. Rubisco enzymatic activity was assayed at 25°C as described previously^{21,34}. Rubisco (0.25 μM) was incubated with M35 or variants (2 μM) for 15 min in buffer A unless indicated otherwise. Rubisco carboxylation activity was determined in buffer A containing RuBP (2 mM), NaHCO₃ (30 mM) and a trace amount of radioactive NaH¹⁴CO₃ (0.33–0.5 mM; specific activity 40–60 mCi/mmol). After carbon fixation by Rubisco for 2 min, reactions were stopped

with formic acid, and the amount of fixed carbon quantified using a HITACHI AccuFLEX LSC-8000 scintillation counter. The activity of purified *SeRbcL₈S₈* is set to 100%.

Liquid-liquid phase separation assay. For analysis of LLPS²⁵, reduced or oxidized M35 was labelled at the N terminus with the fluorophore NT-650-NHS (NT650; NanoTemper) according to the manufacturer's instructions (~0.6 and ~0.4 dye molecules bound per M35 molecule, respectively). Rubisco holoenzyme was labelled at the N terminus with the fluorophore Alexa Fluor 532 NHS ester (ThermoFisher) according to the manufacturer's instructions (~2.6 dye molecules bound per Rubisco holoenzyme). NT650-M35 or A532-Rubisco was mixed with unlabelled M35 or unlabelled Rubisco, respectively, at a ratio of 1:10 (1 μM and 0.25 μM total, respectively). Reactions (20 μl) in buffer A with or without DTT with M35 (1 μM) or Rubisco (0.25 μM) alone or in combination were incubated for 10 min at 25°C before analysis by microscopy. For the M13-2 competition experiments, excess M13-2 (25 μM) was added to the preformed Rubisco-M35 liquid droplets. The 20-μl samples were transferred to an uncoated chambered coverslip (μ-Slide angiogenesis; Ibidi) and illuminated with a coolLED pe-4000 LED source at 635 nm and 460 nm for fluorescence imaging and a 12 V/100 W halogen lamp for bright-field imaging. Images were recorded by focusing on the bottom of the plate using a Leica DMI6000B microscope with a Leica DFC900 GT sCMOS camera and a HCX PL APO 63×/1.20 water objective.

FRAP measurements. FRAP³⁵ experiments for Rubisco condensates formed with M35 (containing 10% NT650-M35, final 2 μM) and unlabelled Rubisco (0.5 μM) were carried out with a Leica TCS SP8 AOBS Confocal Laser Scanning Microscope (HCX PL APO 63×/1.2 water objective, PMT detector). Images before and after photobleaching were recorded in a single focal plane at a 5-s time interval. Bleaching was performed using a 488-nm argon laser at 100% intensity in five repeats with a dwell time of 100 ms. The software Fiji was used for image analysis³⁶.

Dual-colour fluorescence cross-correlation spectroscopy (dcFCCS). dcFCCS was performed as previously described³⁷. Rubisco (0.25 μM, containing 2 nM N-terminally labelled Rubisco with Alexa Fluor 532 NHS ester) and M35 (1 μM, containing 2 nM N-terminally labelled M35 with NT-650-NHS) were mixed in buffer A and incubated for 5 min. To follow the displacement of M35 from Rubisco, preformed Rubisco-M35 complexes as above were incubated with an excess of M13-2 (25 μM) for 5 min. The measurements were performed on a MicroTime 200 inverse time-resolved, confocal fluorescence microscope (PicoQuant) for 30 min. The auto-correlation data were fitted with one triplet one-diffusion equation using the Symphotime software (PicoQuant)³⁷.

Generation of *ccmM* mutants for analysis in vivo. Genetic sequences containing codon modifications, giving rise to serine residues instead of cysteine at positions C261, C279, C377 and C395 of *Se7942 CcmM*, were synthesized (Genscript) and inserted into pUC57-Amp^R. Each 989-bp synthetic sequence was supplied with 5' NcoI and 3' NotI restriction sites bordering the SSUL1 and SSUL2 domain regions of the native *ccmM* gene. These fragments were cloned into the pSE41 expression vector harbouring the native *ccmM* gene sequence and transformed into a $\Delta ccmM$ mutant line^{10,38} (Extended Data Table 2).

Phenotypic analysis. Each *Se7942* strain was first grown over 2–3 days in ~3.0% (v/v) CO₂ and then diluted in fresh modified BG11 medium to an optical density (OD₇₃₀) of 0.1 in triplicate 35-ml cultures. Cultures were grown at 30°C, 80 μmol of photons m⁻² s⁻¹ and bubbled with air or 3% CO₂. Hourly culture measurements (OD₇₃₀) were conducted over a period of 8 h and the maximum growth rate was determined from the slope of logarithmic regressions of the data and transformed into doubling times¹⁰.

Wild-type and mutant *Se7942* strains were assessed for differences in photosynthetic affinity for inorganic carbon (C_i) by estimation of O₂ evolution rates over a range of C_i concentrations (0.01–250 mM) using an Oxylab O₂ electrode (Hansatech Instruments). Cultures were pre-grown at 3% CO₂ and cells collected by centrifugation (5,000g, 5 min) and washed twice in an N₂-sparged version of BG-11 medium³⁹, buffered with 50 mM BisTrisPropane-HCl pH 7.9, containing 20 mM NaCl instead of NaNO₃. Washed cells were assayed at a chlorophyll density of 2 μg ml⁻¹ at 30°C, and light maintained at 170 μmol photons m⁻² s⁻¹ in an initial volume of 1.5 ml. Increasing C_i was supplied as NaHCO₃ from concentrated stock solutions prepared in water.

In vivo CcmM and Rubisco analysis. The presence of CcmM and Rubisco proteins in *Se7942* strains was determined by immunoblot, essentially as described⁴⁰ with the following modifications. Cells pre-grown at 3% CO₂ were harvested by centrifugation and resuspended in TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) containing 1% bacterial protease inhibitor cocktail, 700 mM mannitol and r-lysozyme (Merck) and Cell Lytic B (Sigma-Aldrich) and left shaking for 60 min at 37°C in darkness before being harvested by centrifugation and resuspended in a mannitol free TE buffer. Cells were lysed using an Emulsiflex (Avestin) set at 60 p.s.i. Cell debris was removed by centrifugation (3,000g, 10 min) and the clarified lysates were treated with 25 mM MgSO₄ to generate carboxysome-enriched pellet fractions as described previously^{41,42}. Clarified lysate samples containing approximately

100 ng chlorophyll and equivalent volumes of resuspended Mg^{2+} pellets and Mg^{2+} supernatants were diluted in SDS-PAGE sample buffer and proteins were separated on 4–20% SDS-PAGE gels (Bio-Rad) before transfer to polyvinylidene fluoride (PVDF) membranes using a Trans-Blot Turbo blotting apparatus (Bio-Rad). Membranes were probed with polyclonal rabbit antiserum raised against Se7942 M35, and a polyclonal rabbit antiserum raised against *Nicotiana tabacum* Rubisco¹⁰. Bound antibodies were detected using alkaline phosphatase conjugated secondary antibodies and the Attophos AP fluorescent substrate system (Promega) and visualized by fluorescent detection using a ChemiDoc MP gel imaging system (BioRad).

Clarified cell lysates, carboxysome-enriched Mg^{2+} pellets and their corresponding supernatants were also used for Rubisco assays. These were carried out in 250 μ l reaction volumes in assay buffer (100 mM EPPS pH 8.0, 20 mM $MgCl_2$, 20 mM $NaH^{14}CO_3$; specific activity of $NaH^{14}CO_3$ is ~ 800 cpm $nmol^{-1}$) at 25 °C. Rubisco-containing samples (5 μ l) were activated in the reaction mixture for 5 min before initiation of reactions by addition of 5 μ l 20 mM RuBP. Reactions were allowed to proceed for 5 min and terminated by the addition of 250 μ l 10% (v/v) formic acid. Samples were dried overnight at 60 °C before resuspension in 100 μ l water and 1 ml Ultima Gold scintillation cocktail (PerkinElmer) and counted using a Tricarb 2800TR scintillation counter (PerkinElmer).

Size-exclusion chromatography coupled to multi-angle static light scattering (SEC-MALS). M13-1 (~ 80 μ g) was analysed using static and dynamic light scattering by auto-injection of the sample onto a SEC column (5 μ m, 4.3 \times 300 mm column, Wyatt Technology, product #WTC-030N5) at a flow rate of 0.15 ml min^{-1} in buffer (20 mM MOPS-KOH pH 7.5, 100 mM KCl, 5 mM $Mg(OAc)_2$) at 25 °C. The column was in line with the following detectors: a variable UV absorbance detector set at 280 nm (Agilent 1100 series), the DAWN EOS MALS detector (Wyatt Technology, 690 nm laser) and the Optilab rEXTM refractive index detector (Wyatt Technology, 690 nm laser)⁴³. Masses were calculated using ASTRA software (Wyatt Technology) with the dn/dc value set to 0.185 ml g^{-1} . Bovine serum albumin (Thermo) was used as the calibration standard.

Crystallization. The CcmM SSUL1 construct used for crystallization comprises residues 225–313 of CcmM from *S. elongatus* PCC7942. Because of a cloning artefact, the construct contains an N-terminal 3 amino acid extension of the sequence Ser/Glu/Phe.

CcmM SSUL1 oxidized. Crystals were grown by the sitting-drop vapour diffusion method at 4 °C. Drops containing 500 nl of 25 mg ml^{-1} CcmM SSUL1 in 10 mM Tris-HCl pH 6.8, 10 mM NaCl and precipitant were equilibrated against a large volume of precipitant. The precipitant contained 25.5% PEG-3350 and 0.1 M Na-acetate pH 4.5.

CcmM SSUL1 reduced. Crystals were grown by the sitting-drop vapour diffusion method at 4 °C. Drops containing 500 nl of 25 mg ml^{-1} CcmM SSUL1 in 10 mM Tris-HCl pH 6.8, 10 mM NaCl, 10 mM DTT and precipitant were equilibrated against a large volume of precipitant. The precipitant contained 1.95 M ammonium sulfate and 0.1 M Na-acetate pH 4.5.

The crystals were transferred into cryo solution (precipitant/10% (v/v) glycerol) and subsequently vitrified in liquid nitrogen.

Crystallographic data collection and structure solution. The diffraction data of the crystals of CcmM SSUL1 in the disulfide-bonded state were collected at beamline ID23-2 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France.

The diffraction data of the crystals of CcmM SSUL1 in the thiol-reduced state were collected at the automated beamline ID30A-1 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France.

Diffraction data were integrated with XDS and further processed with POINTLESS⁴⁴, AIMLESS⁴⁵ and CTRUNCATE⁴⁶ as implemented in the CCP4i graphical user interface⁴⁷. The structure of disulfide-bonded CcmM SSUL1 was solved by molecular replacement at 1.65 Å resolution using the program MOLREP⁴⁸ using the SSUL domain of *Nostoc sp.* PCC7120 Rubisco activase-like protein (PDB 6HAS, unpublished structure) as a search template. The asymmetric unit contained two copies of the SSUL1 domain. The model was edited manually using Coot⁴⁹. REFMAC5⁵⁰ was used for model refinement. The model of SSUL1_{ox} contains 97 ordered water molecules and exhibits reasonable stereochemistry with 98.8% of the residues in the favoured regions of the Ramachandran plot according to the criteria of MolProbity⁵¹. The structure of SSUL1 in the thiol-reduced state was solved at 1.2 Å resolution by molecular replacement using the disulfide-bonded model and refined as described above. The asymmetric unit contained two copies of the SSUL1 domain, seven sulfate ions and 162 ordered water molecules. According to MolProbity⁵¹, the model of SSUL1_{red} exhibits excellent stereochemistry with 98.9% of the residues in the favoured regions of the Ramachandran plot. Figures were created with PyMol (<https://www.pymol.org/>) and ESPript⁵².

Transmission electron microscopy. Se7942 cells were grown in modified BG-11 medium³⁹ containing 20 mM HEPES-KOH pH 8.0, at 3% CO_2 , 30 °C and approximately 80 μ mol photons $m^{-2} s^{-1}$ over 2–3 days. Cells in 10 ml culture were fixed

for at least 4 h with an equal volume of 8% formaldehyde/5% glutaraldehyde (ProSciTech) and 200 mM piperazine-N,N'-bis(2-ethanesulfonic acid) pH 7.2 (PIPES; Sigma-Aldrich). Cells were centrifuged at 5,000g for 5 min, washed three times with 100 mM PIPES, and then fixed in 1% (w/v) osmium tetroxide for 4 h. Fixed cells were dehydrated through an ethanol series and embedded in LR white resin (ProSciTech). Ultrathin sections were obtained and stained with 2% (w/v) uranyl acetate and lead citrate and viewed using a Hitachi H7100FA TEM.

Cryo-EM for single-particle analysis. Rubisco (6.25 μ M) and M35 (50 μ M) were mixed in buffer A and incubated at 25 °C for 10 min, then 10 nm Au_{13} -BSA tracer (AURION) was added. Holey carbon supported copper grids (Quantifoil R2/1 300 mesh) were plasma-cleaned for 30 s (Harrick Plasma) before using. All cryo-grids were made using a Vitrobot Mark 4 (FEI); 3 μ l solution was added to a grid at 25 °C and 90% humidity, then semi-automatically blotted and plunge-frozen into liquid ethane.

Grids were first screened on a Talos Arctica (FEI) transmission electron microscope (TEM). Selected grids were transferred to a Titan Krios 300 kV TEM (FEI) equipped with GIF Quantum Energy Filters (Gatan), and imaged with a K2 Quantum direct detector (Gatan) operated in movie mode at 0.15 s per frame, with 0.822 Å pixel size and 7 s exposure. SerialEM⁵³ was used to automatically collect 4,723 movies.

MotionCor2⁵⁴ was first applied to the movies with dose-weighting. Particles (7,520) were picked manually and templates for auto-picking were generated from the 2D classification using RELION 2.1⁵⁵. The selected 2D classes were used in Gautomatch (<http://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch>), yielding 988,648 particles in total (Extended Data Fig. 6d). RELION 2.1 was used for further data processing. The complete data set went through one round of 2D classification to remove bad particles, such as ice and particles on carbon, followed by 3D classification of 663,599 particles (Extended Data Fig. 6d). An initial model was generated using RELION and further refined using 83,746 cleaned particles. No symmetry was applied in these procedures. 3D classification of 663,599 particles resulted in four classes, with only one (class 3) showing a preferred orientation (Extended Data Fig. 6d). Class 1, which comprised 29% of particles, was first selected and refined. The refined map showed extra density (equivalent to two SSUL modules associated with Rubisco) between anti-parallel RbcL dimers (Extended Data Fig. 6d, left). To determine how many SSUL domains bind per site, the extra density was extracted and a focused classification carried out with a reference and a mask covering two SSUL modules. This showed that in three major classes only one SSUL domain was bound per site, either above or below the equator of the Rubisco holoenzyme (Extended Data Fig. 6d, left). Next, we extracted units containing 2 RbcL, 2 RbcS and 2 SSUL for classification and refinement (Extended Data Fig. 6d, main scheme). Before subtraction, particles were first aligned using D4 symmetry and then the symmetry was expanded using the `relion_particle_symmetry_expand` command to generate eightfold aligned particles. The two best resolved classes (class v and vi) contained 78,916 and 103,669 particles, respectively, and were used for final particle refinement with local alignment, resulting in 2.77 and 2.84 Å maps, respectively, after post-processing (Extended Data Fig. 6d). Only class v, with the best resolution, was then used for model building. Local resolution was calculated using ResMap⁵⁶ and RELION. The crystal structures of thiol-reduced SSUL1 and RbcL and RbcS (PDB 1RBL) were placed into the density using Chimera (v 1.12). The model was edited with Coot⁴⁹. The model was refined in reciprocal space against a masked map with REFMAC5⁵⁰. Residues with disordered side-chains were modelled as alanine. The program CONTACT (T. Skarzynski, Imperial College, London) as implemented in the CCP4i graphical user interface was used to analyse the intermolecular interactions using 4 Å as the cut-off distance.

Cryo-electron tomography. For analysis of the Rubisco-M35 network, Rubisco (6.25 μ M) was mixed with M35 (50 μ M) in buffer A. For analysis of the Rubisco-M24-1/2 network, Rubisco (6.25 μ M) was mixed with M24-1/2 (50 μ M) in buffer A. The mixtures were incubated for 10 min at 25 °C. The Rubisco-M24-1/2 reaction was diluted twofold before cryo-grid preparation. Cryo-grids were prepared as described above for single-particle analysis.

Rubisco-M35 grids were pre-screened on a Talos Arctica (FEI) TEM and suitable grids then transferred to a Titan Krios 300 kV TEM (FEI) equipped with a GIF Quantum Energy Filters (Gatan), and imaged with a K2 Summit direct detector (Gatan) operated in movie mode at 5 frames per second. Using SerialEM⁵³, dose symmetric tilt-series were recorded from -60° to $+60^\circ$ (in two halves separated at 0) in 2° increments, with a pixel size of 2.234 Å, a target defocus of -4.5 μ m, and a cumulative electron dose of 60 electrons per Å². For the Rubisco-M24-1/2 complex, single-axis tilt-series were recorded automatically using SerialEM, from -60° to $+60^\circ$ (in two halves separated at 0) in 2° increments on Talos Arctica TEM (FEI), equipped with a Falcon 3EC Direct Electron Detector (FEI) with a pixel size of 1.977 Å. Image frames were first aligned using MotionCor2, then etomo (IMOD version 4.10.15)⁵⁷ was used for alignment and generation of a four-times binned tomogram. Template matching was performed using PyTom⁵⁸. The template was generated from the single-particle reconstruction map and low-pass filtered to

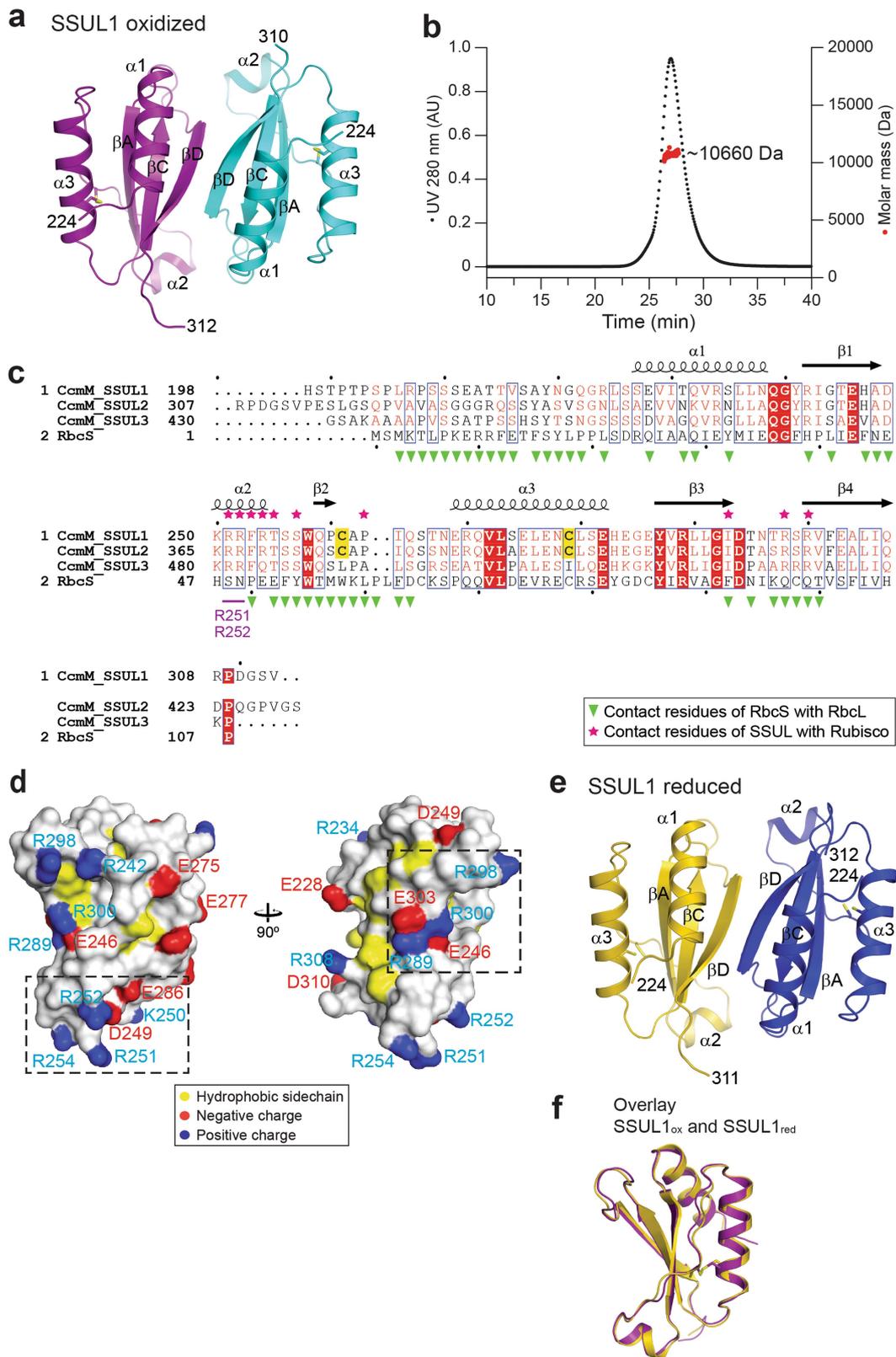
40 Å resolution. Spherical masks were used in template matching. For each tomogram, the highest cross-correlation peaks were exhaustively extracted with a spherical mask of radius of 64 Å. Particles were manually selected, amounting to 2,685 particles for Rubisco-M35 and 1,813 for Rubisco-M24-1/2. Nearest neighbour distances were determined according to the coordinates from template matching with the TOM Toolbox in Matlab and simulated tomogram volumes were generated, according to the coordinates and orientations from template matching. Only particles in the condensate whose distance from the nearest neighbour fitted the plausible range (100–420 Å for Rubisco-M35 and 100–280 Å for Rubisco-M24-1/2) were selected for statistical analysis. Analysis according to Mann-Whitney test showed no significant difference between the measured distances for Rubisco-M35 and Rubisco-M24-1/2 ($P < 0.05$).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The crystallographic models and structure factors for SSUL_{1ox} and SSUL_{1red} have been deposited to wwPDB under accession codes 6HBA (oxidized) and 6HBB (reduced), respectively. The electron density reconstructions and final 2RbcL-2RbcS-SSUL model have been deposited in the Electron Microscopy Data Bank (EMDB) and wwPDB under accession codes EMD-0180 and 6HBC, respectively. Source data for graphs in Figs. 1–3 are provided with the online version of the paper and the source data for the gels shown in Extended Data Fig. 5b are provided in Supplementary Fig. 1.

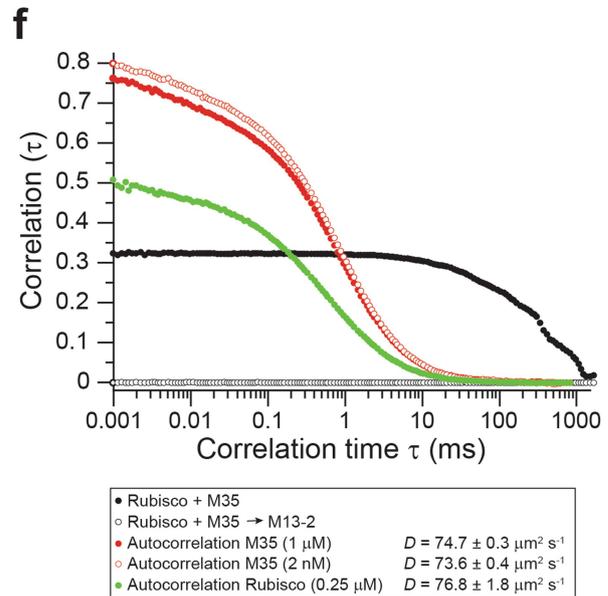
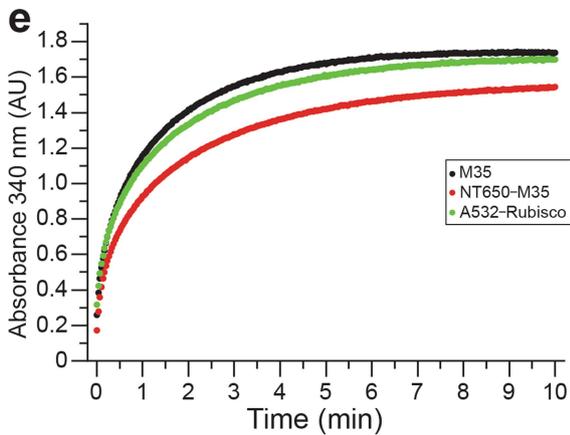
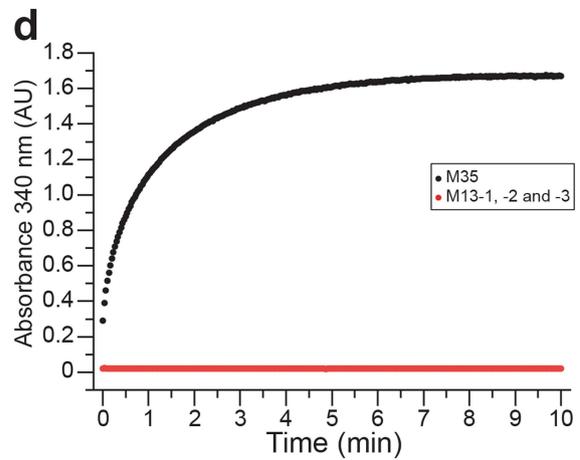
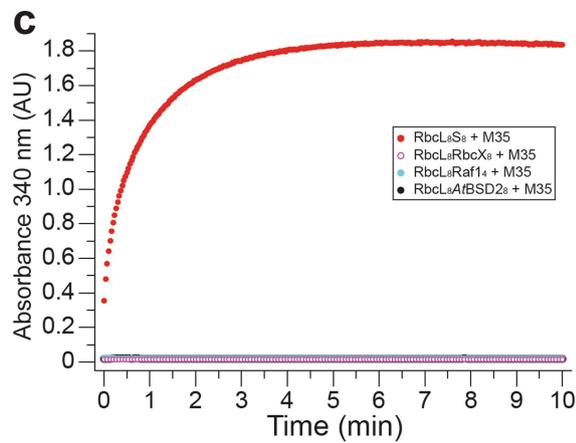
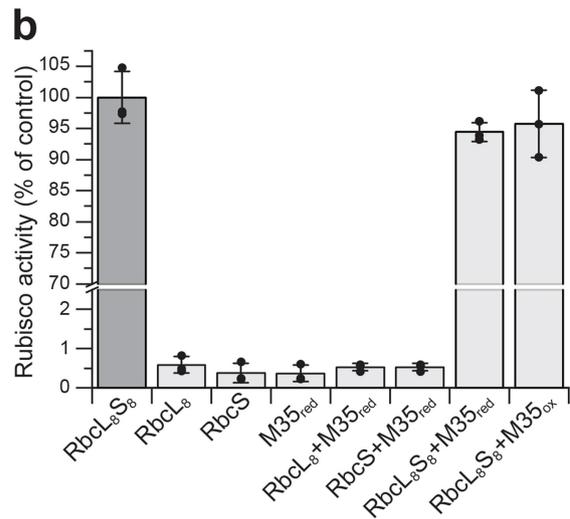
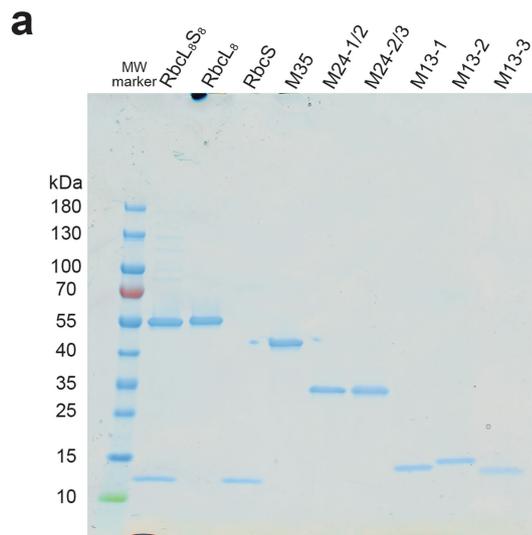
29. Catanzariti, A. M., Soboleva, T. A., Jans, D. A., Board, P. G. & Baker, R. T. An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Sci.* **13**, 1331–1339 (2004).
30. Liu, C. et al. Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* **463**, 197–202 (2010).
31. Saschenbrecker, S. et al. Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell* **129**, 1189–1200 (2007).
32. Baker, R. T. et al. Using deubiquitylating enzymes as research tools. *Methods Enzymol.* **398**, 540–554 (2005).
33. Riddles, P. W., Blakeley, R. L. & Zerner, B. Reassessment of Ellman's reagent. *Methods Enzymol.* **91**, 49–60 (1983).
34. Brinker, A. et al. Dual function of protein confinement in chaperonin-assisted protein folding. *Cell* **107**, 223–233 (2001).
35. Maharana, S. et al. RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **360**, 918–921 (2018).
36. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
37. Gupta, A. J., Haldar, S., Miličić, G., Hartl, F. U. & Hayer-Hartl, M. Active cage mechanism of chaperonin-assisted protein folding demonstrated at single-molecule level. *J. Mol. Biol.* **426**, 2739–2754 (2014).
38. Woodger, F. J., Badger, M. R. & Price, G. D. Sensing of inorganic carbon limitation in *Synechococcus* PCC7942 is correlated with the size of the internal inorganic carbon pool and involves oxygen. *Plant Physiol.* **139**, 1959–1969 (2005).
39. Price, G. D. & Badger, M. R. Ethoxzolamide inhibition of CO₂-dependent photosynthesis in the cyanobacterium *Synechococcus* PCC7942. *Plant Physiol.* **89**, 44–50 (1989).
40. Long, B. M., Tucker, L., Badger, M. R. & Price, G. D. Functional cyanobacterial β-carboxysomes have an absolute requirement for both long and short forms of the CcmM protein. *Plant Physiol.* **153**, 285–293 (2010).
41. Price, G. D. & Badger, M. R. Evidence for the role of carboxysomes in the cyanobacterial CO₂-concentrating mechanism. *Can. J. Bot.* **69**, 963–973 (1991).
42. Price, G. D., Süitemeyer, D., Klughammer, B., Ludwig, M. & Badger, M. R. The functioning of the CO₂ concentrating mechanism in several cyanobacterial strains: a review of general physiological characteristics, genes, proteins, and recent advances. *Can. J. Bot.* **76**, 973–1002 (1998).
43. Wyatt, P. J. Light scattering and the absolute characterization of macromolecules. *Anal. Chim. Acta* **272**, 1–40 (1993).
44. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).
45. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
46. French, G. & Wilson, K. On the treatment of negative intensity observations. *Acta Crystallogr. A* **34**, 517–525 (1978).
47. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1131–1137 (2003).
48. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25 (2010).
49. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
50. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
51. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
52. Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).
53. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
54. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
55. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
56. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
57. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
58. Hrabe, T. et al. PyTom: a python-based toolbox for localization of macromolecules in cryo-electron tomograms and subtomogram analysis. *J. Struct. Biol.* **178**, 177–188 (2012).



Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Crystal structure and sequence analysis of the SSUL domains. **a**, Ribbon representation of the asymmetric unit in the crystal lattice of the SSUL1 domain in the thiol-oxidized state. A view along the approximate dyad axis is shown. The two copies of SSUL1 with nearly identical conformations (with $C\alpha$ r.m.s.d. 0.43 Å) are shown in magenta and cyan. Secondary structure elements and chain termini are indicated. **b**, SEC–MALS analysis of purified SSUL1. The red dotted line across the peak indicates the molar mass and homogeneity of the protein sample. The molar mass is indicated (calculated theoretical mass is 10,537.7 Da). A representative experiment is shown ($n = 2$ independent experiments). **c**, Sequence alignment of SSUL domains 1–3 in CcmM and RbcS from *S. elongatus* PCC7942. Amino acid sequences were aligned using the EBI Clustal-Ω server. Secondary structure elements for the SSUL1 domain are indicated above the sequences. Residues that are similar between SSUL (group 1) and RbcS (group 2) are boxed with blue frames; identical residues are shown in white on a red background. Triangles below

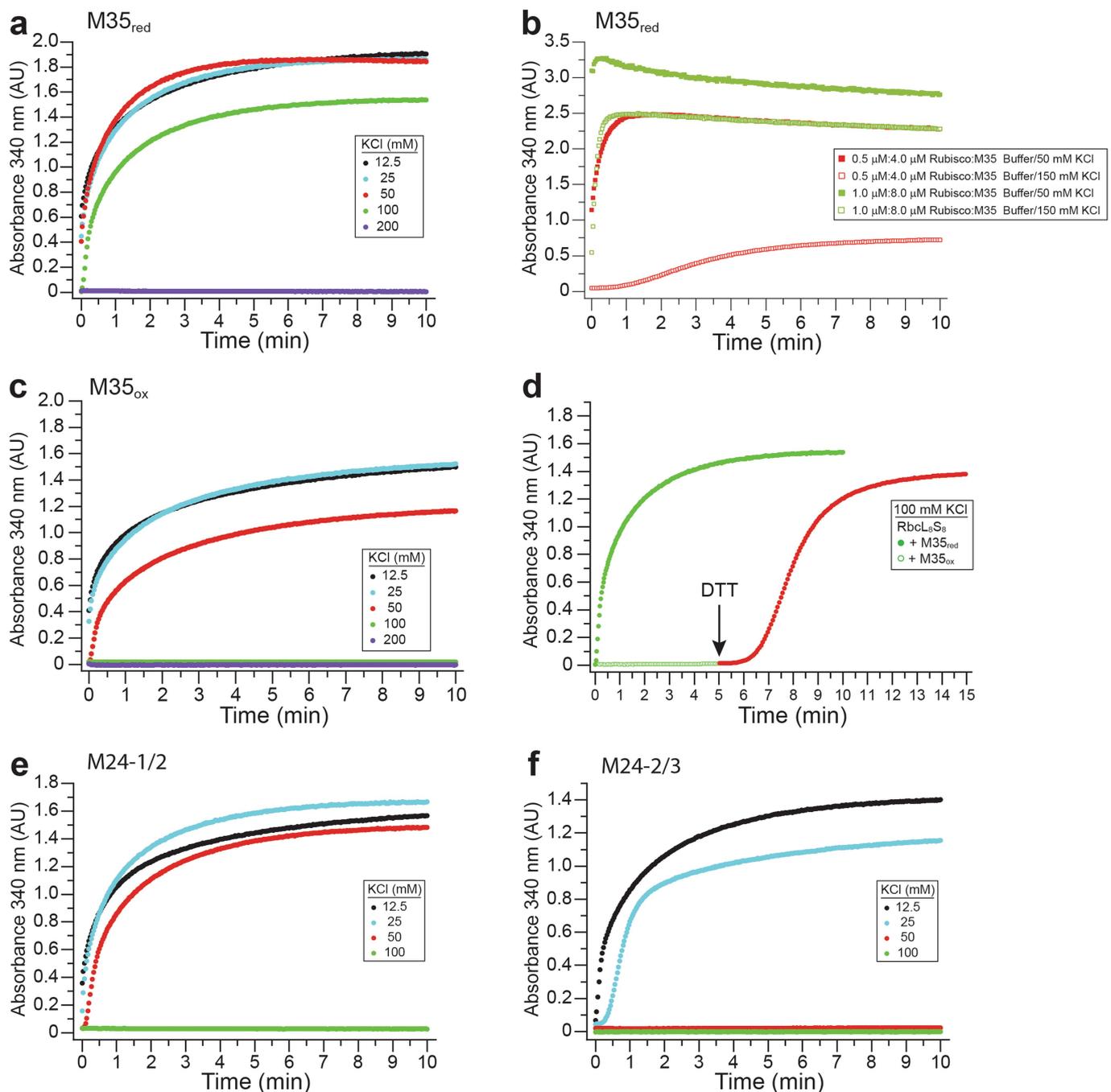
the sequence indicate the contact residues of RbcS with the RbcL₈ core in the Rubisco complex. The mutation site in the SSUL1 module (R251–R252) is indicated. The stars in magenta above the sequence indicate the contact residues of SSUL in the complex with Rubisco. Note that these residues are generally highly conserved in other β-cyanobacteria. The cysteine residues conserved in SSUL1 and SSUL2 are shown in bold with a yellow background. SSUL1 and SSUL2 share 84% identity, while SSUL3 is more divergent (58% identity with SSUL1). **d**, Surface properties of the SSUL1 domain. Hydrophobic side chains are indicated in yellow. Red and blue represent negative and positive charges, respectively. Dashed boxes indicate two areas of charge clusters. **e**, Ribbon representation of the asymmetric unit in the crystal lattice of the SSUL1 domain in the thiol-reduced state. A view along the approximate dyad axis is shown. The two copies of SSUL1 are shown in gold and blue, respectively. Secondary structure elements and chain termini are indicated. **f**, Overlay of SSUL1 structures in the oxidized (magenta) and reduced (gold) forms.



Extended Data Fig. 2 | See next page for caption.

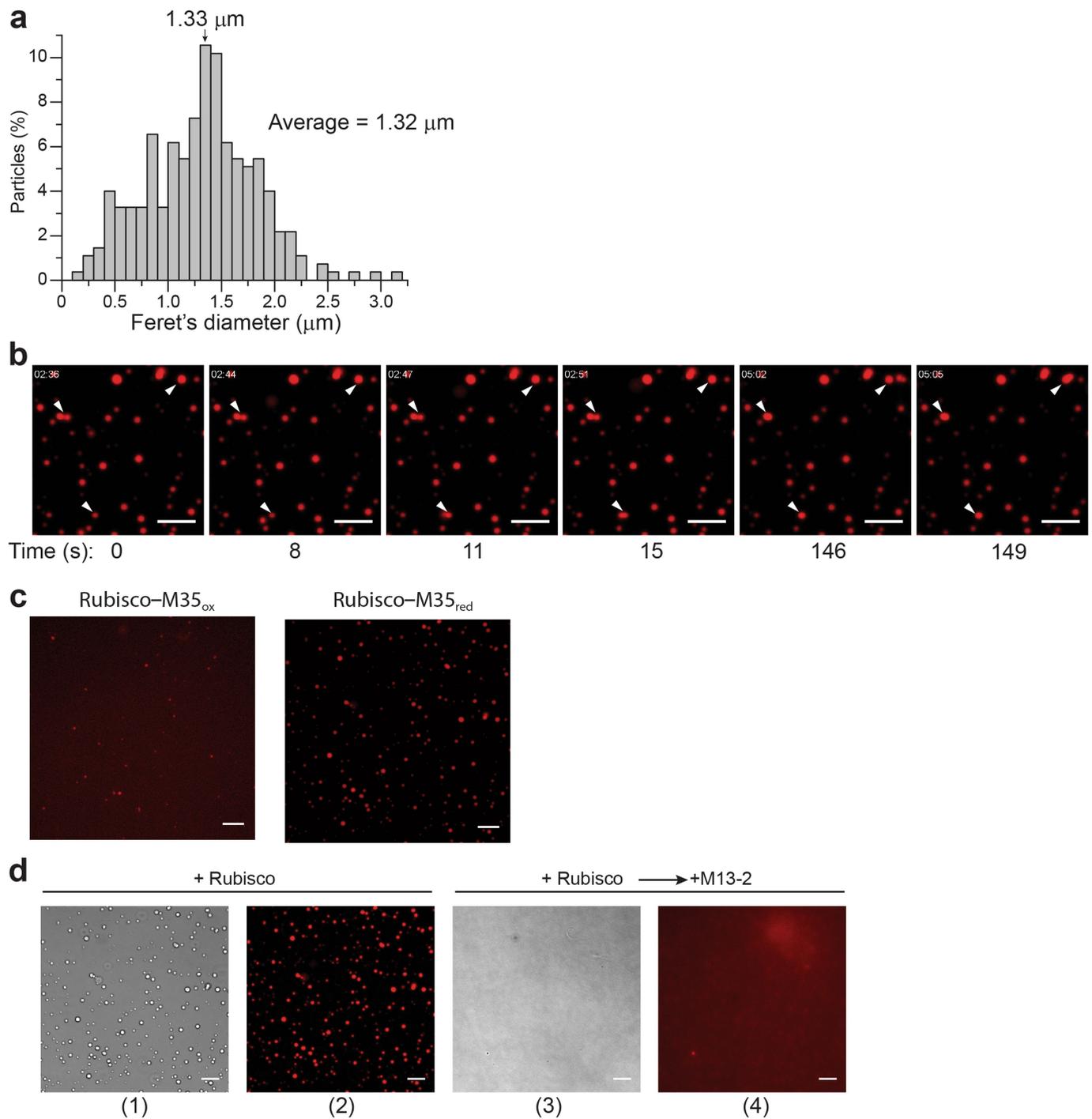
Extended Data Fig. 2 | Properties of the Rubisco–M35 network. **a**, SDS–PAGE of recombinantly expressed and purified proteins. A representative SDS gel is shown ($n = 3$ independent experiments). **b**, Rubisco carboxylation activity of the reactions shown in Fig. 1e, including RbcL₈ core complexes and RbcS in the absence of M35. The activity of purified RbcL₈S₈ is set to 100%, equivalent to a catalytic constant (k_{cat}) of $\sim 4 \text{ s}^{-1}$ per active site (dark grey bar). Data are mean \pm s.d. ($n = 3$ independent experiments). **c**, Kinetics of M35-mediated network formation of Rubisco and Rubisco assembly intermediates by turbidity assay in buffer A at 25 °C. RbcL₈ core complexes (0.25 μM) were incubated with RbcX from *Anabaena* sp. CA (8 μM), or the cognate Raf1 (2 μM) or BSD2 from *Arabidopsis thaliana* (4 μM) in buffer A for 15 min to generate the respective Rubisco assembly intermediates. M35 (2 μM) was then added to start the turbidity measurement at 340 nm. Turbidity of Rubisco with M35 is shown as control. A representative experiment is shown ($n = 3$ independent experiments). **d**, Kinetics of network formation of Rubisco (0.25 μM) with M35 (2 μM) or with combined M13-1, M13-2 and M13-3 (2 μM each) by turbidity assay in buffer A at 25 °C. A representative experiment is shown ($n = 3$ independent experiments). **e**, Kinetics of network formation by turbidity assay at 25 °C of unlabelled

Rubisco (0.25 μM) with NT650–M35 (1 μM) and of unlabelled M35 (1 μM) with A532–Rubisco (0.25 μM). A reaction containing unlabelled Rubisco (0.25 μM) and unlabelled M35 (1 μM) is shown as control. A representative experiment is shown ($n = 3$ independent experiments). **f**, dcFCCS of 0.25 μM Rubisco (containing 2 nM A532–Rubisco) and 1 μM M35 (containing 2 nM NT650–M35). Measurements were taken for 30 min after 5 min incubation in buffer A. Cross-correlation curves are shown. After formation of the Rubisco–M35 network, unlabelled M13-2 (25 μM) was added and cross-correlation measured for 30 min. Auto-correlation curves of 0.25 μM Rubisco (containing 2 nM A532–Rubisco) and 1 μM M35 (containing 2 nM NT650–M35) are also shown. The diffusion coefficients (D) are indicated. The relatively slow rate of diffusion of M35 suggests that the protein is expanded and not as compact as would be expected for a protein of $\sim 35 \text{ kDa}$. To exclude the possibility that the slow diffusion rate of M35 is due to interactions between M35 molecules, we also measured the auto-correlation of M35 at very low concentration (2 nM NT650–M35), where protein–protein interactions are unlikely, and found the same rate of diffusion. Data are mean \pm s.d. ($n = 3$ independent experiments).



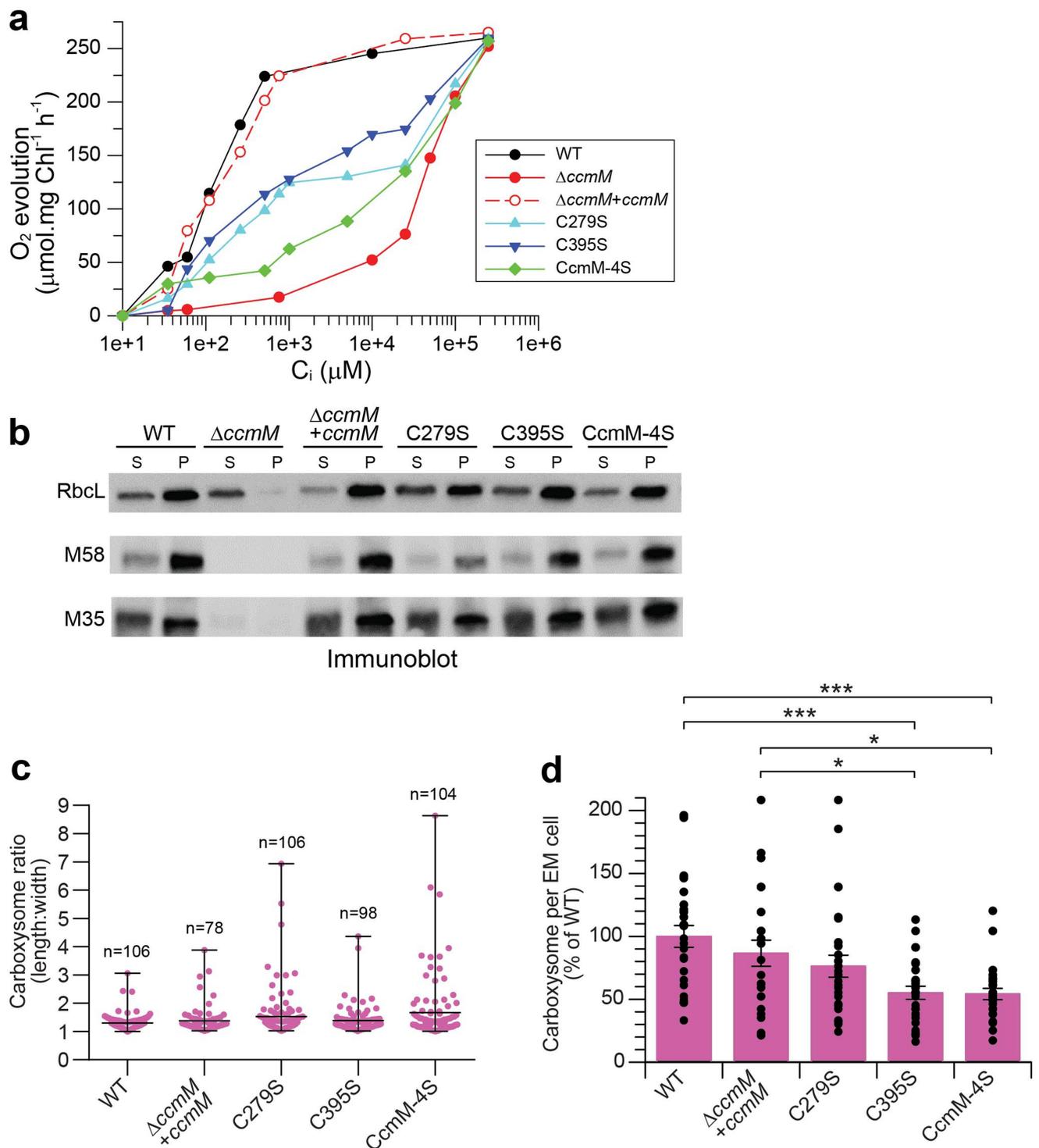
Extended Data Fig. 3 | Salt sensitivity of M35–Rubisco interaction.
a, Kinetics of Rubisco network formation by turbidity assay at different salt concentrations at 25 °C. Rubisco (0.25 μM) in the presence of 2 μM M35_{red} in buffer (50 mM Tris-HCl pH 8.0, 10 mM Mg(OAc)₂, 5 mM DTT) containing different concentrations of KCl at 25 °C. A representative experiment is shown ($n = 3$ independent experiments). **b**, Kinetics of Rubisco network formation as in **a** at 50 mM and 150 mM KCl, and higher concentrations of Rubisco (0.5 or 1.0 μM) and M35 (4.0 or 8.0 μM). A representative experiment is shown ($n = 3$ independent experiments). **c**, Kinetics of Rubisco network formation by turbidity assay at different salt concentrations at 25 °C as in **a** but in the presence of 2 μM M35_{ox}

and in buffer without DTT. A representative experiment is shown ($n = 3$ independent experiments). **d**, M35 function is redox-sensitive. Network formation of Rubisco (0.25 μM) by M35_{ox} (2 μM) was monitored in buffer (50 mM Tris-HCl pH 8.0, 10 mM Mg(OAc)₂) containing 100 mM KCl for 5 min, followed by addition of 5 mM DTT for further 10 min. Rubisco (0.25 μM) with M35_{red} (2 μM) in buffer containing 100 mM KCl and 5 mM DTT is shown as control. A representative experiment is shown ($n = 3$ independent experiments). **e**, **f**, Kinetics of Rubisco network formation by turbidity assay at different salt concentrations at 25 °C as in **a** in the presence of 2 μM reduced M24-1/2 (**e**) or M24-2/3 (**f**). A representative experiment is shown ($n = 3$ independent experiments).



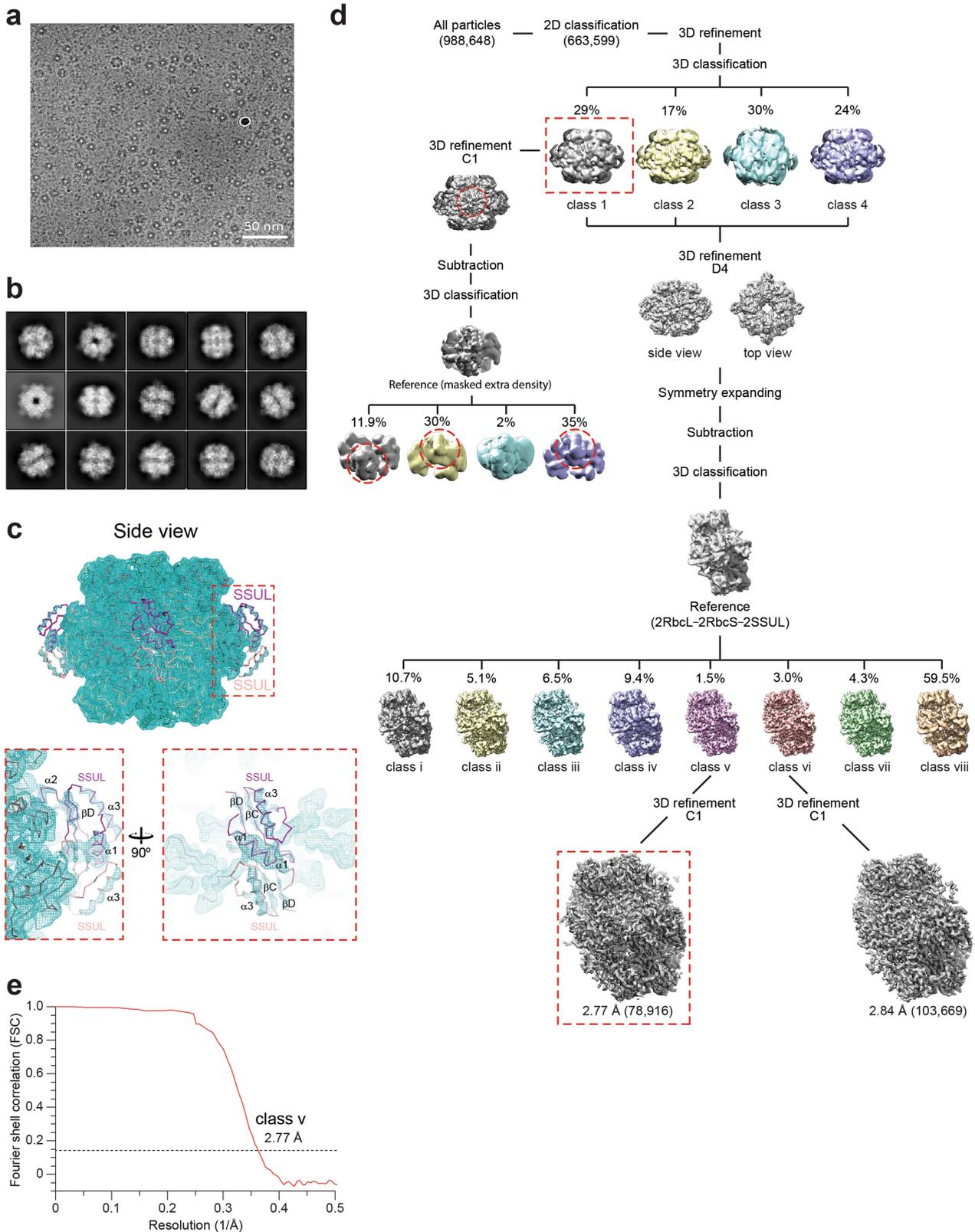
Extended Data Fig. 4 | LLPS of Rubisco by M35_{red} and M35_{ox}. **a**, Size distribution of liquid droplets ($n = 275$) formed by unlabelled Rubisco ($0.25 \mu\text{M}$) and M35 ($1 \mu\text{M}$, containing 10% NT650–M35). The Feret's diameter for the main peak (indicated by arrow) and average size are $1.33 \mu\text{m}$ and $1.32 \mu\text{m}$, respectively. **b**, Time-lapse images of droplet fusion. Reactions containing labelled M35_{red} ($1 \mu\text{M}$, containing 10% NT650–M35) and unlabelled Rubisco ($0.25 \mu\text{M}$) were observed over time. Droplets undergoing fusion are indicated by white arrowheads. Scale bars, $10 \mu\text{m}$. A representative experiment is shown ($n = 3$ independent experiments). **c**, LLPS of Rubisco–M35_{ox} (left) and Rubisco–M35_{red}

(right). NT650–M35_{ox} or NT650–M35_{red} was mixed with unlabelled M35_{ox} or M35_{red}, respectively, at a ratio of 1:10 ($1 \mu\text{M}$ total) and imaged by fluorescence microscopy at 25°C in combination with unlabelled Rubisco ($0.25 \mu\text{M}$). Scale bars, $10 \mu\text{m}$. A representative experiment is shown ($n = 3$ independent experiments). **d**, Rubisco–M35_{red} condensates were generated as in **c** and imaged by fluorescence and bright-field microscopy at 25°C (panels 1 and 2). Dissociation of droplets was observed upon addition of unlabelled M13-2 ($25 \mu\text{M}$) (panels 3 and 4) to preformed droplets as in panels 1 and 2. Scale bars, $10 \mu\text{m}$. A representative experiment is shown ($n = 3$ independent experiments).



Extended Data Fig. 5 | Role of disulfide bonds in SSUL in carboxysome function. **a**, Photosynthetic O₂ evolution in response to external C_i by wild-type *Se7942* and *ccmM* mutant strains. Shown is a representative set of Clarke-type oxygen electrode measurements of C_i-dependent O₂ evolution by different strains of *Se7942* from 10 μM to 250 mM C_i. Cysteine mutants (C279S, C395S and CcmM-4S) had an intermediate CO₂-requiring phenotype, with CcmM-4S having the highest CO₂ requirement. A representative experiment is shown (*n* = 3 biological replicates). **b**, Relative RbcL abundance in carboxysome-enriched pellet fractions in wild-type and *ccmM* mutant strains of *Se7942*. RbcL and CcmM proteins were detected using anti-RbcL and anti-M35 antibodies, respectively. A representative experiment is shown (*n* = 3 biological replicates). **c**, Length:width ratios of carboxysomes in wild-type and *ccmM* mutant strains. Carboxysomes were visualized using TEM, and their lengths and widths measured. These measurements were used to calculate

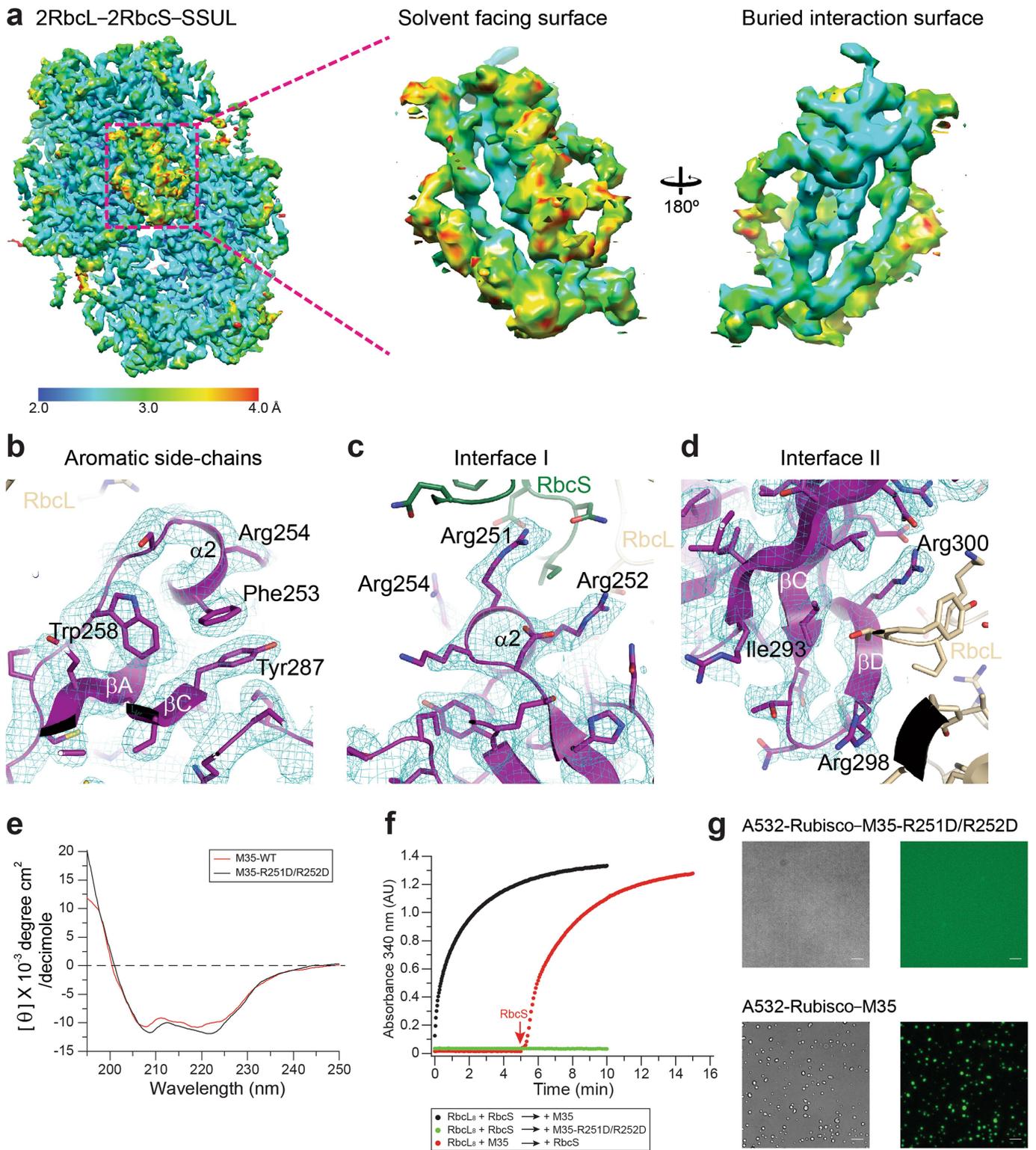
the length:width ratio. Large increases in carboxysome length:width ratio were observed in cysteine mutants, particularly those expressing C279S and CcmM-4S. The numbers of carboxysomes analysed in the different strains are indicated. The horizontal line represents the mean of each data set. Whiskers indicate upper and lower data points of the range. The number of measured values (*n*) is indicated above each data set. **d**, Carboxysomes per cell section for wild-type and *ccmM* mutant strains analysed by TEM. Cell sections (wild-type *n* = 25; Δ*ccmM* + *ccmM* *n* = 23; C279S *n* = 28; C395S *n* = 26; CcmM-4S *n* = 25) were analysed and the data expressed as a percentage of the number of carboxysomes in wild-type ± s.e.m. **P* ≤ 0.05; ****P* ≤ 0.001 (Tukey's multiple comparisons test). *P* values indicated for wild-type and Δ*ccmM*+*ccmM* cells versus cysteine mutants (wild-type versus CcmM-4S *P* = 0.0006, wild-type versus C395S *P* = 0.0008, Δ*ccmM*+*ccmM* versus CcmM-4S *P* = 0.0397, Δ*ccmM*+*ccmM* versus C395S *P* = 0.0464).



Extended Data Fig. 6 | Cryo-EM single-particle reconstruction.

a, Representative micrograph of Rubisco-M35 complexes used for single-particle reconstruction ($n = 4,723$ micrographs from one EM grid). **b**, 2D class averages of complexes in **a** showing extra density of SSUL bound to Rubisco in some classes. **c**, Meshwork representation of unsharpened D4 symmetry-averaged electron density map of the Rubisco-M35 complex (at a contour level of 2.8σ , appropriate for Rubisco), with the D4-averaged final model in backbone trace representation superposed (top). Below, zoomed-in region. Note that there is only ~50% residual

density for the two SSUL modules, except for the region in which the modules overlap in the centre (helices $\alpha 1$). **d**, The workflow for single-particle data processing. The left scheme of the flowchart identified only one SSUL domain bound between two RbcL subunits. The main scheme took the units consisting of two RbcL, two RbcS and two SSUL for focused classification and further refinement to improve the map quality. See Methods for details. **e**, Gold-standard FSC corrected curve (masked and B -factor sharpened) of the final 3D reconstruction. The resolution is ~2.77 Å at the FSC cutoff of 0.143.

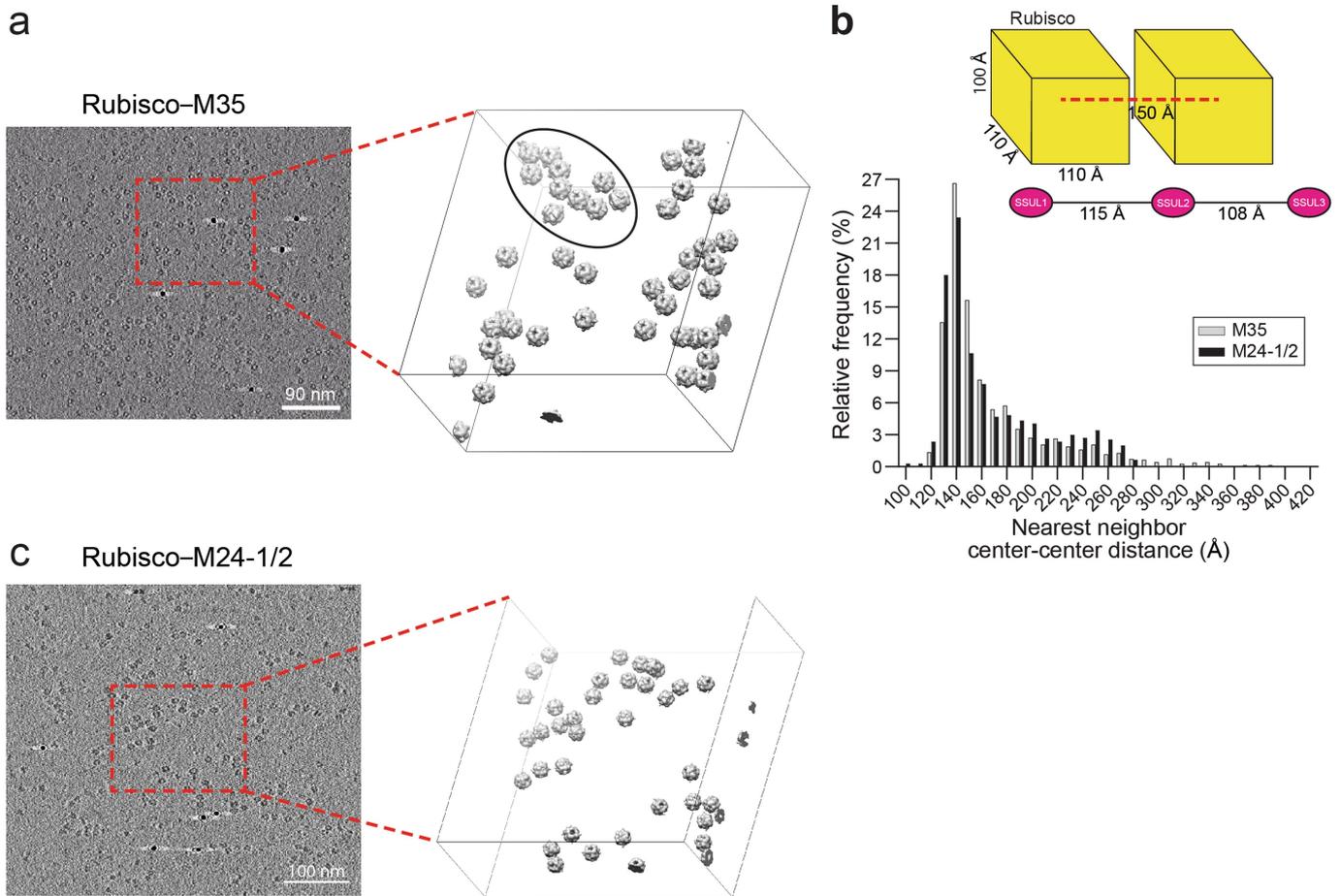


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Cryo-EM single-particle reconstruction.

a, Local resolution for the refined map of class v (2RbcL–2RbcS–SSUL) particles (see Extended Data Fig. 6d). The colour gradient from blue to red indicates local resolution from 2.0 to 4.0 Å. Right, zoomed-in views of the solvent-facing side and the buried interaction side of SSUL. As expected, the elements facing Rubisco have higher resolution. **b**, Quality of the cryo-EM density map of SSUL and the structural model of 2RbcL–2RbcS–SSUL in the vicinity of some aromatic side chains of SSUL1. The density is shown as a meshwork in cyan. The backbone of the structural model is in ribbon representation, and side chains are shown in stick representation. The cryo-EM map is shown at 1.2σ . **c**, **d**, Cryo-EM density map for SSUL at interface I (**c**) and interface II (**d**). **c**, Structural features of interface I. Critical interacting residues (see Fig. 4c): Phe253^{SSUL} forms van der Waals contacts to His353^{RbcL-A}/Glu355^{RbcL-A} and Arg254^{SSUL} forms interactions with Glu351^{RbcL-A} and a backbone–backbone H bond with His353^{RbcL-A}. The side-chain of Arg251^{SSUL} is sandwiched between RbcS residues Gln36/Gly37 and Asn94/Ile95, and forms a salt bridge with Asp93^{RbcS}. Arg252^{SSUL} forms a salt bridge with Asp76^{RbcL-B}. Thr255^{SSUL} and Ser257^{SSUL} form additional contacts with Arg79^{RbcL-B} and Asp76^{RbcL-B}. **d**, Structural features of interface II. Critical interacting residues (see Fig. 4d): Ile293^{SSUL}

contacts Thr30^{RbcL-B}. Arg298^{SSUL} makes van der Waals interactions with Tyr85^{RbcL-B} and His86^{RbcL-B}, and Arg300^{SSUL} with Tyr29/Thr30/Pro31/Lys32 in RbcL-B. **e**, Circular dichroism wavelength scans of wild-type M35 and the mutant M35(R251D/R252D) (3.5 μM each) in buffer (50 mM KH₂PO₄ pH 7.5) measured at 25 °C with 0.1-cm cuvettes using a Jasco 715 CD spectrometer. **f**, Dependence of Rubisco–M35 interaction on interface I contacts and RbcS. RbcL₈ (0.25 μM) and RbcS (2.5 μM) were mixed in buffer A and incubated for 10 min, followed by addition of M35 or M35(R251D/R252D) (1 μM each), and complex formation monitored by turbidity (black and green, respectively). The requirement for RbcS was analysed by addition of RbcS to premixed RbcL₈ and M35. Arrow indicates time of RbcS addition. A representative experiment is shown ($n = 3$ independent experiments). **g**, Inability of M35 mutant R251D/R252D to induce phase separation of Rubisco. Unlabelled M35(R251D/R252D) (1 μM) was mixed with Rubisco (0.25 μM, containing 10% A532–Rubisco) and imaged by bright field (left) and fluorescence microscopy (right) at 25 °C (top). As control, unlabelled M35 (1 μM) was mixed with Rubisco (0.25 μM, containing 10% A532–Rubisco) (bottom). Scale bars, 10 μm. A representative experiment is shown ($n = 3$ independent experiments).



Extended Data Fig. 8 | Cryo-ET of Rubisco-35 and Rubisco-M24-1/2 condensates. a, Analysis of the Rubisco-M35 condensate by cryo-ET. A slice through a tomographic volume of the matrix (left) and Rubisco-SSUL particles, from single-particle reconstruction with C1 symmetry and low-pass filtered to 40 Å, rendered into a 3D matrix (right). The circled cluster of complexes is shown again in Fig. 4e. A representative tomogram is shown ($n = 3$ independent tomograms of the same biological sample). **b,** Distribution of nearest neighbour centre-to-centre distances between Rubisco particles in tomograms of Rubisco-M35 (2,654 particles analysed)

and Rubisco-M24-1/2 (1,405 particles analysed). Data are from $n = 3$ tomograms for each sample. The schematic above the histogram shows the distance between two Rubisco particles relative to the maximum distance spanned by the linkers between SSUL modules in M35. **c,** Analysis of the Rubisco-M24-1/2 condensate by cryo-ET. A slice through a tomographic volume of the matrix (left) and Rubisco-SSUL particles, from single-particle reconstruction with C1 symmetry and low-pass filtered to 40 Å, rendered into a 3D matrix (right). A representative tomogram is shown ($n = 3$ independent tomograms of the same biological sample).

Extended Data Table 2 | Plasmids, proteins and strains used or generated for this study

| Plasmid/Strain | Protein/Strain | Source |
|--|-----------------------|------------|
| pET11a_Syn6301_rbcL | RbcL ₈ | Ref. 31 |
| pET11a_Syn6301_rbcS | RbcS | Ref. 31 |
| pHUE_Syn7942_ccmM_M13-1 | SSUL1 | This study |
| pHUE_Syn6301_ccmM_M35 | M35 | This study |
| pHUE_Syn6301_ccmM_M35_R251D/R252D | R251D/R252D mutant | This study |
| pHUE_Syn6301_ccmM_M35_C261S/C279S/C377S/C395S | M35-4S mutant | This study |
| pHUE_Syn6301_ccmM_M24-1/2 | M24-1/2 | This study |
| pHUE_Syn6301_ccmM_M24-2/3 | M24-2/3 | This study |
| pHUE_Syn6301_ccmM_M13-1 | M13-1 | This study |
| pHUE_Syn6301_ccmM_M13-2 | M13-2 | This study |
| pHUE_Syn6301_ccmM_M13-3 | M13-3 | This study |
| pHUE_Se7942_SSUL1 | SSUL1 | This study |
| pSE4_Se7942_ccmM | Δ ccmM strain | Ref. 38 |
| pSE4_Se7942_ccmM_C279S | C279S mutant strain | This study |
| pSE4_Se7942_ccmM_C395S | C395S mutant strain | This study |
| pSE4_Se7942_ccmM_C261S/C279S/C377S/C395S (-4S) | CcmM-4S mutant strain | This study |
| pHUE_A.thaliana_BSD2 | AtBSD2 | Ref. 22 |
| pHUE_Syn6301_RbcX | RbcX | Ref. 31 |
| pHUE_Se7942_Raf1 | Raf1 | Ref. 21 |

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

SerialEM version 3.5; Spectra Manager; Leica Application Suite X; O2View v 2.10 (Hansatech Chlorolab 2).

Data analysis

CCP4 version 7.0 (POINTLESS, AIMLESS, CTRUNCATE); MOLREP; Coot version 0.8.2; REFMAC5 version 5.8.0155; MolProbity plugin (Phenix version 1.8.2-1309); PyMol (Schrödinger LLC version 1.8); ESPript; RELION 2.1; MotionCor2 version 1.1.0; Gautomatch version 0.56; ResMap version 1.95; Chimera version 1.12; etomo (IMOD version 4.10.15); PyTom version 0.97; Matlab version 2016b; TOM Toolbox (Matlab); Fiji; BioRad Image Lab (v 5.2.1); GraphPad Prism (v 6.07 and 7); OriginPro 2018 SR1 b9.5.1.195; SigmaPlot 14.0; Symphotime (PicoQuant); Astra (Wyatt Technology).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The crystallographic models and structure factors for SSUL1ox and SSUL1red have been deposited to wwPDB under accession codes 6HBA (oxidized) and 6HBB (reduced), respectively. The electron density reconstructions and final 2Rbcl:2RbcS:SSUL model have been deposited in the Electron Microscopy Data Bank (EMDB) and wwPDB under accession codes EMD-0180 and 6HBC, respectively. Source data for graphs in main figures can be found in Graphical Source Data Fig. 1 to 3 and the source data for the gels shown in Extended Data Fig. 5b in Supplementary Fig. 1. Plasmids used and generated in this study are available from the corresponding author upon request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|-----------------|---|
| Sample size | No statistical methods were used to predetermine sample size. For carboxysome numbers in EM cross-sections and dimension analysis, all structures identified as carboxysomes were measured from all available EM images in order to capture all available data. For K0.5 Ci, and growth rate and sub-cellular Rubisco distribution analysis, three replicate analyses for each cell line are reported and represent typical results after repeated measurements. All biochemical experiments were replicated three times. The sample size chosen was generally sufficient to produce positive, statistically significant results. |
| Data exclusions | No data were excluded. |
| Replication | All attempts at replication were successful. For K0.5 Ci, and growth rate and sub-cellular Rubisco distribution analysis, repeated analysis was carried out two to three times to confirm result consistency. |
| Randomization | n.a. |
| Blinding | No blinding during data acquisition and analysis. Blinding was not relevant in this study because of the objective nature of data acquisition and analysis in this study. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| n/a | Involved in the study |
|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |

Methods

| n/a | Involved in the study |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

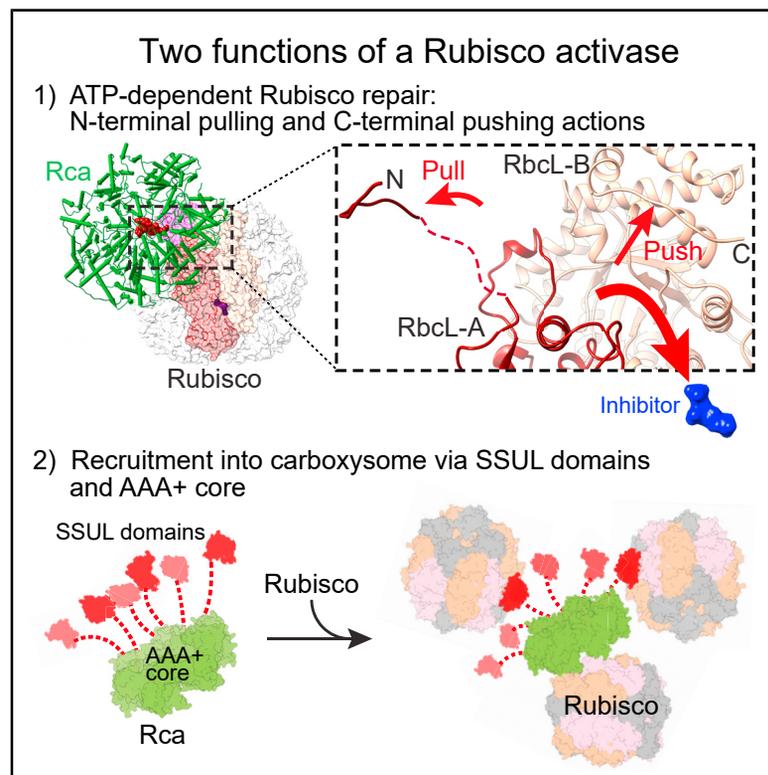
Antibodies

| | |
|-----------------|---|
| Antibodies used | Anti-RbcL and anti-M35 (Long B.M. et al., J. Biol. Chem., 2007), Alkaline phosphatase-conjugated anti-rabbit secondary antibody (A3687, Sigma). All antibodies used at 1:5000 dilution. |
| Validation | Specificity was demonstrated using deletion strains of <i>Synechococcus elongatus</i> PCC7942 (see Extended Data Fig. 5b). |

4.2 Dual functions of a Rubisco activase in metabolic repair and recruitment to carboxysomes

Dual Functions of a Rubisco Activase in Metabolic Repair and Recruitment to Carboxysomes

Graphical Abstract



Authors

Mirkko Flecken, Huping Wang,
Leonhard Popilka, F. Ulrich Hartl,
Andreas Bracher, Manajit Hayer-Hartl

Correspondence

mhartl@biochem.mpg.de

In Brief

The cryo-EM structure of cyanobacterial Rca:Rubisco complex reveals the molecular basis for NosRca-mediated Rubisco repair and carboxysome organization, highlighting distinct modes of action compared to proteobacterial Rca.

Highlights

- Cyanobacterial NosRca, a AAA+ chaperone, functions as a Rubisco activase
- NosRca pulls the N terminus of the Rubisco RbcL subunit into its central pore
- Concomitant pushing on the C terminus of the adjacent RbcL releases inhibitory sugar
- The interaction with Rubisco also mediates recruitment of NosRca to carboxysomes

Article

Dual Functions of a Rubisco Activase in Metabolic Repair and Recruitment to Carboxysomes

Mirkko Flecken,^{1,3} Huping Wang,^{1,3} Leonhard Popilka,^{1,4} F. Ulrich Hartl,^{1,2} Andreas Bracher,^{1,2} and Manajit Hayer-Hartl^{1,5,*}

¹Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany

²Senior author

³These authors contributed equally

⁴Present address: Capgemini Deutschland GmbH, Olof-Palme-Str. 14, 81829 Munich, Germany

⁵Lead Contact

*Correspondence: mhartl@biochem.mpg.de

<https://doi.org/10.1016/j.cell.2020.09.010>

SUMMARY

Rubisco, the key enzyme of CO₂ fixation in photosynthesis, is prone to inactivation by inhibitory sugar phosphates. Inhibited Rubisco undergoes conformational repair by the hexameric AAA+ chaperone Rubisco activase (Rca) in a process that is not well understood. Here, we performed a structural and mechanistic analysis of cyanobacterial Rca, a close homolog of plant Rca. In the Rca:Rubisco complex, Rca is positioned over the Rubisco catalytic site under repair and pulls the N-terminal tail of the large Rubisco subunit (RbcL) into the hexamer pore. Simultaneous displacement of the C terminus of the adjacent RbcL opens the catalytic site for inhibitor release. An alternative interaction of Rca with Rubisco is mediated by C-terminal domains that resemble the small Rubisco subunit. These domains, together with the N-terminal AAA+ hexamer, ensure that Rca is packaged with Rubisco into carboxysomes. The cyanobacterial Rca is a dual-purpose protein with functions in Rubisco repair and carboxysome organization.

INTRODUCTION

The AAA+ (ATPases associated with diverse cellular activities) chaperone Rubisco activase (Rca) serves as a paradigm of conformational enzyme repair (Bhat et al., 2017b; Mueller-Cajar, 2017). The exclusive client of Rca is the photosynthetic key enzyme Rubisco (ribulose-1,5-bisphosphate carboxylase/oxygenase), which is directly or indirectly responsible for all biomass production (Bar-On and Milo, 2019). Since Rca is required for optimal Rubisco function, understanding its mechanism is important in efforts to enhance photosynthesis with the goal of increasing crop yields (Andralojc et al., 2018; Bailey-Serres et al., 2019; Éva et al., 2019; Sharwood, 2017; Singer et al., 2019; Slattery and Ort, 2019).

Rubisco catalyzes the carboxylation of the 5-carbon sugar ribulose-1,5-bisphosphate (RuBP) in the Calvin-Benson-Bassham cycle of photosynthesis. This multistep catalytic reaction produces two molecules of 3-phosphoglycerate as fuel for the synthesis of carbohydrates, fatty acids, and amino acids. The most prevalent form of Rubisco (form I) in plants, algae, and cyanobacteria is a hexadecamer of eight large (RbcL; ~50–55 kDa) and eight small (RbcS, ~12–18 kDa) subunits. The RbcL subunits form a cube-shaped core, consisting of four antiparallel dimers, capped by four RbcS subunits at the top and bottom. Two catalytic sites per RbcL dimer unit are located at the interface between the N-terminal domain of one RbcL and the C-terminal

domain of the other subunit. In order to become functionally competent, the RuBP binding site must be activated by carboxylation of the active-site lysine, which results in formation of a carbamate, and is followed by the binding of a Mg²⁺ ion. This process is commonly referred to as carbamylation (Hartman et al., 1987; Miziorko and Lorimer, 1983). Upon binding of RuBP, the active site is sealed by the so-called “60s loop” from the N-terminal domain of one RbcL subunit and “loop 6” of the C-terminal domain of the adjacent subunit. The N-terminal sequence pins down the 60s loop and the C terminus stretches over loop 6. This arrangement generates the physical environment required for electrophilic attack of RuBP by CO₂.

The requirement of Rca arises from the fact that the complex catalytic reaction of Rubisco is error prone, leading to the generation of tight-binding sugar phosphates that inactivate the enzyme. These “misfire” products include xylulose-1,5-bisphosphate (XuBP) and 2,3-pentodiulose-1,5-bisphosphate (PDBP) (Parry et al., 2008). Moreover, some plants under low light produce the inhibitory 2-carboxy-D-arabinitol-1-phosphate (CA1P) (referred to as “night-time” inhibitor). Binding of RuBP to non-carbamylated Rubisco can also lead to inactive enzyme (Bracher et al., 2017; Parry et al., 2008; Portis et al., 2008). In all these conditions, the release of the inhibitory sugar is mediated by Rca, which couples ATP hydrolysis to structural remodeling of the Rubisco catalytic site. Rca proteins have evolved independently in photosynthetic organisms and are associated

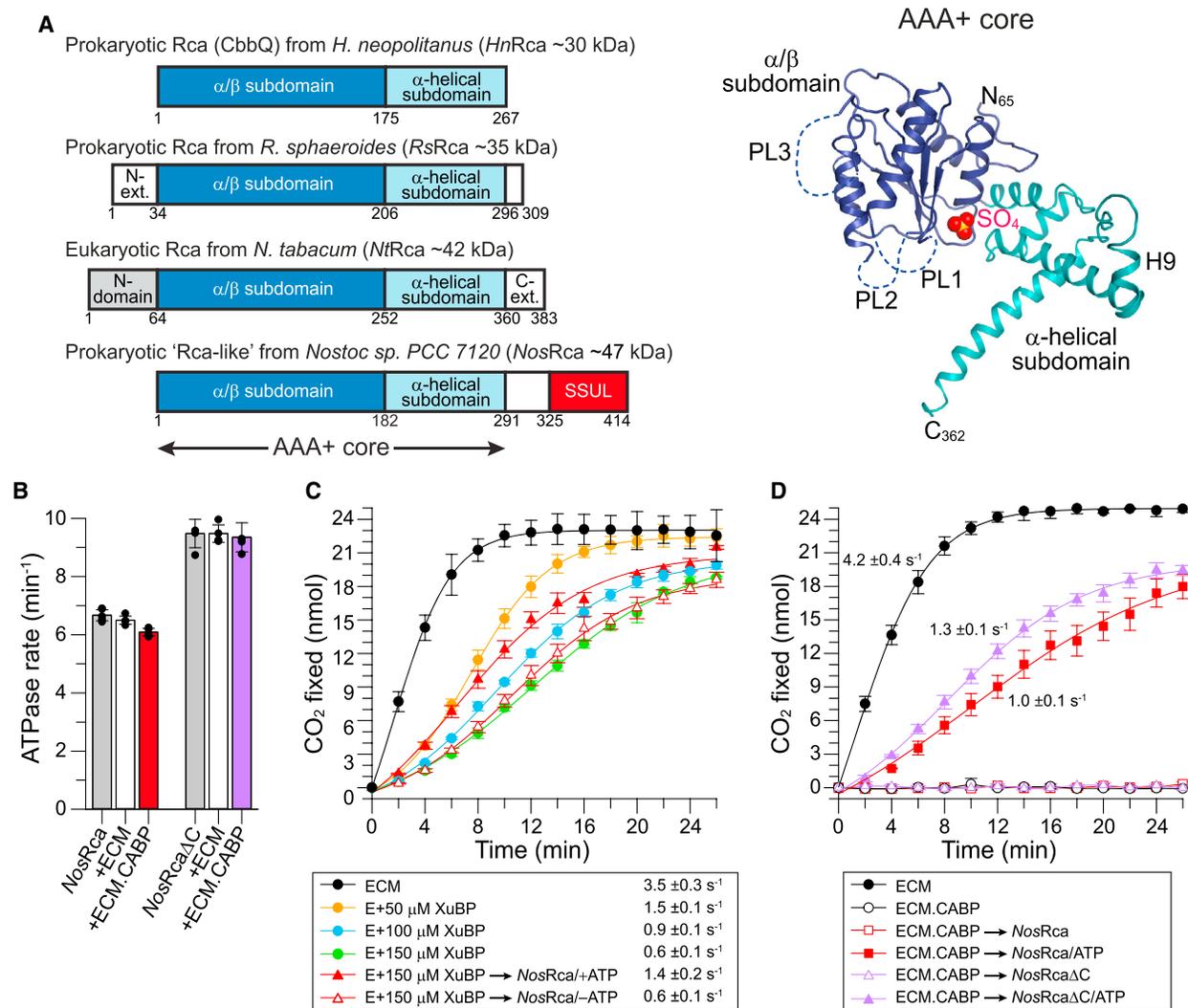


Figure 1. Rubisco Activase Function of *NosRca*

(A) Left: Domain structures of Rca proteins across phylogenetic groups. CbbQ from the chemoautotroph *Halothiobacillus neopolitanus* (*HnRca*), CbbX from *R. sphaeroides* (*RsRca*; prokaryotic, red-type), Rca from *N. tabacum* (*NtRca*; eukaryotic, green-type), and cyanobacterial "Rca-like" (*NosRca*; prokaryotic, green-type) from *Nostoc sp. PCC 7120*. Right: Ribbon representation of Rca from *A. thaliana* (*AtRca*; PDB: 4W5W) (Hasse et al., 2015). α/β - and α -helical subdomains, the pore-loops PL1, PL2, and PL3 and helix H9 are indicated. A bound sulfate ion in the nucleotide binding pocket is represented as space-filling model.

(B) ATPase rates of *NosRca* and *NosRca*ΔC in presence or absence of active (ECM) or inhibited (ECM.CABP) *NosRubisco*. ATPase rates were measured at 0.5 μM Rca (hexamer) and 0.25 μM *NosRubisco* (hexadecamer). Bars represent averages with SD from at least three independent replicates. See STAR Methods for details.

(C) *NosRca* functions as a Rubisco activase. CO₂ fixation assays were performed with Rubisco enzyme (E, 0.25 μM) in presence of increasing concentrations of XuBP and in presence or absence of 0.5 μM Rca (hexamer), 3 mM ATP, and 0.4 mM RuBP. The active *NosRubisco* (ECM) was used as control. Approximate rates of CO₂ fixation were determined from the linear parts of the curves. Data represent averages with SD from at least three independent replicates.

(D) Reactivation of CABP inhibited *NosRubisco* (ECM.CABP) by *NosRca* or *NosRca*ΔC. CO₂ fixation assays were performed as in (C) with ECM.CABP in the presence or absence of ATP. Bars represent averages with SD of at least three independent replicates.

See also Figure S1 and Table S1.

with the red and green lineages of Rubisco (Bhat et al., 2017b; Mueller-Cajar, 2017; Tabita et al., 2008) (Figure 1A). While divergent in sequence, they all share the canonical AAA+ core consisting of an N-terminal α/β -nucleotide binding subdomain and a C-terminal α -helical subdomain (Figure 1A) and function as six-membered ring complexes. The AAA+ chaperones are typically substrate promiscuous and act by threading flexible termi-

nal sequences or loop segments of their target proteins into the hexamer pore (Olivares et al., 2016; Puchades et al., 2020). Threading is mediated by pore-loops (Figure 1A) that face the central pore and engage the substrate peptide, exerting a pulling force (Avellaneda et al., 2020; de la Peña et al., 2018; Dong et al., 2019; Fei et al., 2020; Ripstein et al., 2020; Rizo et al., 2019; Twomey et al., 2019).

A low-resolution cryo-electron microscopy (cryo-EM) structure of the Rca from the proteobacterium *Rhodobacter sphaeroides* (Rs) in complex with red-type Rubisco (form 1C) showed that the RsRca hexamer docks onto Rubisco over one catalytic site at a time (Bhat et al., 2017a). RsRca is positioned so as to allow threading of the extended C-terminal sequence of RsRbcL into the central pore, facilitating opening of the catalytic site and release of the inhibitory sugar (Bhat et al., 2017a; Mueller-Cajar et al., 2011). In contrast, the mechanism of Rca in plants and green algae containing form IB Rubisco (Tabita et al., 2008) remains enigmatic, because form IB RbcL lacks the extended C-terminal sequence of red-type RbcL. Moreover, Rca of plants and green algae have a domain N-terminal to the AAA+ core that is required for Rubisco recognition (Figure 1A), and the so-called specificity helix H9 in the α -helical subdomain confers specificity for Rubisco proteins from solanaceous (nightshade) and non-solanaceous species (Li et al., 2005; Portis et al., 2008; Wachter et al., 2013). The mechanistic analysis of plant Rca has proven difficult, since the N-domain appears to be flexible and the protein, though active as a hexamer, tends to populate a range of dynamic oligomeric states *in vitro* (Blayney et al., 2011; Keown and Pearce, 2014; Stotz et al., 2011).

The closest sequence homolog to plant Rca is the “Rca-like” protein of several cyanobacteria with form IB Rubisco. Rca-like lacks the flexible N-terminal domain, but its AAA+ core is highly homologous to eukaryotic Rca, with ~38% sequence identity and ~70% similarity (Data S1). Consistent with Rca function, the Rca-like protein of *Anabaena* sp. strain CA was shown to be required for elevated Rubisco activity and growth under high light conditions (Li et al., 1999). However, a recent study could not detect Rca activity for the Rca-like protein from *Fremyella diplosiphon* (Lechno-Yossef et al., 2020). A distinguishing feature of Rca-like is the presence of a Rubisco small subunit-like (SSUL) domain, connected to the C terminus of the AAA+ core via a flexible linker (Figure 1A). The homologous SSUL modules of the scaffolding protein CcmM have been shown to bind to Rubisco, inducing condensate formation for packaging into carboxysomes (Ryan et al., 2019; Wang et al., 2019), protein-bounded compartments in which high concentrations of CO₂ are generated for carbon fixation (Badger and Price, 2003; Price et al., 1998; Turmo et al., 2017). Similarly, the Rca-like protein of *F. diplosiphon* interacts with Rubisco via its SSUL domains, and its deletion resulted in impaired regulation of carboxysome biogenesis (Lechno-Yossef et al., 2020).

To obtain insight into the basic mechanism of plant Rca, here we analyzed the structure and function of the Rca-like protein of the cyanobacterium *Nostoc* sp. PCC 7120 (NosRca). We find that NosRca functions as a bona fide Rubisco activase, independent of the SSUL domain (NosRca Δ C). We solved the crystal structure of the NosRca Δ C hexamer and obtained its high-resolution cryo-EM structure in complex with inhibited Rubisco. A key finding is the well-resolved N-terminal amino acid sequence of RbcL inserted into the central pore of the Rca hexamer. Mutational analysis confirmed that plant Rca also engages the N-terminal RbcL sequence of its cognate Rubisco, establishing binding of the N terminus rather than the C terminus as the basic mechanism for activation of form IB Rubisco. We further show that the SSUL domains cooperate with the AAA+ core to ensure

that Rca is packaged with Rubisco into carboxysomes. The cyanobacterial Rca thus combines functions in Rubisco remodeling and carboxysome organization.

RESULTS

NosRca Protein Is a Rubisco Activase

The purified recombinant NosRca protein, with and without nucleotide, behaved as a ~260 kDa complex by size-exclusion chromatography coupled to multiangle light scattering (SEC-MALS), consistent with a hexamer (Figures S1A and S1B; Table S1). NosRca had a constitutive ATPase activity of ~6.5 min⁻¹ per protomer (Figure 1B), which is lower than the ATPase rates reported for plant Rca (~17–45 min⁻¹) (Hazra et al., 2015). We next tested whether NosRca functions as a Rubisco activase, considering that the AAA+ modules of NosRca and the Rca proteins of *Nicotiana tabacum* (NtRca) and *Arabidopsis thaliana* (AtRca) share ~59% sequence identity (Data S2). Because activases are adapted to their cognate Rubisco (Bhat et al., 2017b; Mueller-Cajar, 2017), we recombinantly expressed and purified the Rubisco from *Nostoc* sp. PCC 7120 (NosRubisco) (Figure S1A) using coexpression of *Nostoc* chaperonin (NosGroES/EL; see STAR Methods). The purified NosRubisco, activated with CO₂ and Mg²⁺ (ECM), catalyzed the carboxylation of RuBP at a rate of ~4 molecules of CO₂ per active site s⁻¹ (Figures 1C and 1D). However, as shown previously (Li et al., 1999), binding of RuBP to the non-activated enzyme did not result in inhibition (Figure S1C), in contrast to plant Rubisco (Parry et al., 2008; Stotz et al., 2011; Wang and Portis, 1992). We therefore tested the effect of the misfire sugar phosphate XuBP as a competitive inhibitor (Bracher et al., 2015). A ~60%–80% inhibition of the carboxylation rate was observed at 50 to 150 μ M XuBP in the presence of low RuBP (400 μ M) (Figure 1C). Note that NosRubisco has a lower affinity for XuBP than plant Rubisco (Bracher et al., 2015; Pearce, 2006). Addition of NosRca partially relieved the inhibition in an ATP-dependent manner (Figure 1C), suggesting that NosRca facilitated the release of XuBP from the Rubisco catalytic sites. The effect was only partial because the XuBP remained in the reaction as a competitive inhibitor. However, XuBP is unlikely to be a physiological inhibitor of NosRubisco, consistent with the absence in *Nostoc* sp. PCC 7120 of XuBP phosphatase, which serves to hydrolyze the XuBP upon release (Bracher et al., 2015). Complete and highly efficient inhibition of Rubisco was obtained with 2-carboxyarabinitol-1,5-diphosphate (CABP), a mimic of the carboxyketone intermediate of the carboxylation reaction, at a concentration equivalent to Rubisco active sites (0.25 μ M) (Figure 1D). Strikingly, NosRca efficiently reactivated the inhibited Rubisco in the presence of ATP (Figures 1D). Thus, NosRca is a bona fide Rubisco activase, although the endogenous inhibitory sugar phosphate(s) of NosRubisco remains to be identified. Deletion of the C-terminal 123 residues (linker and the SSUL domain) in NosRca (NosRca Δ C), resulted in a faster ATPase rate (~9.5 min⁻¹) and slightly more efficient Rubisco reactivation (Figures 1B and 1D), indicating that the SSUL domain is not required for activase function. The rate of reactivation was dependent on the concentration of NosRca (Figure S1D). Similar ATPase rates were measured in the presence of activated or inhibited Rubisco (Figure 1B), a property similar to plant Rca (Robinson and Portis, 1989).

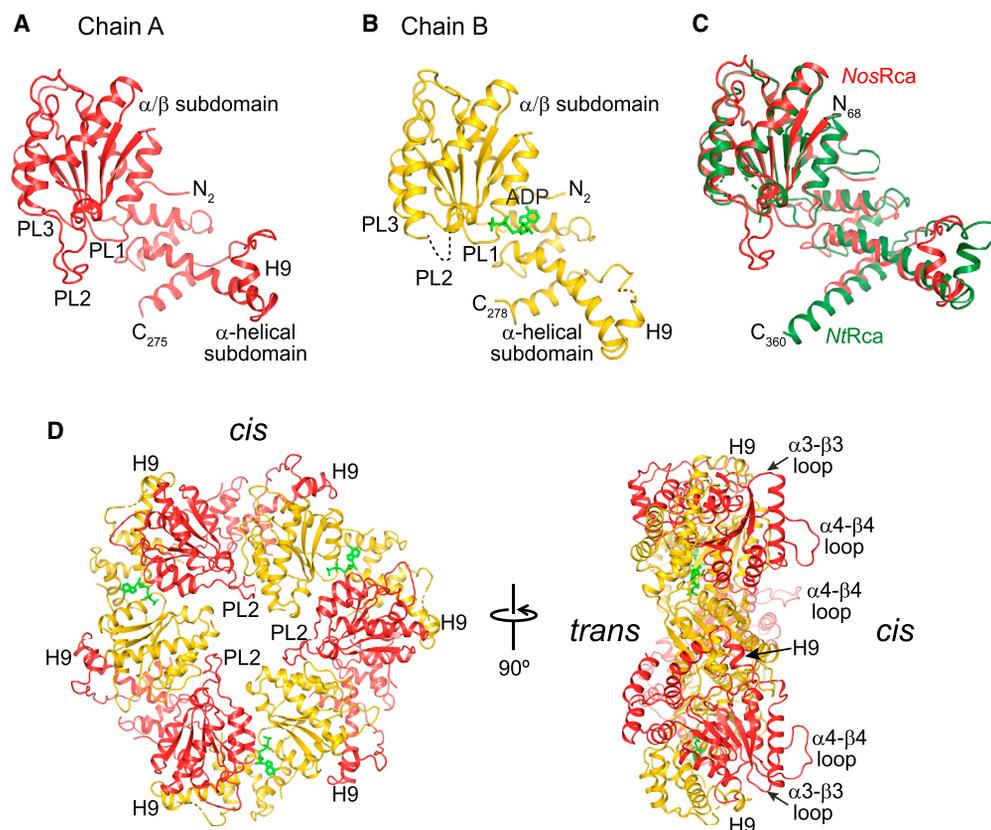


Figure 2. Crystal Structure of NosRca Δ C

(A and B) Ribbon representations of the two crystallographically independent chains of NosRca Δ C in red (A) and yellow (B), respectively. The subdomain structure as well as locations of helix H9 and pore-loops PL1, PL2, and PL3 are indicated. ADP is represented as a wire-frame model in green. Chain termini are indicated. (C) Superposition of chain A (no nucleotide bound) of the asymmetric unit of NosRca Δ C with NtRca (PDB: 3T15) (Stotz et al., 2011). Chain A of NosRca Δ C and NtRca are shown as ribbons in red and green, respectively.

(D) NosRca Δ C hexamer. The hexamer was generated by applying the crystallographic 3-fold rotational symmetry. Helix H9, pore-loop PL2, loop α 3- β 3, and loop α 4- β 4 are indicated. Chain A, red; chain B, yellow. *cis*, surface at which helix H9 is located; *trans*, the opposite surface from which the linkers to the SSUL domains emanate.

See also Figure S2, Data S1 and S2, and Tables S1 and S2.

Crystal Structure of NosRca Δ C

To allow a structural comparison of NosRca with plant Rca, we solved the crystal structure of NosRca Δ C (residues 2–291) at ~ 2.7 Å by Gd-multi-wavelength anomalous diffraction (Gd-MAD) (Figure S1A; Table S2). The resolution of the isomorphous native structure was 2.45 Å (Table S2). The asymmetric unit contained two subunits of NosRca Δ C, with chain A (residues 2–275) having no nucleotide bound and chain B (residues 2–278) having ADP bound (Figures 2A, 2B, and S2A). Note that the nucleotide bound during recombinant expression.

NosRca Δ C shows the typical AAA+ fold topology, consisting of the N-terminal Rossmann-fold α/β subdomain and a smaller C-terminal α -helical subdomain (Figures 2A and 2B). The two NosRca Δ C subunits had nearly identical subdomain conformations but differed in the inter-domain angle between the α/β and the α -helical subdomains by $\sim 11^\circ$ (Figure S2B). Chain A superimposes well with the subunit structures of NtRca (rmsd 1.46 Å for 206 matching C α positions) and AtRca (rmsd 1.96 Å for 219 matching C α positions) (Figures 2C and S2C). The α/β

subdomains separately match almost perfectly (rmsd 0.53 Å and 0.52 Å, respectively). The structural conservation between NosRca and plant Rca includes the so-called specificity helix H9 (α 9) in the α -helical subdomain (Figures 2A and 2B; Data S1 and S2). As in the NtRca and AtRca crystal lattices (Hasse et al., 2015; Stotz et al., 2011), the α/β subdomain and helix H9 of the adjacent subunit form a rigid module (Figure S2D). The orientation of helix H9 to the four-helix bundle of the α -helical subdomain is variable (Figure S2E). In the rhombohedral crystal lattice the NosRca Δ C subunits form a hexamer with alternating ADP-bound and nucleotide-free subunits (Figure 2D). Note that NosRca Δ C is a hexamer in solution (Table S1). We define the surface at which helix H9 is located as *cis* and the opposite surface, from which the linkers to the SSUL domains would emanate, as *trans* (Figure 2D).

The interface between the α/β subdomain and α -helical subdomain of adjacent subunits contains the nucleotide binding site (Figures 2D and S2A). In contrast to the pore-loops in the helical crystal structures of NtRca and AtRca (Hasse et al., 2015; Stotz

et al., 2011), the pore-loops of *NosRca*ΔC are ordered (PL1, residues 67–72; PL2, residues 105–119; PL3, residues 160–165), with the exception of the B chains where PL2 is disordered (Figures 2A, 2B, and 2D). The hexamer pore would be wide enough for a sphere of ~12 Å diameter to pass. In summary, the high degree of structural similarity of *NosRca* to plant Rca supports the identification of *NosRca* as a Rubisco activase. Thus, *NosRca* serves as a model to understand the mechanism of plant Rca.

***NosRca* Binds the N Terminus of RbcL in the Central Pore**

To understand the mechanism of *NosRca*, we sought to capture *NosRca* in the process of remodeling the inhibited Rubisco by cryo-EM. We incubated CABP-inhibited *NosRubisco* (ECM-CABP) with excess *NosRca*ΔC in the presence of ATP, followed after 10 s by addition of the slowly hydrolyzing ATP analog, ATP γ S, a nucleotide replacement strategy (Dong et al., 2019) (Figure 3A; STAR Methods). Cryo-EM analysis and single-particle reconstruction yielded one predominant conformation of the *NosRca*ΔC:Rubisco complex (Figures S3A–S3C). We obtained an EM density map with an overall resolution of ~2.86 Å (Figures S3C and S3D; Table S3), and an overall resolution of the Rubisco-subtracted *NosRca*ΔC of ~3.29 Å (Figures S3C and S3E). *NosRca*ΔC docks onto one corner of the cube-shaped Rubisco (Figure 3B; Video S1). The *cis* surface of the hexamer (Figure 2D) faces Rubisco, with the N-terminal α/β subdomains and the central pore positioned approximately over one substrate binding pocket of an antiparallel RbcL dimer (RbcL-A and RbcL-B; Figure 3B). The engaged catalytic site is in an open conformation with no discernable density for the inhibitor CABP, while the bound CABP was well resolved in all other binding pockets (Figure S4A). Thus, the interaction captured in the cryo-EM structure represents the end point of *NosRca* remodeling before dissociation from Rubisco.

A key feature of the complex was the presence of well-defined density in the *NosRca*ΔC hexamer pore, fitting to a 13-residue peptide in an extended conformation (Figure 3C; Video S1). This finding suggested a mechanism similar to that of red-type proteobacterial Rca, which engages the flexible C terminus of RbcL (Bhat et al., 2017a; Mueller-Cajar et al., 2011). However, the C-terminal sequence of RbcL-B was resolved up to residue Glu471 in the Rubisco structure, with only six residues (residues 472–477) missing in the EM density (Figures S4B and S4C), thus excluding engagement of the C terminus by the *NosRca* central pore. Strikingly, we found that the N-terminal sequence of subunit RbcL-A residues Ser2 to Tyr14 fit into the peptide density spanning the pore (Figure 3C). The next resolved N-terminal residue on RbcL-A was Thr24, with residues 15–23 connecting to the peptide in the pore being disordered (Figures S4B and S4D). In all other RbcL subunits, the first ordered N-terminal residue is Gly13. We noted that the N-terminal sequence of *NosRbcL* consists of alternating small and bulky side chains (Figure 3C). Interestingly, this pattern is conserved at the N terminus of green-type form IB RbcL (Tabita et al., 2008) of plants, green algae, and other cyanobacteria (Figure S4E).

To test the functional importance of the RbcL N terminus for reactivation, we generated a truncated *NosRubisco* mutant

lacking the first 12 residues (*NosL₈S₈*ΔN) (Figure S1A). *NosL₈S₈*ΔN exhibited wild-type carboxylation activity but could no longer be reactivated by *NosRca* upon inhibition with CABP (Figure 3D). To determine whether this mechanism is conserved in plants, we produced the analogous N-terminal RbcL truncation in Rubisco of *N. tabacum* (*NtL₈S₈*ΔN) by recombinant expression in the presence of chloroplast chaperonins and assembly factors (Aigner et al., 2017; Lin et al., 2019; Wilson et al., 2019) (Figure S1A; STAR Methods). *NtL₈S₈*ΔN was carboxylation active but the RuBP-inhibited enzyme (E.RuBP) could not be reactivated by *NtRca* in contrast to wild-type Rubisco (Figures 3E and S1A; STAR Methods). Thus, consistent with their high structural homology, engagement of the N-terminal RbcL sequence is important for remodeling by both *NosRca* and plant Rca.

Pore-Loop Interactions with the Rubisco N-Terminal Peptide

The Rubisco-bound *NosRca*ΔC hexamer adopted a right-handed helical “split-washer” conformation, with subunit 1 (Rca1; yellow) located at the bottom and Rca6 (blue) at the top of the spiral (Figure 4A). Nucleotide density was present in all six nucleotide binding pockets of *NosRca*, ADP at the Rca1-Rca2 interface, and ATP (or ATP γ S) in the other sites (Figure S5A). The ATP/ATP γ S was positioned close to the well-resolved side chains of the arginine residues Arg169 and/or Arg172 from the subsequent subunit (Figure S5A). These arginines have been implicated in ATP-hydrolysis in other AAA+ proteins, functioning as the so-called “arginine fingers” (Puchades et al., 2020). The EM density of both arginines is less well defined at the Rca6-Rca1 interface, the split ends of the *NosRca* spiral (Figure S5A), perhaps reflecting mixed conformations due to ongoing ATP hydrolysis.

The three pore-loops, PL1 (residues 67–72), PL2 (residues 105–119), and PL3 (residues 160–165) (Data S2) were resolved in all *NosRca* subunits (Figures S5B–S5D), adopting a staircase arrangement. Close examination of the bound peptide revealed that the small side chains of RbcL-A residues Ala4, Thr6, Thr8, and Thr10 point into successive binding pockets in the central pore formed between PL1 and PL2 of adjacent subunits Rca2 to Rca6 (Figures 4B, S5B, and S5C). The ADP-bound Rca1 is the only subunit not in contact with the peptide. The bulkier side chains of Gln5, Lys7, Gln9, and Lys11 point into the pore solvent channel (Figure 4B). Note that Ser2 and Tyr3, protruding from the *trans* side of the pore, also conform to this pattern of alternating small and bulky side chains, while the pattern does not continue beyond residue 12 (Figure S4E).

The bound peptide buries a substantial surface area of 1,018 Å² on *NosRca*. The contacts are mediated by van der Waals interactions from both PL1 and PL2 (Figure 4C). In addition, the peptide backbone is in hydrogen bond distance to the carbonyl group of Pro69 of PL1 in Rca3 to Rca6 (Figures 4C and S5E). Interestingly, the highly conserved aromatic residue Tyr116 in PL2, which is essential in plant Rca (Tyr188 in *NtRca*) (Stotz et al., 2011) (Data S2), does not contact the peptide (Figure S5F). In the ATP-bound subunits Rca2 to Rca6, the hydroxyl group of Tyr116 is in hydrogen bond distance to the amide of Phe110 in the same subunit, and to Gln121 in the subsequent

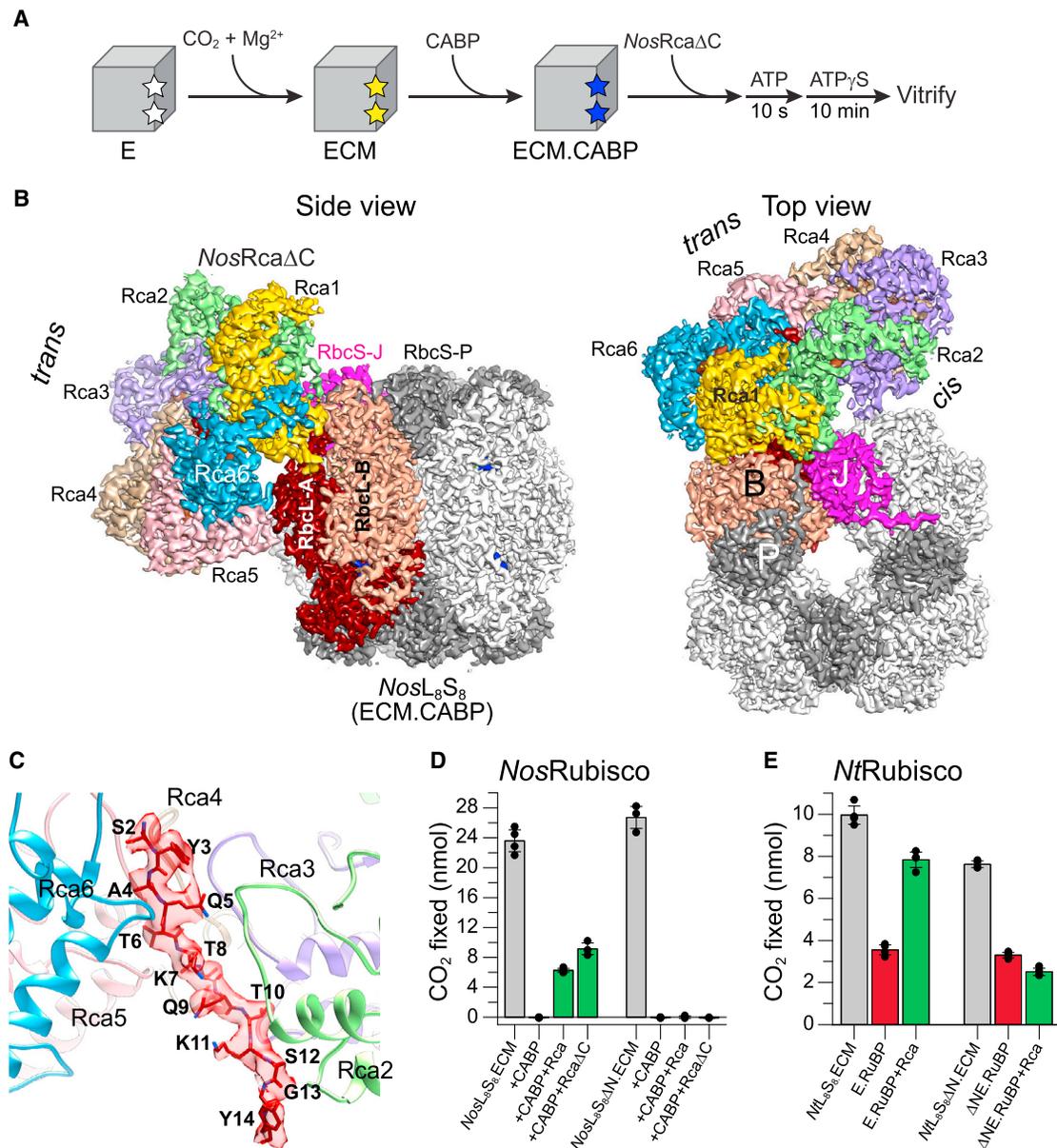


Figure 3. Cryo-EM Structure of the *NosRcaΔC*:Rubisco Complex

(A) Procedure for the preparation of the *NosRcaΔC*:Rubisco complex analyzed by cryo-EM.

(B) Cryo-EM reconstruction of the *NosRcaΔC*:Rubisco complex in two perpendicular views, with the side view aligned to the 2-fold axis of Rubisco and the top view to the 4-fold axis of Rubisco. *NosRca* subunits are labeled Rca1 to Rca6. The antiparallel RbcL dimer subunits RbcL-A and RbcL-B as well as the RbcS subunits RbcS-J and RbcS-P in contact with *NosRca* are indicated.

(C) Cryo-EM density of the N-terminal RbcL-A peptide bound inside the *NosRcaΔC* hexamer pore, interacting with Rca subunits (ribbon representation) Rca2 to Rca6. Amino acids in single letter code.

(D) Reactivation of *NosRubisco* by *NosRca* depends on the N terminus of RbcL. CO₂ fixation assays were performed as in Figure 1D with CABP-inhibited *NosRubisco* (*NosL₈S₈*) and *NosL₈S₈ΔN* in the presence or absence of *NosRca* or *NosRcaΔC* and ATP/1 mM RuBP for 8 min. Bars represent averages with SD from at least three independent experiments.

(E) Reactivation of *NtRubisco* by *NtRca* depends on the N terminus of RbcL. CO₂ fixation assays were performed with RuBP-inhibited *NtL₈S₈* (E.RuBP) and *NtL₈S₈ΔN* (Δ NE.RuBP) (0.25 μ M active sites) in the presence or absence of *NtRca* (0.5 μ M hexamer) and ATP/1 mM RuBP for 8 min. Bars represent averages with SD from at least three independent experiments. See STAR Methods for details.

See also Figures S3 and S4 and Table S3.

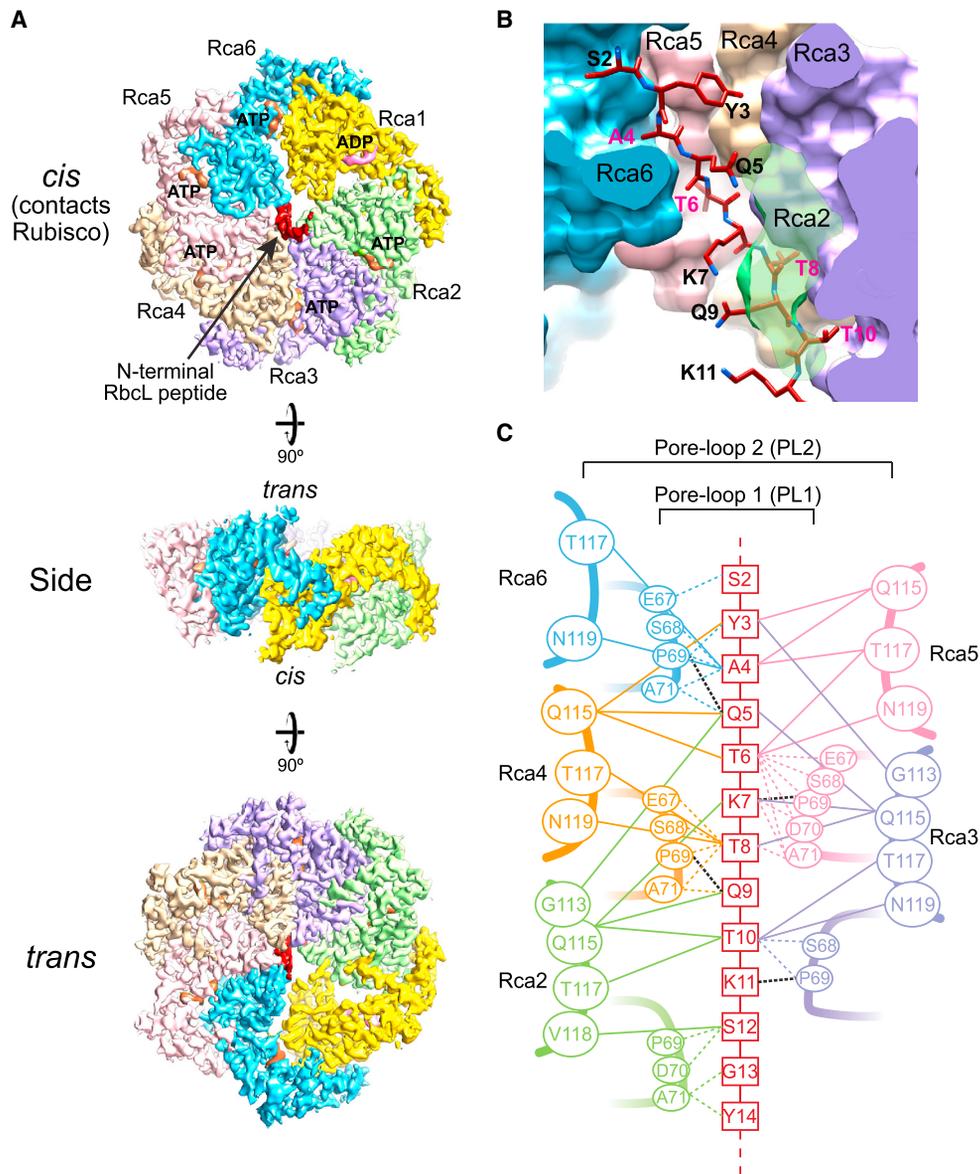


Figure 4. NosRca Interacts via Its Central Pore with the N-Terminal RbcL Peptide

(A) Cryo-EM density map of NosRca Δ C in the NosRca Δ C:Rubisco complex in *cis* (surface in contact with Rubisco) and *trans* view. ADP in pink, ATP/ATP γ S in sienna, and the bound N-terminal peptide from RbcL-A in red. NosRca subunits are color coded as in Figure 3B.

(B) Cavities in the NosRca central pore. Small cavities formed between pore-loops PL1 and PL2 from adjacent Rca subunits (Rca2 to Rca6) accommodate the side-chains of the alternating small amino acids of the RbcL N-terminal peptide. The NosRca central pore is in surface representation and the RbcL N-terminal peptide in stick representation.

(C) Interactions (van der Waals and putative hydrogen bonds) between residues of the RbcL N-terminal peptide and the staggered pore-loops PL1 (dashed colored lines) and PL2 (solid colored lines) of the Rca subunits. The distance cutoff was 4 Å. Putative hydrogen bonds with the peptide backbone are indicated by black dashed lines.

See also Figure S5 and Data S2.

subunit, resulting in a network of interactions that rigidify the central pore (Figure S5G). In contrast, the position of Tyr116 is ill-defined in the ADP-bound Rca1 (Figure S5F). PL3, which does not contact the bound peptide (Figure S5D), is positioned between the nucleotide binding site and PL2, suggesting that it may be involved in allosteric regulation via the spatially adjacent Arg169 and Arg172.

Details of the NosRca Δ C:Rubisco Interface and Remodeling Mechanism

Beyond engaging the RbcL N-terminal sequence, NosRca contacts Rubisco at three regions covering a total of 1,445 Å² (Figures 5A–5E; Video S1). Interface I (419 Å²) is formed between Rca5/Rca6 and RbcL-A and has high surface shape complementarity (Figures 5A and 5B). Rca5 residues Leu244 and

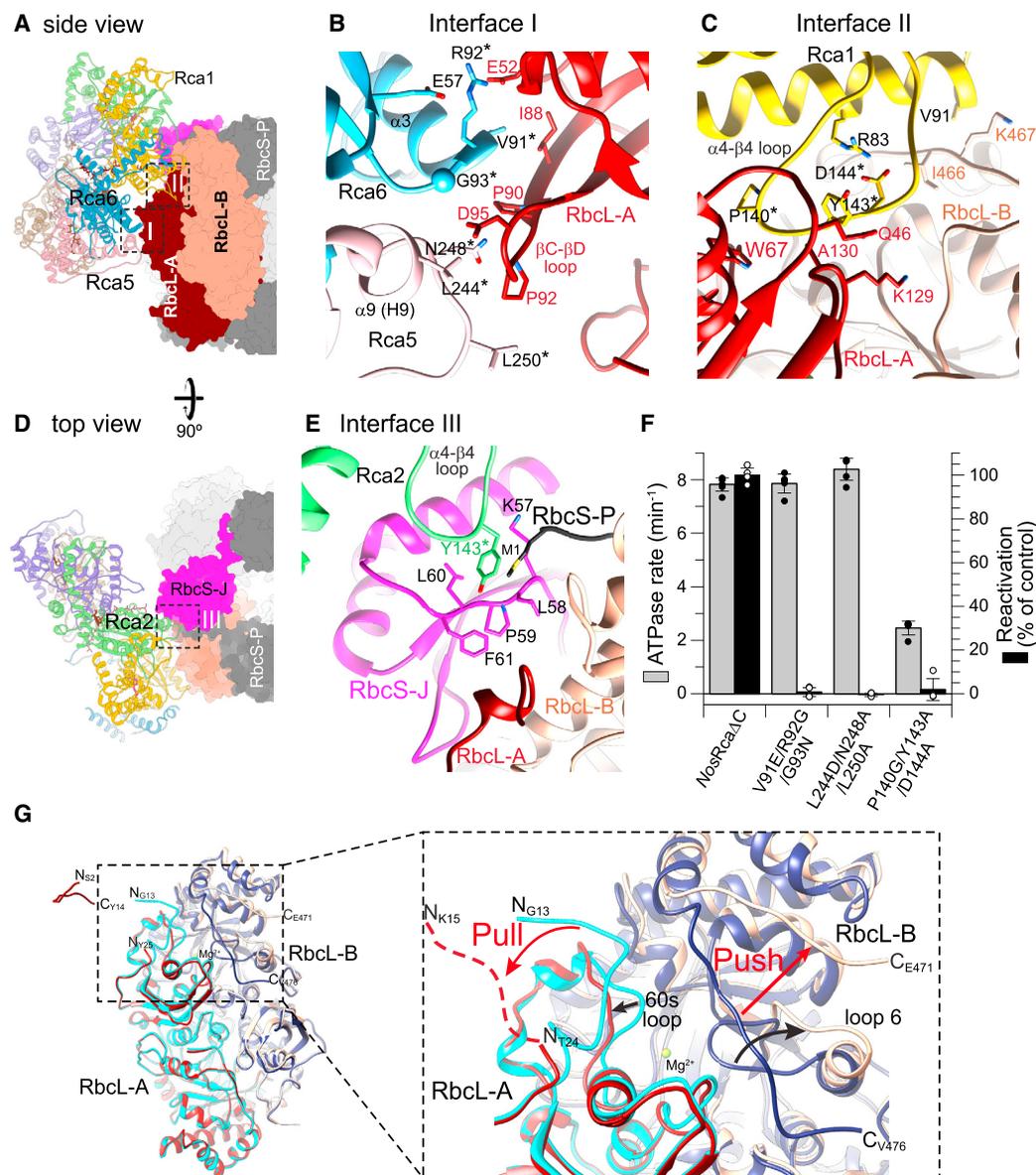


Figure 5. Structural Details of the *NosRca:Rubisco* Interaction

(A–E) Interfaces of the *NosRcaΔC:Rubisco* complex. Overview of the complex, highlighting interfaces I and II with *NosRca* in ribbon representation and Rubisco in surface representation (A). Interacting amino acid residues in interface I (B) and interface II (C). In interface I, Rca5 and Rca6 (subunits color-coded as in Figure 3B) interact with RbcL-A. Helix $\alpha 9$ (H9) in Rca5 and helix $\alpha 3$ in Rca6 are indicated. In interface II, the $\alpha 4$ - $\beta 4$ loop of Rca1 makes contacts with RbcL-A. Overview of the complex highlighting interface III (D). In interface III, the residue Tyr143 of the $\alpha 4$ - $\beta 4$ loop of Rca2 makes contacts with RbcS-J and RbcS-P (E). Side chains of contact residues are shown in stick representation. *, residue analyzed by mutation.

(F) Mutational analysis of *NosRca* interface residues. ATPase rates of *NosRcaΔC* and mutants (*NosRcaΔC* V91E/R92G/G93N, *NosRcaΔC* P140G/Y143A/D144A, *NosRcaΔC* L244D/N248A/L250A) and reactivation of CABP-inhibited *NosRubisco* were measured as in Figures 1B and 3D, respectively. ATPase rates were measured in the absence of Rubisco at 20 mM KCl (see STAR Methods). CO₂ fixation was measured for 8 min and set to 100% for wild-type *NosRcaΔC*. Error bars represent SD of at least three independent experiments.

(G) Structural remodeling of the Rubisco substrate binding pocket. Superposition of the binding pocket, formed at the interface between the N- and C-terminal domains of the antiparallel RbcL subunits, in the closed (RbcL-A, light blue; RbcL-B, dark blue) and the *NosRca*-engaged open state (RbcL-A, red; RbcL-B, peach). The remodeled regions, including the 60s loop of RbcL-A and loop 6 of RbcL-B as well as the pull and push actions of *NosRca* are indicated.

See also Videos S1 and S2.

Asn248 of the specificity helix H9, as well as Leu250 and Rca6 residues Val91 and Gly93 make van der Waals contacts with Glu52, Ile88, Pro90, Pro92, and Asp95 of RbcL-A (Figure 5B).

Arg92 in helix $\alpha 3$ of Rca6 forms a salt bridge with Glu52 of RbcL-A, and an intramolecular van der Waals interaction with Glu57 (Figure 5B). Interfaces II and III involve the $\alpha 4$ - $\beta 4$ loop of

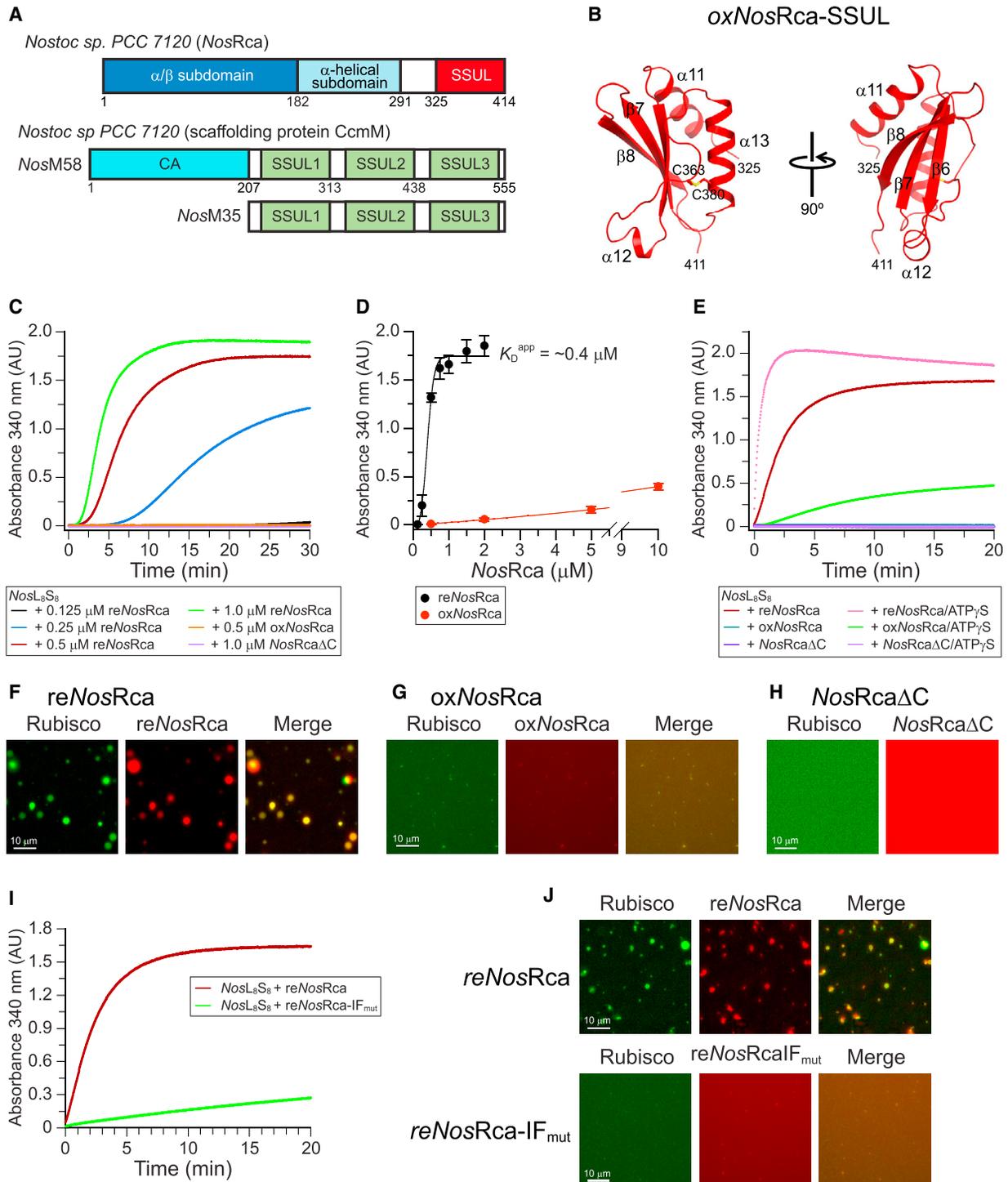


Figure 6. Cooperation of SSUL Domains and AAA+ Core in Packaging NosRca into Carboxysomes

(A) Domain structures of NosRca and NosCcmM (full-length NosM58 and the shorter isoform NosM35). CA, carbonic anhydrase domain.

(B) Crystal structure of the oxidized SSUL domain of NosRca in ribbon representation. Chain termini, secondary structure elements, and the disulfide bond (Cys360-Cys380) are indicated.

(C) Formation of NosRca-Rubisco condensate mediated by the SSUL domains analyzed by turbidity assay. Data shown are for Rubisco (NosL₆S₈; 0.25 μM) with increasing concentrations of reduced NosRca (reNosRca) or with 0.5 μM oxidized NosRca (oxNosRca) or 1 μM NosRcaΔC. Representative results from at least three independent experiments are shown.

(legend continued on next page)

Rca1 and Rca2, respectively (Figures 5C–5E). In interface II (600 Å²), subunit Rca1 inserts between the RbcL-A and RbcL-B antiparallel dimer (Figure 5C). The conformation of the protruding α 4- β 4 loop is rigidified by an intramolecular salt bridge between Arg83 and Asp144, with Pro140 having a further stabilizing effect on the loop. The α 4- β 4 loop contacts Trp67, which is part of the substrate binding pocket and is located in the 60s loop (residues 60–69) of RbcL-A. In addition, Tyr143 of Rca1 makes van der Waals contacts with Gln46, Lys129, and Ala130 of RbcL-A (Figure 5C). Note that there is also a steric interaction between Val91 of Rca1 and the C-terminal residues Ile466 and Lys467 of RbcL-B, which may contribute to keep the substrate pocket open for unhindered inhibitor release (Figure 5C). In interface III (426 Å²) the α 4- β 4 loop of Rca2 contacts the RbcS subunits J and P (Figures 5D and 5E). Specifically, Tyr143 of Rca2 engages in van der Waals interactions with the β 1- α 2 loop (K57–F61) of RbcS-J and Met1 of RbcS-P (Figure 5E). Residues Arg92, Gly93, and Asn248 of interface I, as well as Pro140 and Tyr143 of interface II/III, are conserved in plant Rca (Data S1), providing additional support for a conserved mechanism of remodeling of green-type form IB Rubisco.

Next, we performed a mutational analysis of *NosRca* Δ C to test the functional significance of these interactions. The mutant proteins preserved the hexameric state (Table S1). The interface I mutants V91E/R92G/G93N (breaking the Arg92–Glu52 salt bridge and the van der Waals interactions with Ile88 and Pro90) and L244D/N248A/L250A (disrupting the interaction with Asp95 and Pro92) preserved wild-type ATPase activity but failed to reactivate the CABP-inhibited Rubisco (Figures 5F and S1A). The interface II/III mutant P140G/Y143A/D144A (disrupting the interaction of the α 4- β 4 loop with Rubisco and destabilizing the loop) also abolished reactivation but additionally showed a \sim 70% reduced ATPase activity (Figures 5F and S1A). Note that the mutated residues are distant from the ATPase site, suggesting an indirect effect. Mutational analyses of plant Rca have also implicated residues in helix α 3, helix α 9 (H9), and the α 4- β 4 loop in the interaction with Rubisco (Li et al., 2005; Ott et al., 2000; Portis et al., 2008; Shivhare and Mueller-Cajal, 2017; Shivhare et al., 2019), supporting the functional significance of the structurally defined interface regions in the *NosRca* Δ C:Rubisco complex.

How does *NosRca* remodel Rubisco? The *NosRca* Δ C:Rubisco complex represents a post-remodeling state, as the Rubisco substrate binding pocket that is engaged by Rca is in an

open conformation with no inhibitor bound (Figure S4A). Super-position of the open and closed conformations of the substrate binding pockets revealed the structural changes effected by *NosRca* (Figure 5G): (1) the N-terminal residues 15–23 of RbcL-A are disordered but are well resolved in all other subunits, suggesting that binding the N-terminal sequence (residues 2–14) in the Rca pore destabilizes this region; (2) as a consequence, the 60s loop of RbcL-A, which contributes to stabilizing the substrate binding pocket, is displaced (Figure 5G); (3) in addition, the Rca hexamer—with the bulk of Rca1 α 4- β 4 loop acting as a wedge—displaces the C-terminal peptide of the adjacent RbcL subunit from its closed position over the substrate pocket (RbcL-B) (Video S2), thereby disrupting its interaction with loop 6. As a result, loop 6 is retracted from its closed position over the substrate pocket (Figure 5G). The combined displacement of the 60s loop and loop 6, both of which point to the solvent in the post-remodeling state, results in the complete opening of the substrate binding site, facilitating inhibitor release.

Function of the SSUL Domains and the AAA+ Core in Recruitment to Carboxysomes

A distinguishing feature of the *NosRca* subunit is the presence of the C-terminal SSUL domain (Figure 1A), which is connected to the AAA+ core (residues 2–291) by a \sim 35 residue flexible linker. The last common structured residue in the EM density map of all *NosRca* Δ C subunits is Asp281. The SSUL domain shares 54%–64% sequence identity with the three SSUL modules of the CcmM protein (*NosM58*) and its shorter variant M35 (*NosM35*) (Long et al., 2010) (Figures 6A and S6A). The SSUL modules of CcmM function to concentrate Rubisco into a dense condensate prior to encapsulation into the β -carboxysome (Ryan et al., 2019; Wang et al., 2019). Thus, it seemed plausible that the SSUL domains recruit *NosRca* into the Rubisco condensate (Lechno-Yossef et al., 2020). Notably, while in M58 and M35 three to five SSUL modules are linked to provide the multi-valency for Rubisco network formation, in *NosRca* each subunit of the hexamer carries its own SSUL, presumably generating high avidity for Rubisco.

The crystal structure of *NosRca*-SSUL (*NosSSUL*, residues 325–414; Figure 6A), solved by MAD at 1.4 Å resolution (Figures 6B, S6B, and S6C; Table S2), confirmed the high degree of similarity to the structure of the SSUL1 module in CcmM of *Synechococcus elongatus* PCC 7942 (SeSSUL) (rmsd 1.02–1.04 Å for matching C α positions) (Figure S6D). Like SeSSUL, *NosSSUL*

(D) Apparent binding affinity (K_D^{app}) to Rubisco. Turbidity was measured as in (C) at increasing concentrations of re*NosRca* or ox*NosRca*. Turbidity values reached after 10 min are plotted. Data represent averages with SD of three independent replicates.

(E) Nucleotide and redox-dependence of the *NosRca*-Rubisco interaction. Rubisco (0.25 μ M) was combined with re*NosRca*, ox*NosRca*, or *NosRca* Δ C (0.5 μ M) in the presence or absence of ATP γ S (2 mM). Turbidity assays were performed as in (C). Representative results from at least three independent experiments are shown.

(F–H) Formation of *NosRca*-Rubisco condensate as observed by fluorescence microscopy. Fluorophore labeled proteins (A405-re*NosRca*, A405-ox*NosRca*, A405-*NosRca* Δ C, and A532-*NosRubisco*) were mixed 1:10 with the respective unlabeled proteins and used at total concentrations of *NosRca* and Rubisco of 0.5 μ M and 0.25 μ M, respectively. Rubisco was mixed with either re*NosRca* (F), ox*NosRca* (G), or *NosRca* Δ C (H) in the presence of 2 mM ATP γ S. See STAR Methods for details.

(I) *NosRca*-Rubisco condensate formation analyzed as in (C). Rubisco (0.25 μ M) was combined with re*NosRca* (top) or re*NosRca*-IF_{mut} (0.5 μ M). Representative results from at least three independent experiments are shown.

(J) Inability of re*NosRca*-IF_{mut} to form a condensate with Rubisco, observed by fluorescence microscopy. Fluorophore-labeled A532-re*NosRubisco* (top) and A405-re*NosRca*-IF_{mut} (bottom) were used as described in (F–H) without addition of nucleotide.

See also Figures S6 and S7, and Table S3.

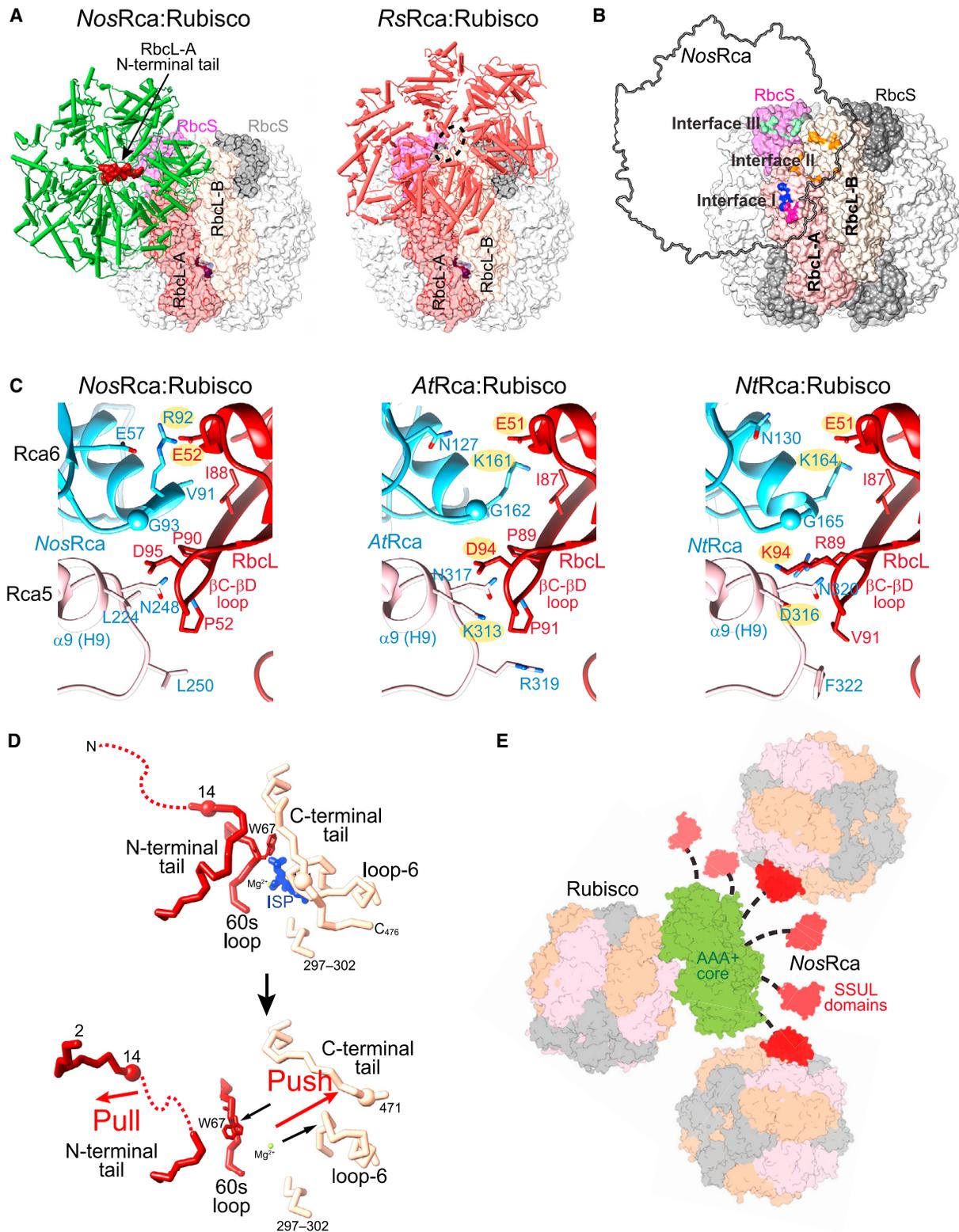


Figure 7. Mechanism of *NosRca* in Rubisco Reactivation and Recruitment into Carboxysome

(A) Comparison of *NosRca* Δ C (left) and *RsRca* (right) in complex with their cognate Rubisco. In both structures, Rubisco is shown in surface representation in the same orientation. The antiparallel RbcL dimer and the RbcS subunits involved in the interaction with Rca are indicated. The black-dashed circle indicates the position of the central pore of *RsRca*. *RsRca:Rubisco* model, EMDb: EMD-3701/PDB: 3ZUH and PDB: 5NV3.

(legend continued on next page)

contains a disulfide bond (Cys363 to Cys380) in the crystal (Figure 6B). The redox state of SeSSUL has been shown to regulate the affinity for Rubisco in condensate formation, with higher affinity in the reducing cytosol and lower affinity in the oxidizing carboxysome interior (Wang et al., 2019).

To investigate whether NosRca interacts with Rubisco in a manner similar to SeCcmM, we used a turbidity assay to monitor Rubisco network formation (Wang et al., 2019). We incubated Rubisco (0.25 μ M hexadecamer) with increasing concentrations of reduced NosRca (reNosRca; 0.125 to 1.0 μ M hexamer) (Figures 6C and S6E). Turbidity at 340 nm developed within minutes dependent on NosRca concentration (Figure 6C), consistent with a recent report for the Rca-like protein from *Fremyella diplosiphon* (Lechno-Yossef et al., 2020). No turbidity was observed with NosRca Δ C (Figure 6C), indicating that the interaction with Rubisco is mediated by the SSUL domains. reNosRca displayed an apparent affinity (K_D^{app}) for Rubisco of \sim 0.4 μ M, similar to the previously analyzed interaction of reduced CcmM (SeM35) with Rubisco (\sim 0.2 μ M) (Wang et al., 2019). In contrast, incubation of oxidized NosRca (oxNosRca) with Rubisco resulted in only low turbidity (Figure 6C), precluding estimation of apparent binding affinity (Figure 6D). Note that oxNosRca also formed ring complexes and was functionally active (Figures S6F and S6G). Cryo-EM analysis resolved the interaction of reNosRca with Rubisco mediated by NosSSUL at 8.2 \AA resolution (Figures S7A–S7D). Docking of the NosSSUL crystal structure and the NosRubisco model into the low-resolution EM density map revealed that the SSUL module binds close to the equatorial region of the hexadecamer in a groove between RbcL dimers (Figure S7E), as previously shown for the SeM35-Rubisco complex (Wang et al., 2019).

Interestingly, addition of ATP γ S to reNosRca resulted in enhanced turbidity with Rubisco (Figure 6E). However, this effect was still completely SSUL-dependent (Figure 6E), suggesting that the AAA+ core modulates the interaction by contributing additional valency. Consistent with such an enhancing effect, addition of ATP γ S also resulted in observable binding of 0.5 μ M oxNosRca to Rubisco (Figure 6E).

To determine whether NosRca and Rubisco can undergo demixing into phase separated droplets, we N-terminally labeled Rubisco with Alexa Fluor 532 (Rubisco-A532) and NosRca (reNosRca, oxNosRca, or NosRca Δ C) with Alexa Fluor 405 (see STAR Methods). Upon incubation of reNosRca with Rubisco (labeled:unlabeled 1:10) in the presence of ATP γ S, liquid-liquid phase separation (LLPS) into fluorescent droplets was readily observed (Figure 6F). Consistent with the turbidity measurements, only very small droplets were observed with oxNosRca

(Figure 6G). No droplet formation occurred with NosRca Δ C and Rubisco (Figure 6H). The phenomenon of LLPS observed *in vitro* indicates that NosRca engages in multivalent interactions with Rubisco.

To assess whether the AAA+ core of NosRca contributes directly to LLPS, we generated a mutant of NosRca with disrupted interface I (mutations V91E/R92G/G93N and L244D/N248A/L250A combined) (NosRca-IF_{mut}). reNosRca-IF_{mut} was a stable hexamer in solution (Figure S1A; Table S1). The ability of reNosRca-IF_{mut} to induce Rubisco condensate formation was strongly reduced (Figure 6I), and only very small droplets were observed, despite the presence of the SSUL domains (Figure 6J). Thus, both the SSUL domains and the AAA+ core contribute critical valency to ensure packaging of NosRca into carboxysomes together with Rubisco.

DISCUSSION

Distinct Rca Mechanisms for Green- and Red-type Rubisco

Our structural and functional analysis of cyanobacterial NosRca, a close homolog of plant Rca, revealed a mechanism of metabolic repair for form 1B Rubisco that differs from the previously described mechanism of the proteobacterial red-type RsRca (Bhat et al., 2017b; Mueller-Cajar et al., 2011). NosRca, like RsRca, docks onto inhibitor-bound Rubisco with its hexamer pore positioned over one of the eight catalytic sites. In contrast to RsRca, which engages the extended RbcL C terminus (Bhat et al., 2017a), NosRca is rotated clockwise by \sim 50 degrees and binds the RbcL N terminus (residues 2–14) in its central pore (Figure 7A). The cryo-EM structure of the complex apparently represents an end state of the remodeling cycle. Presumably, ATP-dependent peptide threading comes to a halt when this stably locked position is reached, with the staggered pore-loops PL1 and PL2 of adjacent Rca subunits providing binding pockets for sequence-specific peptide capture. As suggested for other AAA+ proteins, peptide threading may be mediated by sequential ATP hydrolysis around the hexamer ring (de la Peña et al., 2018; Dong et al., 2019; Fei et al., 2020; Ripstein et al., 2020; Rizo et al., 2019; Twomey et al., 2019; Wang et al., 2020). Insertion of the RbcL N terminus into the Rca central pore is likely of general relevance in organisms with green-type form 1B Rubisco (cyanobacteria, green algae, and plants). Note that in plant Rca the mechanism of reactivation may be modulated by a flexible N-terminal domain of \sim 60 residues, not present in NosRca (Esau et al., 1996; Stotz et al., 2011).

(B) Overview of the surface areas on NosRubisco interacting with NosRca Δ C via interfaces I (red and dark blue), II (orange), and III (light green). NosRca Δ C is shown as an outline (Video S1).

(C) Structural models for interface I in the AtRca:AtRubisco (middle) and NtRca:NtRubisco (right) complexes. Interface I in the NosRca:NosRubisco complex is shown on the left, with protein backbones in ribbon representation and side chains of interacting amino acids in stick representation. Sequence differences to *Nostoc* and side-chain conformers were manually adjusted.

(D) Structural rearrangement of the Rubisco substrate binding pocket by NosRca from the closed state with bound inhibitory sugar phosphate (ISP) to the open state (ISP released) (Video S2).

(E) Cooperation of the SSUL domains and the AAA+ core for packaging of NosRca into carboxysomes with Rubisco. The model is based on the interactions observed in the cryo-EM reconstructions of NosRca Δ C:Rubisco and NosSSUL-Rubisco. The dashed lines represent the \sim 35 residue flexible linker between the AAA+ core (green) and the SSUL domains (red).

See also Videos S1 and S2.

Why have distinct Rca mechanisms of reactivation, employing N- or C-terminal interaction, evolved for green- and red-type Rubisco proteins, considering that the catalytic site is essentially conserved? Upon binding of RuBP, the Rubisco active site is closed by both the 60s loop from the N-terminal domain of one RbcL subunit and loop 6 of the C-terminal domain of the adjacent subunit. The N-terminal sequence pins down the 60s loop and the C terminus stretches over loop 6, generating the solvent-sealed environment required for catalysis. *NosRbcL* and *NtRbcL* lack the extended C terminus of red-type RbcL (~12 residues) and have only 2 and 4 residues, respectively, beyond the conserved Asp (D474 in *NosRbcL*) that forms a critical attachment point to stably close the catalytic site (Duff et al., 2000). Thus, pulling on the flexible N terminus has evolved as an alternative means for Rca to open the catalytic pocket of inhibitor-bound form IB Rubisco.

Mechanism of Form 1B Rubisco Metabolic Repair

The complex of *NosRca* Δ C with inhibitor-bound Rubisco, stabilized by a nucleotide substitution strategy, represents a post-remodeling state in which the substrate binding site under repair is open and the inhibitor released. In addition to insertion of the N terminus of one RbcL (RbcL-A) into the Rca hexamer pore, destabilization of the binding pocket involves extensive interactions with Rubisco via three interface regions (Figure 7B). Underscoring the sequence homology to plant Rca, critical residues of these interfaces are conserved (Data S1). Interface I formed by Rca5/Rca6 with RbcL-A includes the specificity helix H9 on Rca5 and the β C- β D loop on RbcL (Figure 7C). In plants, helix H9 functions in distinguishing Rubisco proteins from solanaceae and non-solanaceae (Portis et al., 2008; Wachter et al., 2013). Notably, the RbcL residues Pro89 and Asp94, which mediate Rca specificity in non-solanaceous plants (e.g., *A. thaliana* and *Spinacia oleracea*), are conserved in *NosRbcL* (Pro90 and Asp95) (Li et al., 2005; Ott et al., 2000) (Figure 7C). Pro90 contacts the *NosRca* backbone at Val91 and Gly93 in the α 3- β 3 loop of Rca6. This loop is one residue shorter in plant Rca (Data S1 and S2). Asp95 in *NosRbcL* makes van der Waals contacts to Leu244 in helix H9 of *NosRca*5 (Figure 7C). Modeling of the *AtRca*:Rubisco complex based on the *NosRca*:Rubisco structure suggests that Asp94 in *AtRbcL* is in electrostatic interaction with Lys313 in *AtRca* (the residue equivalent to Leu244 in *NosRca*) (Figure 7C). This interaction mediates the Rca-Rubisco specificity of non-solanaceous plants, consistent with previous mutational analyses (Li et al., 2005; Ott et al., 2000; Portis et al., 2008). In tobacco, a solanaceous plant, this electrostatic interaction is reversed, with Lys94 on *NtRbcL* interacting with Asp316 in *NtRca* (Figure 7C), and the charge reversal switching the specificity from non-solanaceous to solanaceous (Portis et al., 2008). Another conserved interaction in interface I is the salt bridge Arg92 in *NosRca*6 with Glu52 in RbcL, which is equivalent to the salt bridges of Lys161 to Glu51 and Lys164 to Glu51 in *A. thaliana* and tobacco, respectively (Figure 7C).

Interfaces II and III involve the α 4- β 4 loop of Rca1 and Rca2. In interface II, the α 4- β 4 loop of Rca1 inserts between the RbcL-A and RbcL-B subunits of the antiparallel dimer containing the catalytic site under repair. In interface III, the α 4- β 4 loop

of Rca2 makes contact with RbcS subunits J and P. As a result of these combined interactions, the catalytic site is converted to a completely open state. Specifically, the N-terminal tail and the 60s loop of RbcL-A and the C-terminal tail and loop 6 of the adjacent RbcL-B subunit are displaced (Figure 7D). The interaction of *NosRca* with RbcS is intriguing. RbcS contributes indirectly to stabilizing the sugar substrate in the catalytic site pocket (Bracher et al., 2011) and thus the interaction with *NosRca*, in addition to providing an anchoring point, may also prevent dissociation of RbcS while the catalytic site is being destabilized.

Recruitment of *NosRca* to Carboxysomes

Another important feature of *NosRca* is the presence of a Rubisco SSUL domain flexibly attached to the C terminus of the AAA+ core in each Rca subunit. As shown for the Rca-like protein of *Fremyella diplosiphon* (Lechno-Yossef et al., 2020), these SSUL domains mediate an alternative mode of interaction with Rubisco to ensure that *NosRca* becomes encapsulated with Rubisco during carboxysome biogenesis. They bind into a groove close to the equatorial region of Rubisco between the RbcL dimer units, as shown for the SSUL domains of the carboxysomal scaffolding protein CcmM (Ryan et al., 2019; Wang et al., 2019). *NosSSUL* also contains the cysteine pair previously described for the SSUL module in CcmM (Wang et al., 2019), and thus its interaction with Rubisco is redox regulated, resulting in reduced affinity under oxidizing conditions.

In contrast to CcmM, which contains three to five SSUL modules as a linear fusion, *NosRca* contains six separate SSUL modules, one on each subunit, providing multi-valency for Rubisco binding (Figure 7E). Interestingly, we found that the AAA+ core of *NosRca* contributes critical additional valency. Thus, the seemingly unrelated interactions of *NosRca* in Rubisco repair and binding to Rubisco via the SSUL domains are linked and must cooperate to ensure Rca packing into carboxysomes (Figure 7E). Upon transfer into the oxidizing environment of the carboxysome, the SSUL-mediated association would then be released in favor of the functional interaction for Rubisco repair via the AAA+ core. It will be interesting to understand the interplay of *NosRca* and CcmM during carboxysome biogenesis.

The acquisition of the SSUL domain by the *NosRca* AAA+ core probably occurred in the context of the appearance of carboxysomes about 350 million years ago, long after the primary endosymbiotic event leading to the evolution of chloroplasts (Rae et al., 2013). Efforts to introduce carboxysomes into chloroplasts in order to increase Rubisco efficiency (Hanson et al., 2016; Henancy and Jonikas, 2020; Long et al., 2016; Rae et al., 2017) will have to consider coexpression of cyanobacterial Rca or attachment of the SSUL domain to plant activase.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

- Lead Contact
- Materials Availability
- Data and Code Availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Strains
 - Plasmids and proteins
- **METHOD DETAILS**
 - Protein expression and purification
 - CABP synthesis
 - Enzymatic assays
 - Turbidity assay
 - Liquid-liquid phase separation
 - Size-exclusion chromatography coupled to multi-angle static light scattering (SEC-MALS)
 - Electron microscopy and reconstruction
 - Crystallization and data collection
 - Structure analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cell.2020.09.010>.

ACKNOWLEDGMENTS

The technical assistance in the cloning and purification of proteins by S. Gaertner and R. Lange is gratefully acknowledged. We thank F. Bonneau and C. Basquin (Department of Structural Cell Biology) for help with the screening of a suitable buffer for protein purification, G. Thieulin-Pardo for providing *NtRcaR294V* and *NtRubisco* as well as advice on Rca assays, R.H. Wilson for advice on expression of cyanobacterial Rubisco, X. Yan for advice on fluorescence microscopy, P. Wendler (University Potsdam) for initial analysis of *NosRca* by negative-stain EM, D. Cobessi (Institut de Biologie Structurale, Grenoble) for help with X-ray data collection of *NosSSUL*, and D. Bollschweiler for advice with cryo-EM data collection. We are grateful to the staff at FIP (French beamline for Investigation of Proteins) beamline BM30A at the European Synchrotron Radiation Facility (ESRF) and at beamline X06DA at the Swiss Synchrotron Light Source (SLS) in Villigen, Switzerland. Expert support by the Imaging, Crystallization, and cryo-EM facilities of the Max Planck Institute of Biochemistry is acknowledged.

AUTHOR CONTRIBUTIONS

M.F. and H.W. planned and performed most of the experiments. M.F. established and performed all the activase assays and carried out turbidity assays together with H.W. M.F. was responsible for construct development and protein purification and performed the synthesis of CABP. H.W. performed the cryo-EM structural analysis, including sample screening, grid preparation, data acquisition, and single-particle reconstructions, and fluorescence microscopy. L.P. contributed to cloning and protein purification and obtained crystals of *NosRcaΔC* and *NosSSUL*. A.B. supervised the structural analyses and solved the crystal structures. M.H.-H. conceived the project and participated in data interpretation with F.U.H. and the other authors. M.H.-H. and F.U.H. wrote the manuscript with input from H.W., A.B., and M.F.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 16, 2020

Revised: July 13, 2020

Accepted: September 1, 2020

Published: September 25, 2020

REFERENCES

- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221.
- Aigner, H., Wilson, R.H., Bracher, A., Calisse, L., Bhat, J.Y., Hartl, F.U., and Hayer-Hartl, M. (2017). Plant RuBisCo assembly in *E. coli* with five chloroplast chaperones including BSD2. *Science* 358, 1272–1278.
- Andralojc, P.J., Carmo-Silva, E., Degen, G.E., and Parry, M.A.J. (2018). Increasing metabolic potential: C-fixation. *Essays Biochem.* 62, 109–118.
- Avellaneda, M.J., Franke, K.B., Sunderlikova, V., Bukau, B., Mogk, A., and Tans, S.J. (2020). Processive extrusion of polypeptide loops by a Hsp100 disaggregase. *Nature* 578, 317–320.
- Badger, M.R., and Price, G.D. (2003). CO₂ concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J. Exp. Bot.* 54, 609–622.
- Bailey-Serres, J., Parker, J.E., Ainsworth, E.A., Oldroyd, G.E.D., and Schroeder, J.I. (2019). Genetic strategies for improving crop yields. *Nature* 575, 109–118.
- Baker, R.T., Catanzariti, A.M., Karunasekara, Y., Soboleva, T.A., Sharwood, R., Whitney, S., and Board, P.G. (2005). Using deubiquitylating enzymes as research tools. *Methods Enzymol.* 398, 540–554.
- Bar-On, Y.M., and Milo, R. (2019). The global mass and average rate of rubisco. *Proc. Natl. Acad. Sci. USA* 116, 4738–4743.
- Barta, C., Carmo-Silva, A.E., and Salvucci, M.E. (2011). Rubisco activase activity assays. *Methods Mol. Biol.* 684, 375–382.
- Bhat, J.Y., Miličić, G., Thieulin-Pardo, G., Bracher, A., Maxwell, A., Ciniawsky, S., Mueller-Cajar, O., Engen, J.R., Hartl, F.U., Wendler, P., and Hayer-Hartl, M. (2017a). Mechanism of enzyme repair by the AAA+ chaperone Rubisco activase. *Mol. Cell* 67, 744–756.e6.
- Bhat, J.Y., Thieulin-Pardo, G., Hartl, F.U., and Hayer-Hartl, M. (2017b). Rubisco Activases: AAA+ Chaperones Adapted to Enzyme Repair. *Front. Mol. Biosci.* 4, 20.
- Biyani, N., Righetto, R.D., McLeod, R., Caujolle-Bert, D., Castano-Diez, D., Goldie, K.N., and Stahlberg, H. (2017). Focus: The interface between data collection and data processing in cryo-EM. *J. Struct. Biol.* 198, 124–133.
- Blayney, M.J., Whitney, S.M., and Beck, J.L. (2011). NanoESI mass spectrometry of Rubisco and Rubisco activase structures and their interactions with nucleotides and sugar phosphates. *J. Am. Soc. Mass Spectrom.* 22, 1588–1601.
- Bracher, A., Starling-Windhof, A., Hartl, F.U., and Hayer-Hartl, M. (2011). Crystal structure of a chaperone-bound assembly intermediate of form I Rubisco. *Nat. Struct. Mol. Biol.* 18, 875–880.
- Bracher, A., Sharma, A., Starling-Windhof, A., Hartl, F.U., and Hayer-Hartl, M. (2015). Degradation of potent Rubisco inhibitor by selective sugar phosphatase. *Nat. Plants* 1, 14002.
- Bracher, A., Whitney, S.M., Hartl, F.U., and Hayer-Hartl, M. (2017). Biogenesis and metabolic maintenance of Rubisco. *Annu. Rev. Plant Biol.* 68, 29–60.
- Catanzariti, A.M., Soboleva, T.A., Jans, D.A., Board, P.G., and Baker, R.T. (2004). An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Sci.* 13, 1331–1339.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12–21.
- Chifflet, S., Torriglia, A., Chiesa, R., and Tolosa, S. (1988). A method for the determination of inorganic phosphate in the presence of labile organic phosphate and high concentrations of protein: application to lens ATPases. *Anal. Biochem.* 168, 1–4.
- Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 50, 760–763.

- Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011.
- de La Fortelle, E., and Bricogne, G. (1997). [27] Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multi-wavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494.
- de la Peña, A.H., Goodall, E.A., Gates, S.N., Lander, G.C., and Martin, A. (2018). Substrate-engaged 26S proteasome structures reveal mechanisms for ATP-hydrolysis-driven translocation. *Science* **362**, eaav0725.
- Dong, Y., Zhang, S., Wu, Z., Li, X., Wang, W.L., Zhu, Y., Stoilova-McPhie, S., Lu, Y., Finley, D., and Mao, Y. (2019). Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. *Nature* **565**, 49–55.
- Duff, A.P., Andrews, T.J., and Curmi, P.M. (2000). The transition between the open and closed states of rubisco is triggered by the inter-phosphate distance of the bound biphosphate. *J. Mol. Biol.* **298**, 903–916.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132.
- Esau, B.D., Snyder, G.W., and Portis, A.R., Jr. (1996). Differential effects of N- and C-terminal deletions on the two activities of rubisco activase. *Arch. Biochem. Biophys.* **326**, 100–105.
- Éva, C., Oszvald, M., and Tamás, L. (2019). Current and possible approaches for improving photosynthetic efficiency. *Plant Sci.* **280**, 433–440.
- Evans, P.R. (1997). Scala. Joint CCP4 and ESF-EACBM Newsletter **33**, pp. 22–24.
- Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82.
- Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214.
- Fei, X., Bell, T.A., Jenni, S., Stinson, B.M., Baker, T.A., Harrison, S.C., and Sauer, R.T. (2020). Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate. *eLife* **9**, e52774.
- French, G., and Wilson, K. (1978). On the treatment of negative intensity observations. *Acta Crystallogr. A* **34**, 517–525.
- Gouet, P., Courcelle, E., Stuart, D.I., and Métoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308.
- Hanson, M.R., Lin, M.T., Carmo-Silva, A.E., and Parry, M.A. (2016). Towards engineering carboxysomes into C3 plants. *Plant J.* **87**, 38–50.
- Hartman, F.C., Soper, T.S., Niyogi, S.K., Mural, R.J., Foote, R.S., Mitra, S., Lee, E.H., Machanoff, R., and Larimer, F.W. (1987). Function of Lys-166 of *Rhodospirillum rubrum* ribulosebiphosphate carboxylase/oxygenase as examined by site-directed mutagenesis. *J. Biol. Chem.* **262**, 3496–3501.
- Hasse, D., Larsson, A.M., and Andersson, I. (2015). Structure of *Arabidopsis thaliana* Rubisco activase. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 800–808.
- Hazra, S., Henderson, J.N., Liles, K., Hilton, M.T., and Wachter, R.M. (2015). Regulation of ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) activase: product inhibition, cooperativity, and magnesium activation. *J. Biol. Chem.* **290**, 24222–24236.
- Hennacy, J.H., and Jonikas, M.C. (2020). Prospects for engineering biophysical CO₂ concentrating mechanisms into land plants to enhance yields. *Annu. Rev. Plant Biol.* **71**, 461–485.
- Kabsch, W. (2010). XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132.
- Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637.
- Karkehabadi, S., Peddi, S.R., Anwaruzzaman, M., Taylor, T.C., Cederlund, A., Genkov, T., Andersson, I., and Spreitzer, R.J. (2005). Chimeric small subunits influence catalysis without causing global conformational changes in the crystal structure of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Biochemistry* **44**, 9851–9861.
- Keown, J.R., and Pearce, F.G. (2014). Characterization of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase activase isoforms reveals hexameric assemblies with increased thermal stability. *Biochem. J.* **464**, 413–423.
- Kerner, M.J., Naylor, D.J., Ishihama, Y., Maier, T., Chang, H.C., Stines, A.P., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M., and Hartl, F.U. (2005). Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **122**, 209–220.
- Kleywegt, G.T., and Jones, T.A. (1994). A super position. CCP4/ESF-EACBM Newslett. *Protein Crystallog.* **31**, 9–14.
- Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.
- Larimer, F.W., and Soper, T.S. (1993). Overproduction of *Anabaena* 7120 ribulose-bisphosphate carboxylase/oxygenase in *Escherichia coli*. *Gene* **126**, 85–92.
- 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.
- Li, L.A., Zianni, M.R., and Tabita, F.R. (1999). Inactivation of the monocistronic rca gene in *Anabaena variabilis* suggests a physiological ribulose biphosphate carboxylase/oxygenase activase-like function in heterocystous cyanobacteria. *Plant Mol. Biol.* **40**, 467–478.
- Li, C., Salvucci, M.E., and Portis, A.R., Jr. (2005). Two residues of rubisco activase involved in recognition of the Rubisco substrate. *J. Biol. Chem.* **280**, 24864–24869.
- Lin, M.T., Stone, W.D., Chaudhari, V., and Hanson, M.R. (2019). Enzyme kinetics of tobacco Rubisco expressed in *Escherichia coli* varies depending on the small subunit composition. *bioRxiv*.
- 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.
- Long, B.M., Rae, B.D., Rolland, V., Förster, B., and Price, G.D. (2016). Cyanobacterial CO₂-concentrating mechanism components: function and prospects for plant metabolic engineering. *Curr. Opin. Plant Biol.* **31**, 1–8.
- Mastrorade, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51.
- Matsumura, H., Mizohata, E., Ishida, H., Kogami, A., Ueno, T., Makino, A., Inoue, T., Yokota, A., Mae, T., and Kai, Y. (2012). Crystal structure of rice Rubisco and implications for activation induced by positive effectors NADPH and 6-phosphogluconate. *J. Mol. Biol.* **422**, 75–86.
- Mziorko, H.M., and Lorimer, G.H. (1983). Ribulose-1,5-bisphosphate carboxylase-oxygenase. *Annu. Rev. Biochem.* **52**, 507–535.
- Mueller-Cajar, O. (2017). The diverse AAA+ machines that repair inhibited Rubisco active sites. *Front. Mol. Biosci.* **4**, 31.
- Mueller-Cajar, O., Stotz, M., Wendler, P., Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Structure and function of the AAA+ protein CbbX, a red-type Rubisco activase. *Nature* **479**, 194–199.
- Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367.
- Newman, J., Branden, C.I., and Jones, T.A. (1993). Structure determination and refinement of ribulose 1,5-bisphosphate carboxylase/oxygenase from *Synechococcus* PCC6301. *Acta Crystallogr. D Biol. Crystallogr.* **49**, 548–560.
- Olivares, A.O., Baker, T.A., and Sauer, R.T. (2016). Mechanistic insights into bacterial AAA+ proteases and protein-remodelling machines. *Nat. Rev. Microbiol.* **14**, 33–44.
- Ott, C.M., Smith, B.D., Portis, A.R., Jr., and Spreitzer, R.J. (2000). Activase region on chloroplast ribulose-1,5-bisphosphate carboxylase/oxygenase. Nonconservative substitution in the large subunit alters species specificity of protein interaction. *J. Biol. Chem.* **275**, 26241–26244.
- Panjikar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M.S., and Tucker, P.A. (2005). Auto-rickshaw: an automated crystal structure determination platform

- as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 449–457.
- Pape, T., and Schneider, T.R. (2004). HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Cryst.* **37**, 843–844.
- Parry, M.A.J., Keys, A.J., Madgwick, P.J., Carmo-Silva, A.E., and Andralojc, P.J. (2008). Rubisco regulation: a role for inhibitors. *J. Exp. Bot.* **59**, 1569–1580.
- Pearce, F.G. (2006). Catalytic by-product formation and ligand binding by ribulose biphosphate carboxylases from different phylogenies. *Biochem. J.* **399**, 525–534.
- Petterson, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612.
- Pierce, J., Tolbert, N.E., and Barker, R. (1980). Interaction of ribulosebiphosphate carboxylase/oxygenase with transition-state analogues. *Biochemistry* **19**, 934–942.
- Portis, A.R., Jr., Li, C., Wang, D., and Salvucci, M.E. (2008). Regulation of Rubisco activase and its interaction with Rubisco. *J. Exp. Bot.* **59**, 1597–1604.
- Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1131–1137.
- Price, G.D., Sültemeyer, D., Klughammer, B., Ludwig, M., and Badger, M.R. (1998). The functioning of the CO₂ concentrating mechanism in several cyanobacterial strains: a review of general physiological characteristics, genes, proteins, and recent advances. *Can. J. Bot.* **76**, 973–1002.
- Puchades, C., Sandate, C.R., and Lander, G.C. (2020). The molecular principles governing the activity and functional diversity of AAA+ proteins. *Nat. Rev. Mol. Cell Biol.* **21**, 43–58.
- 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.
- 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.
- Ripstein, Z.A., Vahidi, S., Houry, W.A., Rubinstein, J.L., and Kay, L.E. (2020). A processive rotary mechanism couples substrate unfolding and proteolysis in the ClpXP degradation machinery. *eLife* **9**, e52158.
- Rizo, A.N., Lin, J., Gates, S.N., Tse, E., Bart, S.M., Castellano, L.M., DiMaio, F., Shorter, J., and Southworth, D.R. (2019). Structural basis for substrate gripping and translocation by the ClpB AAA+ disaggregase. *Nat. Commun.* **10**, 2393.
- Robinson, S.P., and Portis, A.R., Jr. (1989). Adenosine triphosphate hydrolysis by purified rubisco activase. *Arch. Biochem. Biophys.* **268**, 93–99.
- Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221.
- 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.
- Scheres, S.H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682.
- Servaites, J.C. (1985). Crystalline ribulose biphosphate carboxylase/oxygenase of high integrity and catalytic activity from *Nicotiana tabacum*. *Arch. Biochem. Biophys.* **238**, 154–160.
- Sharwood, R.E. (2017). Engineering chloroplasts to improve Rubisco catalysis: prospects for translating improvements into food and fiber crops. *New Phytol.* **213**, 494–510.
- Sheldrick, G.M. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479–485.
- Shivhare, D., and Mueller-Cajar, O. (2017). In vitro characterization of thermo-stable CAM Rubisco activase reveals a Rubisco interacting surface loop. *Plant Physiol.* **174**, 1505–1516.
- Shivhare, D., Ng, J., Tsai, Y.-C.C., and Mueller-Cajar, O. (2019). Probing the rice Rubisco-Rubisco activase interaction via subunit heterooligomerization. *Proc. Natl. Acad. Sci. USA* **116**, 24041–24048.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539.
- Singer, S.D., Soolanayakanahally, R.Y., Foroud, N.A., and Kroebel, R. (2019). Biotechnological strategies for improved photosynthesis in a future of elevated atmospheric CO₂. *Planta* **251**, 24.
- Slattery, R.A., and Ort, D.R. (2019). Carbon assimilation in crops at high temperatures. *Plant Cell Environ.* **42**, 2750–2758.
- Stotz, M., Mueller-Cajar, O., Ciniawsky, S., Wendler, P., Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Structure of green-type Rubisco activase from tobacco. *Nat. Struct. Mol. Biol.* **18**, 1366–1370.
- Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234.
- Tabita, F.R., Satagopan, S., Hanson, T.E., Kree, N.E., and Scott, S.S. (2008). Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J. Exp. Bot.* **59**, 1515–1524.
- Terwilliger, T.C. (2000). Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 965–972.
- Thévenaz, P., Ruttimann, U.E., and Unser, M. (1998). A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* **7**, 27–41.
- Turmo, A., Gonzalez-Esquer, C.R., and Kerfeld, C.A. (2017). Carboxysomes: metabolic modules for CO₂ fixation. *FEMS Microbiol. Lett.* **364**.
- Twomey, E.C., Ji, Z., Wales, T.E., Bodnar, N.O., Ficarro, S.B., Marto, J.A., Engen, J.R., and Rapoport, T.A. (2019). Substrate processing by the Cdc48 ATPase complex is initiated by ubiquitin unfolding. *Science* **365**, eaax1033.
- Vagin, A.A., and Isupov, M.N. (2001). Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 1451–1456.
- Wachter, R.M., Salvucci, M.E., Carmo-Silva, A.E., Barta, C., Genkov, T., and Spreitzer, R.J. (2013). Activation of interspecies-hybrid Rubisco enzymes to assess different models for the Rubisco-Rubisco activase interaction. *Photosynth. Res.* **117**, 557–566.
- Wang, Z.Y., and Portis, A.R. (1992). Dissociation of ribulose-1,5-bisphosphate bound to ribulose-1,5-bisphosphate carboxylase/oxygenase and its enhancement by ribulose-1,5-bisphosphate carboxylase/oxygenase activase-mediated hydrolysis of ATP. *Plant Physiol.* **99**, 1348–1353.
- 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.
- Wang, L., Myasnikov, A., Pan, X., and Walter, P. (2020). Structure of the AAA protein Msp1 reveals mechanism of mislocalized membrane protein extraction. *eLife* **9**, e54031.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018).

SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46 (W1), W296–W303.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the tidyverse. *J. Open Source Softw.* 4, 1686.

Wilson, R.H., Thieulin-Pardo, G., Hartl, F.U., and Hayer-Hartl, M. (2019). Improved recombinant expression and purification of functional plant Rubisco. *FEBS Lett.* 593, 611–621.

Wyatt, P.J. (1993). Light scattering and the absolute characterization of macromolecules. *Anal. Chim. Acta* 272, 1–40.

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* 14, 331–332.

STAR★METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|------------------------------------|---------------------------------------|
| Bacterial Strains | | |
| <i>E. coli</i> BL21 STAR | Invitrogen | Cat#C601003 |
| <i>E. coli</i> BL21 (DE3) | Stratagene | Cat#200131 |
| <i>E. coli</i> DH5 α | ThermoFischer | Cat#18265017 |
| <i>Nostoc</i> sp. PCC 7120 | Culture Collection of Algae (UTEX) | Cat#2576 |
| Biological Samples | | |
| <i>Nicotiana tabacum</i> Rubisco (<i>NtRubisco</i>) | This study | (Servaites, 1985; Stotz et al., 2011) |
| Chemicals, Peptides, and Recombinant Proteins | | |
| Alexa Fluor TM 532 NHS-Ester | ThermoFisher | Cat#A20101MP |
| Alexa Fluor TM 405 NHS-Ester | ThermoFisher | Cat#A30000 |
| RuBP | Sigma-Aldrich | Cat#R0878 |
| XuBP | This study | (Bracher et al., 2015) |
| CABP | This study | (Pierce et al., 1980) |
| ATP | Sigma-Aldrich | Cat#A3377 |
| ATP γ S | Sigma-Aldrich | Cat#A1388 |
| NaHCO ₃ | Sigma-Aldrich | Cat#1.06329 |
| NaH ¹⁴ CO ₃ | Perkin-Elmer | Cat#NEC086H005MC |
| K ¹⁴ CN | ARC | Cat#0136A |
| <i>NosRubisco</i> | This study | N/A |
| <i>NosRubisco</i> Δ N | This study | N/A |
| <i>NosRca</i> | This study | N/A |
| <i>NosRca</i> mutant (V91E/R92G/G93N/L244D/N248A/L250A) | This study | N/A |
| <i>NosRca</i> Δ C | This study | N/A |
| <i>NosRca</i> Δ C mutant (V91E/R92G/G93N) | This study | N/A |
| <i>NosRca</i> Δ C mutant (P140G/Y143A/D144A) | This study | N/A |
| <i>NosRca</i> Δ C mutant (L244D/N248A/L250A) | This study | N/A |
| <i>NosSSUL</i> | This study | N/A |
| <i>NtRubisco</i> Δ N | This study | N/A |
| <i>NtRca</i> (R294V) | This study | (Stotz et al., 2011) |
| Deposited Data | | |
| Coordinates and structure factors: <i>NosSSUL</i> | This study | PDB: 6HAS |
| Coordinates and structure factors: <i>NosRca</i> Δ C, native data | This study | PDB: 6Z1E |
| Coordinates and structure factors: <i>NosRca</i> Δ C, Gd-MAD | This study | PDB: 6Z1D |
| Coordinates: <i>NosRca</i> Δ C:Rubisco complex | This study | PDB: 6Z1F |
| Cryo-EM density map: <i>NosRca</i> Δ C:Rubisco complex | This study | EMDB: EMD-11028 |
| Cryo-EM local density map of <i>NosRca</i> Δ C in <i>NosRca</i> Δ C:Rubisco complex | This study | EMDB: EMD-11575 |
| Coordinates: <i>NosRca</i> :Rubisco complex | This study | PDB: 6Z1G |
| Cryo-EM density map: <i>NosRca</i> :Rubisco complex | This study | EMDB: EMD-11029 |

(Continued on next page)

Continued

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Oligonucleotides | | |
| Oligo sequences used in this study | Table S4 | N/A |
| Recombinant DNA | | |
| pET11a-NosGroSEL-H ₆ ubi-NosLS | This study | N/A |
| pET11a-NosGroSEL-H ₆ ubi-NosΔN12-S | This study | N/A |
| pHUE-NosRca | This study | N/A |
| pHUE-NosRca mutant (V91E/R92G/G93N/L244D/N248A/L250A) | This study | N/A |
| pHUE-NosRcaΔC | This study | N/A |
| pHUE-NosRcaΔC (V91E/R92G/G93N) | This study | N/A |
| pHUE-NosRcaΔC (P140G/Y143A/D144A) | This study | N/A |
| pHUE-NosRcaΔC (L244D/N248A/L250A) | This study | N/A |
| pHUE-NosRca-SSUL | This study | N/A |
| pET11a-AtCpn60αβ-AtCpn20-NtRbcL | (Lin et al., 2019) | N/A |
| pCDF-NtXSR ₁ R ₂ B ₂ | (Lin et al., 2019) | N/A |
| pET11a-AtCpn60αβ-AtCpn20-H ₆ ubi-NtRbcLΔN9 | This study | N/A |
| pHUE-NtRca(R294V) | (Stotz et al., 2011) | N/A |
| pBAD33-EcGroSEL | (Kerner et al., 2005) | N/A |
| Software and Algorithms | | |
| Aimless | (Evans and Murshudov, 2013) | https://www.ccp4.ac.uk/html/aimless.html |
| Auto-Rickshaw | (Panjikar et al., 2005) | http://www.embl-hamburg.de/Auto-Rickshaw/ |
| Buccaneer | (Cowtan, 2006) | https://www.ccp4.ac.uk/html/cbuccaneer.html |
| CCP4i | (Collaborative Computational Project, Number 4, 1994) | https://www.ccp4.ac.uk/ccp4i_main.php |
| Clustal Omega | (Sievers et al., 2011) | https://www.ebi.ac.uk/Tools/msa/clustalo/ |
| Coot | (Emsley and Cowtan, 2004) | https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/ |
| CTFFIND-4.1 | (Rohou and Grigorieff, 2015) | https://grigoriefflab.umassmed.edu/ctffind4 |
| DSSP | (Kabsch and Sander, 1983) | https://swift.cmbi.umcn.nl/gv/dssp/ |
| ESPrpt | (Gouet et al., 1999) | http://esprpt.ibcp.fr/ESPrpt/ESPrpt/ |
| Focus | (Biyani et al., 2017) | https://github.com/C-CINA/focus |
| Fiji | (Schindelin et al., 2012) | https://imagej.net/Welcome |
| Gautomatch | https://www.mrc-lmb.cam.ac.uk/kzhang/ | https://www.mrc-lmb.cam.ac.uk/kzhang/ |
| HKL2MAP | (Pape and Schneider, 2004) | http://webapps.embl-hamburg.de/hkl2map/ |
| Lsqman | (Kleywegt and Jones, 1994) | http://xray.bmc.uu.se/usf/lsqman_man.html |
| MolProbity | (Chen et al., 2010) | https://www.phenix-online.org/documentation/reference/molprobity_tool.html |
| Molrep | (Vagin and Isupov, 2001) | https://www.ccp4.ac.uk/html/molrep.html |
| MotionCor2 | (Zheng et al., 2017) | https://emcore.ucsf.edu/ucsf-software |
| Origin | OriginLab | https://www.originlab.com/ |
| Phenix | (Adams et al., 2010) | https://www.phenix-online.org/ |
| PISA | (Krissinel and Henrick, 2007) | http://www.ccp4.ac.uk/pisa/ |
| Pointless | (Evans, 2006) | N/A |

(Continued on next page)

Continued

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|-------------------------------------|---|
| PyMol | Schrödinger, LLC | http://www.pymol.org/ |
| R | R Core Team | https://www.R-project.org/ |
| Refmac5 | (Murshudov et al., 2011) | http://www.ccp4.ac.uk/html/refmac5.html |
| Relion | (Scheres, 2012) | https://www3.mrc-lmb.cam.ac.uk/relion/index.php?title=Main_Page |
| Resolve | (Terwilliger, 2000) | https://solve.lanl.gov/ |
| Scala | (Evans, 1997) | https://www.ccp4.ac.uk/html/scala.html |
| SerialEM | (Mastronarde, 2005) | https://bio3d.colorado.edu/SerialEM/ |
| Sharp | (de La Fortelle and Bricogne, 1997) | https://www.globalphasing.com/sharp/ |
| Shelx | (Sheldrick, 2010) | http://shelx.uni-ac.gwdg.de/ |
| StackReg | (Thévenaz et al., 1998) | https://github.com/fiji-BIG/StackReg/blob/master/src/main/java/StackReg_java |
| SWISS-MODEL | (Waterhouse et al., 2018) | https://swissmodel.expasy.org/ |
| tidyverse | (Wickham et al., 2019) | https://www.tidyverse.org/ |
| Truncate | (French and Wilson, 1978) | https://www.ccp4.ac.uk/html/truncate.html |
| UCSF-Chimera | (Pettersen et al., 2004) | https://www.cgl.ucsf.edu/chimera/ |
| XDS | (Kabsch, 2010) | http://xds.mpimf-heidelberg.mpg.de/ |

RESOURCE AVAILABILITY

Lead Contact

M. Hayer-Hartl (mhartl@biochem.mpg.de).

Materials Availability

Requests for resources and reagents should be directed to M. Hayer-Hartl (mhartl@biochem.mpg.de).

Data and Code Availability

The models and the electron density maps for *NosRca*ΔC:Rubisco and *NosRca*:Rubisco have been deposited to the wwPDB database under PDB/EMDB accession codes 6Z1F/EMD-11028 and 6Z1G/EMD-11029, respectively. The local electron density map for *NosRca*ΔC in the *NosRca*ΔC:Rubisco complex has been deposited under EMDB accession code EMD-11575. The crystallographic models and structure factors for *NosRca*ΔC-Gd complex, *NosRca*ΔC and *NosSSUL* have been deposited to the PDB database under accession codes 6Z1D, 6Z1E and 6HAS, respectively.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Strains

E. coli DH5α (ThermoFisher) were used for the amplification of plasmid DNA. Positive clones were selected and cultivated in LB medium at 37°C for 12 h. Both *E. coli* BL21 STAR and BL21 (DE3) (ThermoFisher) were used for recombinant protein production. Culture conditions for protein expression are provided in ‘Method Details’.

The cyanobacterium *Nostoc* sp. PCC 7120 (Culture Collection of Algae) was used to obtain genomic DNA for the cloning of *Nostoc* genes encoding *groES-EL*, *rbcL-X-S* and *rca*. *Nostoc* sp. PCC 7120 was cultured in BG-11 medium at 30°C and 50 rpm in continuous light.

Plasmids and proteins

The oligos used for amplification and generation of plasmid are listed in Table S4.

Genomic DNA

Nostoc sp. PCC 7120 cells cultured to high density (1 ml) were pelleted by centrifugation (10,000 x g for 10 min). The cell pellet was resuspended in 100 mL buffer (50 mM Tris-HCl pH 8.0/50 mM NaCl) and cells lysed by 5 cycles of heating (3 min at 95°C) and snap freezing in liquid nitrogen. The lysate was centrifuged (20,000 x g for 10 min) and 1 μL of supernatant was used as template in PCR reactions. The *groES-EL* operon was amplified using oligos 25/26, the *rbcL-X-S* operon using oligos 33/34 and the *rca* gene using oligos 49/50 (Table S4). The final plasmids were assembled by PCR and Gibson assembly as described below.

Plasmids

pET11a-*NosGroSEL* was generated by amplifying the *groES-groEL* operon by PCR from genomic DNA of *Nostoc* sp. PCC 7120 (*NosGroSEL*) and cloning into the pET11a vector via Gibson assembly. The Rubisco genes from *Nostoc* sp. PCC 7120 (*NosrbcL* and *NosrbcS*) were also amplified from genomic DNA by PCR and inserted into the pET28b vector. The His₆-ubiquitin tag (H₆ubi) was amplified from the pHUE plasmid (Baker et al., 2005; Catanzariti et al., 2004) and inserted as a N-terminal fusion to *NosrbcLS* to obtain pET28b-H₆ubi-*NosLS*. Finally, H₆ubi-*NosLS* was inserted C-terminally into pET11a-*NosGroSEL* to obtain pET11a-*NosGroSEL*-H₆ubi-*NosLS*. The N-terminal deletion mutant pET11a-*NosGroSEL*-H₆ubi-*Nos*ΔN12-S was generated by deleting the first 12 residues (MSYAQTKTQTKS) at the N terminus of *NosRbcL* by PCR. The plasmid pET11a-*AtCpn60αβ-AtCpn20-NtRbcL*, a kind gift from the group of M.R. Hanson (Lin et al., 2019), was modified to generate pET11a-*AtCpn60αβ-AtCpn20-H6ubi-NtRbcL*ΔN9 by deletion of the first nine N-terminal residues (PQTETKASV) of *NtRbcL* (*NtRbcL*ΔN9), and subsequently introducing the His₆-ubiquitin tag (H₆ubi) from the pHUE plasmid as a N-terminal fusion to *NtRbcL*ΔN9 (H₆ubi-*NtRbcL*ΔN9).

pHUE-*NosRca* was generated by amplification of the *rca* gene from genomic DNA (*Nostoc* sp. PCC 7120) and subsequent cloning of residues 2-414 into the pHUE vector by using the *Sac*II and *Eco*RI restriction sites. To generate pHUE-*NosRca*ΔC, residues 292 and 293 in pHUE-*NosRca* were changed to two consecutive stop codons by site-directed mutagenesis. All mutations in pHUE-*NosRca* and pHUE-*NosRca*ΔC were introduced by QuikChange mutagenesis (Agilent) resulting in the following constructs: pHUE-*NosRca* mutant (V91E/R92G/G93N/L244D/N248A/L250A); pHUE -*NosRca*ΔC (V91E/R92G/G93N); pHUE -*NosRca*ΔC (P140G/Y143A/D144A); pHUE -*NosRca*ΔC (L244D/N248A/L250A).

pHUE-*NosRca*-SSUL was generated by cloning residues 325-414 (SSUL domain) of *NosRca* into the pHUE plasmid by PCR, and subsequent Gibson assembly (NEB).

Proteins

*Nt*Rubisco from *Nicotiana tabacum* leaves and *NtRca*(R294V) recombinantly expressed in *E. coli*, were purified as previously described (Servaites, 1985) (Stotz et al., 2011).

METHOD DETAILS

Protein expression and purification

Protein concentrations were determined spectrophotometrically at 280 nm.

*NosRca*ΔC and *NosSSUL* for crystallography

E. coli BL21, harboring arabinose-inducible pBAD33-*EcGroSEL* and pHUE-*NosRca* plasmids, was used for the expression and purification of *NosRca*ΔC (residues 2-291) as a H₆-ubiquitin (H₆Ubi) fusion protein. Cells were grown in Luria-Bertani (LB) media at 37°C/180 rpm until OD₆₀₀ 0.3-0.4. The chaperonin GroES/EL was induced by addition of 1% (w/v) L-arabinose. Two h after induction, the temperature was reduced to 18°C and expression of H₆Ubi-*NosRca*ΔC was induced by addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 20 h. Cells were resuspended in ethanolamine (ETA) lysis buffer (50 mM ETA pH 8.0/300 mM NaCl/10 mM imidazole/5% (v/v) glycerol) containing 1 mM PMSF, 10 mM 2-mercaptoethanol (2-ME), 1 g L⁻¹ lysozyme and 5 U mL⁻¹ benzonase, and lysed by sonication (15 × 15 s bursts with 75 s intermissions on ice). The supernatant obtained after high speed centrifugation (20 000 × g for 30 min at 4°C) was loaded on a gravity-flow nickel-nitrilotriacetic acid (Ni-NTA) metal affinity column (QIAGEN), washed with 10 column volumes (CV) of ETA lysis buffer pH 9.2 containing 25 mM imidazole, and the protein eluted with ETA lysis buffer pH 9.2/200 mM imidazole. Fractions containing the protein were pooled and the H₆Ubi moiety was cleaved by Usp2 (Baker et al., 2005; Catanzariti et al., 2004) at 4°C overnight. After dialysis against 50 mM ETA pH 9.2/10 mM NaCl, *NosRca*ΔC was loaded on a Mono Q HR 16/10 column (GE). The protein was eluted with a 10 CV gradient (0.01 – 0.5 M NaCl), concentrated and applied to a size-exclusion chromatography column (Superdex 200 10/300 GL; GE) equilibrated in 20 mM Tris-HCl pH 8.0/50 mM NaCl/5 mM MgCl₂. Fractions containing the *NosRca*ΔC were concentrated by ultrafiltration using Vivaspin MWCO 10000 (GE) and glycerol added to 5% final prior to aliquoting and flash freezing in liquid N₂. For all other studies the purification strategy was modified as stated below.

NosSSUL (residues 325-414) was recombinantly expressed in *E. coli* from the pHUE-*NosRca*-SSUL plasmid as a H₆Ubi fusion protein. Cells were grown in LB media at 30°C/180 rpm until OD₆₀₀ 0.3-0.4, and then *NosSSUL* expression induced by addition of 0.2 mM IPTG and the cells shifted to 22°C/120 rpm for 18 h. Cells were harvested and resuspended in ETA lysis buffer (50 mM ETA pH 9.2/300 mM NaCl/10 mM imidazole/5% (v/v) glycerol) containing 1 mM PMSF, 10 mM 2-mercaptoethanol (2-ME), 1 g L⁻¹ lysozyme and 5 U mL⁻¹ benzonase, and lysed by sonication (15 × 15 s bursts with 75 s intermissions on ice). Purification of *NosSSUL* was carried out essentially as above for *NosRca*ΔC. Subsequent to dialysis after the first Ni-NTA column, *NosSSUL* was applied to a second Ni-NTA resin to remove Usp2, H₆Ubi and any uncleaved protein. The flow through was concentrated and applied to a size-exclusion chromatography column (HiLoad 16/60 Superdex 75; GE) equilibrated in 50 mM ETA pH 9.2/300 mM NaCl/1 mM DTT/5% glycerol. Fractions containing the *NosSSUL* were concentrated by ultrafiltration using Vivaspin MWCO 3000 (GE), aliquoted and flash frozen in liquid N₂.

NosRca, *NosRca*ΔC and mutants

NosRca, *NosRca*ΔC and mutants were expressed and purified from *E. coli* BL21 STAR cells harboring arabinose-inducible pBAD33-*EcGroSEL* and the respective pHUE plasmid. Briefly, cells were grown in 2xYT media containing 10 mM KCl (Larimer and Soper, 1993) at 37°C/180 rpm until OD₆₀₀ 0.3-0.4. GroSEL was induced by addition of 0.4% (w/v) L-arabinose. One h after chaperonin induction, the temperature was reduced to 18°C. When the cells had equilibrated to 18°C (~1 h), induction of protein was started by

addition of 0.2 mM IPTG and allowed to proceed for 18 h/120 rpm. Cells were harvested and incubated in buffer A (50 mM MMT pH 8.0/300 mM KCl/10 mM MgCl₂/5% glycerol) containing 1 g L⁻¹ lysozyme/2.5 U mL⁻¹/SmdNase//complete protease inhibitor cocktail (Roche) for 1 h prior to lysis using EmulsiFlex C5 (Avestin, Inc). Note MMT is a composite buffer consisting of DL-malic acid, MES monohydrate and Tris base in the molar ratio 1:2:2. After high speed centrifugation (40 000 x g/40 min/4°C) the supernatant was loaded on to a gravity TALON metal affinity column (Takara), equilibrated and washed with 10 CV buffer A/20 mM imidazole. The bound protein was eluted with buffer A pH 8.4/200 mM imidazole, and diluted 3-fold in buffer A containing 5 mM 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS)/10% glycerol/5 mM 2-ME to a final protein concentration < 0.5 g L⁻¹. The H₆Ubi moiety was cleaved by Usp2 overnight at 10°C. The cleaved protein was buffer exchanged on a HiPrep 26/10 desalting column (GE) to 50 mM MMT pH 8.4/10 mM KCl/5% glycerol, and subsequently loaded on a MonoQ column (GE). The protein was eluted with a 10 CV gradient (0.01 – 0.5 M KCl), concentrated to ~5 mL and applied onto a size-exclusion chromatography column (HiLoad 16/60 Superdex 200; GE) equilibrated in buffer B (50 mM MMT pH 8.4/100 mM KCl/10 mM MgCl₂/5% glycerol). The protein containing fractions were concentrated by ultrafiltration using Vivaspin MWCO 3000 (GE), aliquoted and flash frozen in liquid N₂. To generate reduced NosRca (reNosRca), 5 mM DTT was added to all buffers after Usp2 cleavage. The oxidized NosRca (oxNosRca) purified in the absence of DTT was allowed to further air oxidize on ice for 48 h after the final column prior to concentrating, aliquoting and flash freezing.

NosRubisco and mutants

NosRubisco was expressed in *E. coli* BL21 STAR cells harboring the IPTG-inducible *Nostoc* chaperonin on pET11a-NosGroSEL-H₆ubi-NosLS plasmid, in which the cleavable H₆Ubi motif is attached at the N terminus of NosRbcL. Cells were grown in 2 x YT/10 mM KCl media (Larimer and Soper, 1993) at 37°C/180 rpm until OD₆₀₀ 0.6 – 0.8. NosGroES/EL and H₆ubi-NosRbcLS were induced by addition of 0.5 mM IPTG and cells shifted to 22°C at 120 rpm for 22 h. Cells were lysed, cell debris removed and the supernatant loaded onto a TALON resin column as described above. After 10 CV washes in buffer A/20 mM imidazole the H₆Ubi-tagged Rubisco complex was eluted with buffer A pH 8.4/200 mM imidazole. Usp2-mediated digestion was performed overnight in presence of 5 mM 2-ME at 10°C. After removal of imidazole on a HiPrep 26/10 desalting column (GE) equilibrated in buffer A, the buffer-exchanged protein eluate was applied to a TALON resin column for removal of Usp2, the cleaved H₆Ubi moiety and any uncleaved protein. The flow through was concentrated to ~5 mL and applied onto a size-exclusion chromatography column as above, equilibrated in buffer B containing 30 mM NaHCO₃. The NosRubisco containing fractions were concentrated by ultrafiltration using Vivaspin MWCO 30000 (GE), aliquoted and flash frozen in liquid N₂.

NtRubiscoΔN

NtRubiscoΔN, with 9 residues deleted at the N terminus of the RbcL subunit, was expressed in *E. coli* BL21 STAR cells harboring the IPTG-inducible plasmids pET11a-AtCpn60αβ-AtCpn20-H₆ubi-ΔN9NtRbcL and pCDF-NtXSR₁R₂B₂ (kind gift from MR Hanson). Cells were grown in 0.1 L ZYP-5052 auto-induction media (Lin et al., 2019; Studier, 2005) at 37°C/160 rpm for 6 h and then transferred to 23°C/120 rpm for 22 h. Harvested cells were re-suspended and incubated in buffer A containing 1 g L⁻¹ lysozyme, 2.5 U mL⁻¹ SmdNase, complete protease inhibitor cocktail (Roche) for 1 h prior to lysis by sonication. After high speed centrifugation (40 000 x g/40 min/4°C) the supernatant was loaded on a HiTrap (GE) TALON Crude (Takara) metal affinity column, washed with 10 CV of buffer A prior to elution by a 10 CV linear imidazole gradient 0 to 200 mM for separation of H₆Ubi-tagged Rubisco from AtCpn60αβ. Fractions containing only H₆Ubi-tagged ΔN9NtRubisco were selected by immunoblotting against NtRbcL and AtCpn60αβ. The pooled fractions were subjected to digestion by Usp2 (8 h/10°C) in the presence of 5 mM 2-ME. The reaction was then buffer exchanged to 50 mM MMT pH 8.4/100 mM KCl/5% glycerol on a HiPrep 26/10 desalting column (GE), and applied to a gravity TALON (Takara) metal affinity column for removal of Usp2, the cleaved H₆Ubi moiety and any uncleaved protein. The flow through was concentrated by ultrafiltration using Vivaspin MWCO 30000 (GE), aliquoted and flash frozen in liquid N₂.

CABP synthesis

2-Carboxyarabinitol-1,5-diphosphate (CABP) was synthesized according to (Pierce et al., 1980). In brief, 100 μmol RuBP was incubated with a 2-fold molar excess of KCN in 5 mL 0.1 M Tris-acetate pH 8.3 for 48 h at 25°C. The racemic mixture of 2-carboxyarabinitol-1,5-diphosphate (CRBP) and CABP was treated with the cation exchange resin AG50W-X8 (H⁺), filtered and freeze-dried. To separate the enantiomers, the lactonized products were dissolved in 3 mM HCl and applied to a 120 mL AG1X8 (Cl⁻) column equilibrated in 3 mM HCL, eluted over a 4 L gradient from 0 – 0.4 M LiCl, and then the 50 mL fractions were assayed for total phosphate (Chifflet et al., 1988). CABP containing fractions were pooled and reduced to a volume of 50 mL in a rotary evaporator at 30°C. Addition of 3-fold molar excess of barium acetate precipitated the CABP as barium salt (1 h at –20°C). The precipitate was collected by centrifugation (5000 x g for 20 min at 4°C) and redissolved by addition of acid-washed AG50W-X8 resin. After filtration, the purified CABP was freeze-dried and redissolved in 50 mM Bicine-NaOH pH 9.3. Complete saponification was ensured by incubation on ice for 24 h, before aliquoting and flash freezing in liquid N₂.

Enzymatic assays

All assays were performed at 25°C.

Rubisco reactivation

CO₂ fixation by Rubisco and Rubisco activase activity were measured as previously described (Barta et al., 2011; Esau et al., 1996) in buffer 50 mM MMT pH 8.4/50 mM KCl/30 mM NaH^{14/12}CO₃ (14 Bq nmol⁻¹)/10 mM MgCl₂/3 mM phosphocreatine/50 U mL⁻¹

phosphocreatine kinase containing RuBP (0.4 or 1 mM) and ATP (3 mM) as indicated in the figure legends. The carbamylated Rubisco (ECM) was obtained by preincubating Rubisco in 100 mM MMT pH 8.4/60 mM NaHCO₃/20 mM MgCl₂ for 10 min. The uncarbamylated Rubisco (E) was obtained by buffer exchange of the purified Rubisco into 100 mM MMT pH 8.4/50 mM KCl/4 mM EDTA. The inhibited *Nos*Rubisco (E.XuBP, E.RuBP and ECM.CABP) and inhibited *Nt*Rubisco (E.RuBP) were formed by addition of the respective inhibitory sugars and incubation for 30–60 min. Time course experiments were initiated by addition of ECM, ECM.CABP, E, E.RuBP or E.XuBP (0.25 μM active sites) to the reaction assay including RuBP (0.4 mM unless otherwise indicated), Rca and ATP (3 mM) as indicated. The data were fitted in SigmaPlot using the equation Sigmoid, 4 Parameter $f = y_0 + a / (1 + \exp(-(x-x_0)/b))$.

Dose dependent reactivation of *Nos*Rubisco (ECM.CABP) was measured in presence of ATP (3 mM) and at the indicated *Nos*RcaΔC concentrations (0.125, 0.25, 0.5, 1.25 μM hexamer). The single time-point experiments for the amount of CO₂ fixed was analyzed at 8 min after initiation of the reaction and conducted in the presence of 1 mM RuBP to ensure steady state kinetics for the ECM control.

To estimate specific activity and account for background counts, corresponding samples were measured in 5 or 6 replicates per assay.

ATPase assay

ATP hydrolysis was enzymatically coupled to NADH oxidation and measured spectrophotometrically at 25°C (Barta et al., 2011; Mueller-Cajar et al., 2011). The assay was started by addition of *Nos*Rca, *Nos*RcaΔC, or mutants (*Nos*Rca V91E/R92G/G93N/ L244D/ N248A/L250A, *Nos*RcaΔC P140G/Y143A/D144A, *Nos*RcaΔC V91E/R92G/G93N, *Nos*RcaΔC L244D/N248A/L250A) at 0.5 μM (hexamer) in assay buffer (100 mM MMT pH 8.4/20 mM KCl/10 mM MgCl₂/5 mM DTT/2 mM phosphoenolpyruvate/0.3 mM NADH/2 mM ATP/~30 U mL⁻¹ pyruvate kinase/45 U mL⁻¹ lactic dehydrogenase). ECM and ECM.CABP (0.25 μM hexadecamer) were prepared as described above and added as indicated. To suppress turbidity due to condensate formation of activated or inhibited Rubisco (ECM or ECM.CABP, respectively) in the presence of *Nos*Rca, the salt concentration was increased to 200 mM KCl in the assay buffer.

Turbidity assay

Measurements were performed at 25°C in buffer C (50 mM MMT pH 8.4/50 mM KCl/10 mM MgCl₂) for ox*Nos*Rca and in the presence of additional 5 mM DTT for re*Nos*Rca /*Nos*RcaΔC /re*Nos*Rca-IF_{mut}. Reactions (100 μL) containing Rubisco (0.25 mM) and different concentrations of *Nos*Rca (re*Nos*Rca or ox*Nos*Rca) or *Nos*RcaΔC (0.5 μM) or re*Nos*Rca-IF_{mut} (0.5 μM) in the absence or presence of 2 mM ATPγS as indicated in the figure legends were mixed rapidly by vortexing, and absorbance at 340 nm was monitored as a function of time on a Jasco V-560 spectrophotometer.

Liquid-liquid phase separation

For analysis of LLPS, Rubisco holoenzyme was labeled at the N terminus with the fluorophore Alexa Fluor 532 NHS ester (ThermoFisher) according to manufacturer's instructions (~2 dye molecules bound per Rubisco holoenzyme). re*Nos*Rca, ox*Nos*Rca, *Nos*RcaΔC and *Nos*Rca-IF_{mut} was labeled at the N terminus with the fluorophore Alexa Fluor 405 NHS ester (ThermoFisher) (~2.2, 2.4, 1.7 and 3 dye molecules bound per Rca hexamer, respectively). Labeled protein was mixed with unlabeled protein, at a ratio of 1:10. Reactions (20 μL) in buffer C containing 5 mM DTT in the presence or absence of 2 mM ATPγS with re*Nos*Rca, ox*Nos*Rca, *Nos*RcaΔC or *Nos*Rca-IF_{mut} (0.5 μM) and Rubisco (0.25 μM) were incubated for 5 min at 25°C and then transferred to an uncoated chambered coverslip (μ-Slide angiogenesis; Ibidi) for another 5 min before analysis. Images were illuminated with a Lumencor SPECTRA X Light Engine at 398 nm and 558 nm for fluorescence imaging. Images were recorded by focusing on the bottom of the plate using a Leica Thunder Widefield 2 microscope with Leica DFC9000 GTC camera and a HC PL APO 63x/1.47 oil objective.

Size-exclusion chromatography coupled to multi-angle static light scattering (SEC-MALS)

Purified proteins at 2 mg mL⁻¹ was analyzed using static and dynamic light scattering by auto-injection of the sample onto a SEC column (5 mm, 4.6x300 mm column, Wyatt Technology, product # WTC-030N5) at a flow rate of 0.2 mL min⁻¹ in buffer 50 mM MMT pH 8.4/100 mM KCl/10 mM MgCl₂ at 25°C in the presence or absence of nucleotide (1 mM). The column was in line with the following detectors: a variable UV absorbance detector set at 280 nm (Agilent 1100 series), the DAWN EOS MALS detector (Wyatt Technology, 690 nm laser) and the Optilab rEXTM refractive index detector (Wyatt Technology, 690 nm laser) (Wyatt, 1993). Molecular masses were calculated using the ASTRA software (Wyatt Technology) with the dn/dc value set to 0.185 mL g⁻¹. Bovine serum albumin (Thermo) was used as the calibration standard.

Electron microscopy and reconstruction

Cryo-EM

All cryo-grids were prepared with a Vitrobot Mark 4 (FEI). A volume of 3 μL of the sample was applied to a glow-discharged grid (Quantifoil R2/1 300 mesh) at 25°C and 90% humidity, then semi-automatically blotted and plunge-frozen into liquid ethane.

To capture the interaction between *Nos*RcaΔC and CABP-inhibited Rubisco (*Nos*RcaΔC:Rubisco), *Nos*Rubisco (1 μM) was first carbamylated by incubation in buffer C containing 10 mM NaHCO₃ for 10 min at 25°C, then inhibited with 8 μM CABP at 25°C for 1 h. The *Nos*RcaΔC:Rubisco (*Nos*RcaΔC:ECM.CABP) complex was formed following a published nucleotide-substitution strategy (Dong et al., 2019). Specifically, ECM.CABP (0.5 μM) was incubated with *Nos*RcaΔC (10 μM) at 25°C in the presence of ATP (2 mM) for 10 s, followed by the addition ATPγS (2 mM), and incubated at 25°C for another 10 min before preparing the cryo-grids

as stated above. The cryo-grids were initially screened on a Talos Arctica (FEI) transmission electron microscope (TEM). Selected grids were transferred to a Titan Krios 300 kV TEM (FEI) equipped with GIF Quantum Energy Filters (Gatan), and a K3 direct detector (Gatan). 9,042 movies were automatically collected by SerialEM (Mastrorarde, 2005) using a pixel size of 0.8512 Å. The total exposure time of 2.8 s was divided into 31 frames with an accumulated dose of 60 electrons per Å² and a defocus range of $-0.7 \mu\text{m}$ to $-2.5 \mu\text{m}$.

For Preparing the NosRca:Rubisco complex, NosRubisco (1.25 μM) was mixed with NosRca (15 μM) in buffer C containing 5 mM DTT and cryo-grids prepared as above. The cryo-grids were screened on a Glacios transmission electron microscope (Thermo Scientific), equipped with K2 summit direct electron detector (Gatan), operated at 200 keV. Selected grid on stage was used for data collection directly with K2 summit. Exposure times of 12 s were divided into 40 frames with an accumulated dose of 47 electrons per Å². 1,570 movies were automatically collected by SerialEM (Mastrorarde, 2005) with a pixel size of 1.885 Å and a defocus range of $-1 \mu\text{m}$ to $-4.5 \mu\text{m}$.

Image processing

For the NosRcaΔC:Rubisco (NosRcaΔC:ECM·CABP) dataset, on-the-fly processing during data collection was performed with MotionCorr2 and CTFIND-4.1, as implemented in the Focus software (Biyani et al., 2017). Only micrographs with good particle quality, with an estimated maximum resolution below 5 Å, were kept for further data processing with RELION 3.0. A total of 519,151 particles were auto-picked by Gautomatch (<https://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch>) and extracted at a pixel size of 3.4048 Å (four-fold binned). The first round of 2D-classification was used to exclude ice contaminations and classes with no structural features. The selected particles were next refined according to a Rubisco reference converted from the dataset of PDB: 1RBL (Newman et al., 1993), re-centered, and then subjected to a second round of 2D-classification. Free NosRcaΔC classes were excluded, which resulted in the selection of 106,831 particles (Figure S3). 3D-classification with the Rubisco reference showed only one class of the NosRcaΔC hexamer in complex with Rubisco, containing 30,607 particles, with a resolution of 8.28 Å. These particles were re-extracted at full resolution (0.8512 Å pixel size) and subjected to “polishing” with RELION to generate “shiny” particles. To deal with multiple NosRcaΔC hexamers bound per Rubisco, we followed a previously published symmetry-expansion procedure (Wang et al., 2019). In detail, particles were first aligned with *D*₄ symmetry. Then *D*₄ symmetry was released and yielded 8-fold particles by the symmetry-expanding command, `relion_particle_symmetry_expand`. A focused 3D-classification in NosRcaΔC with symmetry-released particles revealed one class with NosRcaΔC hexamer occupancy, comprising of 21,149 particles (Figure S3). These particles generated the final map of the NosRcaΔC:ECM·CABP complex at 2.86 Å resolution, determined by gold-standard Fourier shell correlation (FSC) with a cutoff at 0.143. Particle subtraction of the Rubisco signal improved the alignment accuracy of the NosRcaΔC hexamer, and generated a local NosRcaΔC map at 3.29 Å resolution.

The raw movies of the NosRca:Rubisco dataset were first processed with MotionCorr2 (Zheng et al., 2017) with dose-weighting. CTFIND-4.1 (Rohou and Grigorieff, 2015) estimated the CTF parameters for each micrograph. 298,336 Particles were picked by Gautomatch (<https://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch>). Two rounds of 2D-classification excluded ice contaminations and classes with no structural features, and resulted in 45,859 clean particles (Figure S7). These particles were used to generate a reference for 3D-classification in RELION 3.0. (Scheres, 2012) with the 3D initial model module. 3D-classification revealed one class with detailed Rubisco features with a resolution of 8.28 Å (27,527 particles). To deal with multiple SSUL domains bound per Rubisco, we followed the same symmetry-expansion procedure previously reported (Wang et al., 2019). Particles were first aligned with *D*₄ symmetry. Each asymmetric unit L₂S₂SSUL was processed as an individual particle, which is achieved by the symmetry-expanding command, `relion_particle_symmetry_expand`, and particle subtraction. A focused classification with a SSUL mask resulted in one class of particles with detailed SSUL feature. 32,128 particles from this class were selected and subjected to final local refinement. Post-processing improved the map resolution to 8.2 Å.

Model building

NosRcaΔC:ECM·CABP – NosRcaΔC model building was initiated by rigid-body fitting the NosRcaΔC subdomains from the crystal structure into the cryo-EM density, followed by manual editing using Coot (Emsley and Cowtan, 2004). This model was refined in reciprocal space with REFMAC5 (Murshudov et al., 2011). The ECM.CABP model was generated with SWISS-MODEL (Waterhouse et al., 2018) based on the coordinates of Rubisco from *Chlamydomonas reinhardtii* (PDB: 1UZH) (Karkehabadi et al., 2005). This model was placed into the cryo-EM density using Chimera (Pettersen et al., 2004), followed by manual editing using Coot. Residues with disordered side-chains were truncated at C-β. This model was refined in reciprocal space with REFMAC5 (Murshudov et al., 2011), using non-crystallographic symmetry restraints.

NosRca:Rubisco – First, the crystal structure of thiol-reduced NosSSUL and the NosRubisco model (solved from the NosRcaΔC:ECM-CABP complex) were placed into the density using Chimera, followed by manual editing using Coot. The resulting model was refined in reciprocal space with REFMAC5, using jelly-body restraints. The used structure factors were calculated from a masked map.

The models and the electron density maps for NosRcaΔC:Rubisco and NosRca:Rubisco have been deposited to the wwPDB database under PDB/EMDB accession codes 6Z1F/EMD-11028 and 6Z1G/EMD-11029, respectively. The local electron density map for NosRcaΔC in the NosRcaΔC:Rubisco complex has been deposited under EMDB accession code EMD-11575.

Negative-stain EM

Reduced or oxidized NosRca (0.125 μM), in buffer C with or without 5 mM DTT, respectively, was incubated on a glow-discharged grid (Quantifoil Carbon support films 200 mesh) at 25°C for 1 min, then stained with 2% (w/v) uranyl acetate. Images were recorded on

Titan Halo (FEI) equipped with Falcon 3EC Direct Electron Detector (ThermoFisher) with a pixel size of 1.85 Å by SerialEM (Mastrorade, 2005). A targeted defocus of $-2\ \mu\text{m}$ was obtained by the auto-focus function in SerialEM.

Crystallization and data collection

The *NosRcaΔC* construct used for crystallization includes residues 2–291. The *NosSSUL* construct used for crystallization includes residues 325–414.

NosRcaΔC – Crystals were grown by the hanging-drop vapor diffusion method at 4°C. Drops containing 2 μL of a 1:1 mixture of 4.9 mg mL⁻¹ *NosRcaΔC* in buffer 20 mM Tris-HCl pH 8.0/50 mM NaCl/5 mM MgCl₂ and precipitant were equilibrated against 500 μL precipitant. The precipitant contained 2.2 or 2.3 M Na-acetate pH 7.0.

NosSSUL – Crystals were grown by the hanging-drop vapor diffusion method at 4°C. Drops containing 3 μL of a 1:1 mixture of 5.2 mg mL⁻¹ *NosSSUL* in buffer 50 mM ETA pH 9.2/300 mM NaCl/1 mM DTT/5% glycerol and precipitant were equilibrated against 500 μL precipitant. The precipitant contained 26% PEG-3350 and 50 mM MES-NaOH pH 6.0.

For cryo-mounting, the crystals were transferred into a cryo-solution that was precipitant containing 15% glycerol in addition and subsequently cryo-cooled by dipping into liquid nitrogen.

Crystallographic data collection, structure solution and refinement

NosRcaΔC – The diffraction data of the crystals of *NosRcaΔC* were collected by the oscillation method at beamline X06DA at the Swiss Synchrotron Light Source (SLS) in Villigen, Switzerland. The diffraction data from *NosRcaΔC* crystals were integrated with XDS and further processed with POINTLESS (Evans, 2006), SCALA (Evans, 1997) and CTRUNCATE (French and Wilson, 1978). The structure was solved at 3.4 Å from a Gadolinium (GdCl₃) derivative by Gd-multi-wavelength anomalous diffraction (MAD) using ShelxC/D/E (Sheldrick, 2010) as implemented in the Hkl2map GUI (Pape and Schneider, 2004). The seven Gd sites were refined and phases calculated with SHARP (de La Fortelle and Bricogne, 1997). The map, calculated after density modification with RESOLVE (Terwilliger, 2000), assuming a solvent content of 55%, revealed features of secondary structure elements. A preliminary model was auto-built with Buccaneer (Cowtan, 2006), and missing portions added manually using COOT (Emsley and Cowtan, 2004). REFMAC5 was used for initial model refinement (Murshudov et al., 2011). The final refinement was performed with phenix.refine (Adams et al., 2010) using translation-libration-screw (TLS) parametrization of B-factors. The native structure was solved by molecular replacement. The final models contain two copies of *NosRcaΔC* per asymmetric unit. One chain (chain B) has ADP bound. Residues 276–291 were disordered in chain A. In chain B, residues 105–115, 248–254 and 279–291 are missing. Residues facing solvent channels with disordered side-chains were modeled as alanine. The model of the native structure contains 14 ordered water molecules and exhibits reasonable stereochemistry with 96.6% of the residues in the favored regions of the Ramachandran plot according to the criteria of MolProbity (Chen et al., 2010).

NosSSUL – The native diffraction data of the crystals of *NosSSUL* were collected at the automated beamline ID30A-1 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. MAD data for a presumed Pt-derivative were collected at beamline BM30A at ESRF. The diffraction data from *NosSSUL* crystals were integrated with XDS and further processed with POINTLESS (Evans, 2006), AIMLESS (Evans and Murshudov, 2013) and CTRUNCATE (French and Wilson, 1978) as implemented in the CCP4i graphical user interface (Potterton et al., 2003). The structure of *NosSSUL* was solved by MAD using the Auto-Rickshaw platform (Panjikar et al., 2005). The anomalous scatterers were a bound Ni²⁺ atom from protein purification and presumably ordered Cl⁻ ions. The chemical environment of the sites was not compatible with PtCl₄²⁻. The asymmetric unit contained two copies of the SSUL domain. The model was edited manually using Coot (Emsley and Cowtan, 2004). REFMAC5 was used for model refinement (Murshudov et al., 2011). The model contains 203 ordered water molecules and exhibits reasonable stereochemistry with 99.5% of the residues in the favored regions of the Ramachandran plot according to the criteria of MolProbity (Chen et al., 2010).

Structure analysis

The quality of the structural models was analyzed with the program Molprobity (Chen et al., 2010). Coordinates were aligned with Lsqkab and Lsqman (Kleywegt and Jones, 1994). Molecular interfaces were analyzed with PISA (Krissinel and Henrick, 2007) and Contact, as implemented in the CCP4i graphical user interface (Potterton et al., 2003). Figures were created with Chimera (Pettersen et al., 2004), PyMol (<http://www.pymol.org/>) and ESPript (Gouet et al., 1999).

QUANTIFICATION AND STATISTICAL ANALYSIS

The oligomeric state of proteins in solution was analyzed using the program Astra version 4.90.08 (Wyatt Technology). Data in Figures 1B–1D, 3E, 5F, and S6E represent the standard deviation (SD) of at least three independent experiments ($n = 3$). Data in Figures 6C, 6D, and 6G are representative curves of three independent experiments.

Supplemental Figures

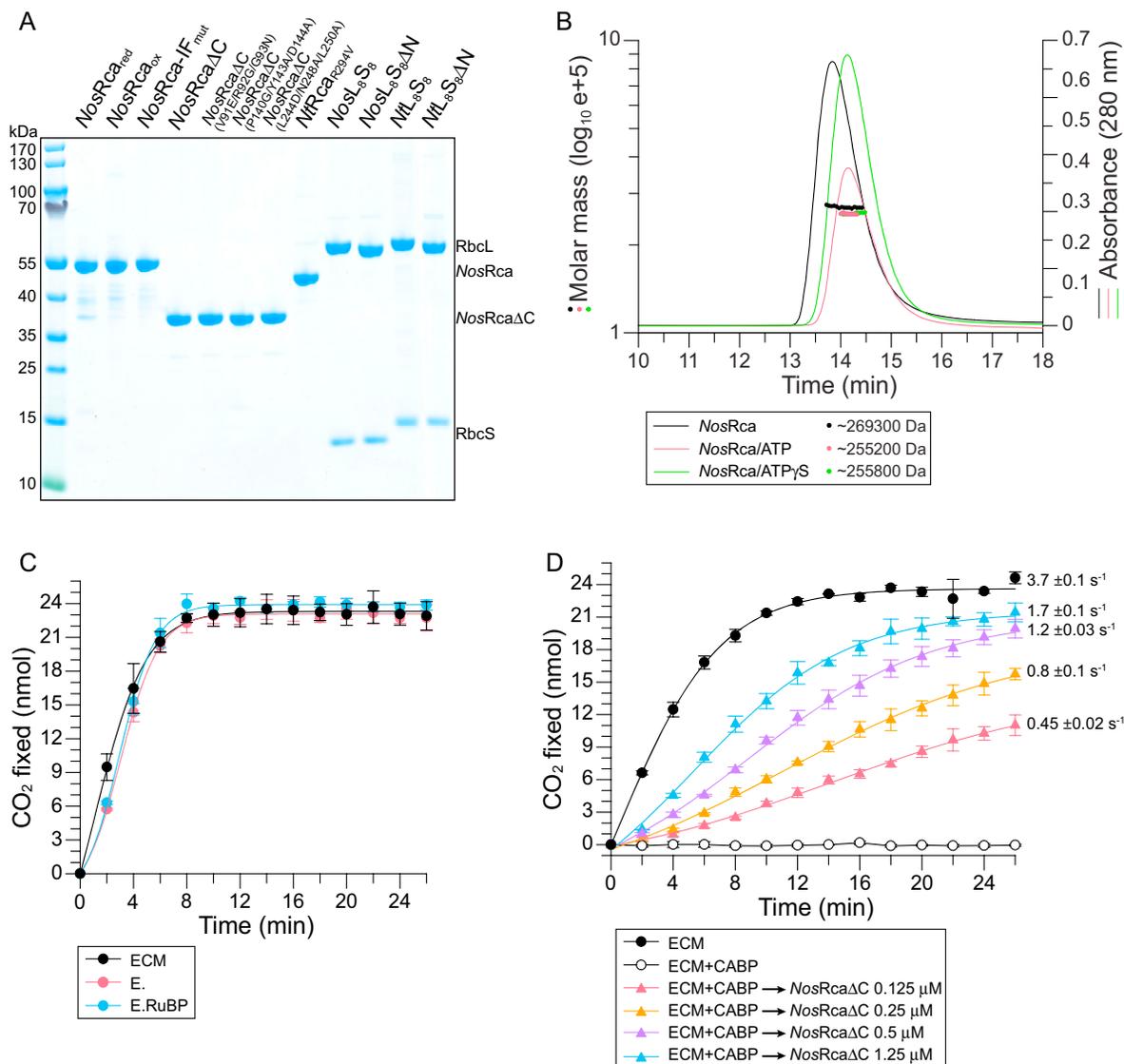


Figure S1. Rubisco Activase Function of NosRca, Related to Figures 1, 2, 5, and 6 and Table S1

(A) Purified proteins used in this study. Protein concentrations were determined spectrophotometrically. 1.5 μg protein was analyzed by SDS-PAGE.

(B) *NosRca* is a hexamer in solution. SEC-MALS analysis of *NosRca* in the absence or presence of nucleotide (1 mM ATP or ATPγS). The chromatographic absorbance traces at 280 nm wavelength are shown. The molecular mass determined for the protein peaks by static light scattering is indicated.

(C) Non-carbamylated *NosRubisco* is not inhibited by RuBP. CO₂ fixation assays were performed with carbamylated (ECM), non-carbamylated (E) and non-carbamylated Rubisco with RuBP (E.RuBP) as in Figure 1C. Error bars represent SD of at least three independent replicates.

(D) Dependence of Rubisco reactivation on *NosRcaΔC* concentration. CABP inhibited *NosRubisco* (ECM.CABP) was incubated with increasing concentrations of *NosRcaΔC* (0.125, 0.25, 0.5 and 1.25 μM hexamer) in the presence of 3 mM ATP, and CO₂ fixation measured as in Figure 1D. Approximate rates of CO₂ fixation were determined from the linear parts of the curves. Error bars represent SD of three independent replicates.

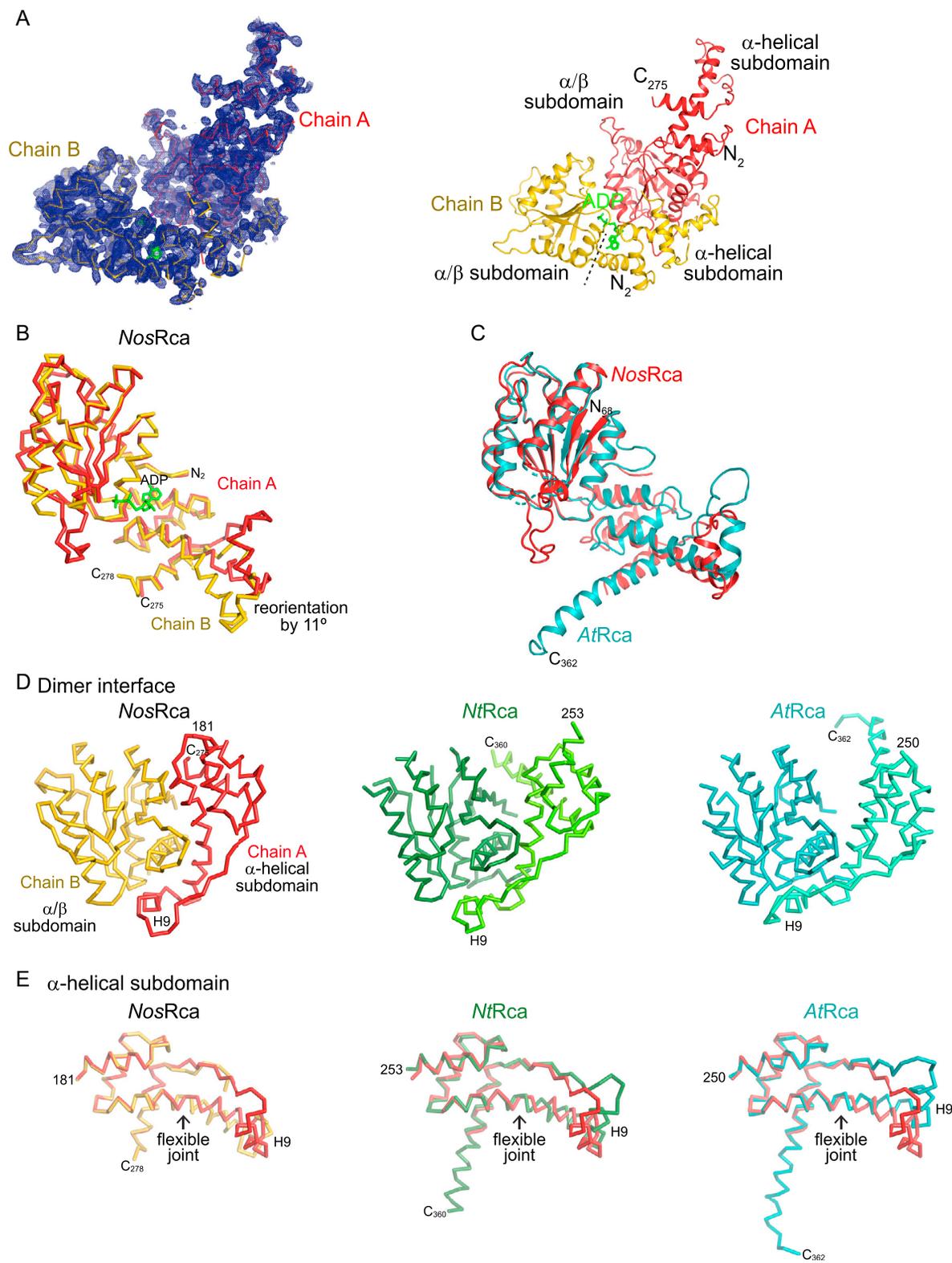


Figure S2. Crystal Structure of *NosRca*ΔC, Related to Figure 2, Table S2, and Data S1 and S2

(A) Left: Unbiased experimental electron density at 2.7 Å resolution for the Gd^{3+} complex of *NosRca*ΔC (for data collection and refinement statistics, see Table S2). The density after Gd-MAD phasing and density modification contoured at 1.5 σ is shown as a blue meshwork. Chains A and B of the asymmetric unit are shown as $C\alpha$ traces in red and yellow, respectively. ADP is represented as a wire-frame model in green. Right: Final model of the asymmetric unit in the *NosRca*ΔC (legend continued on next page)

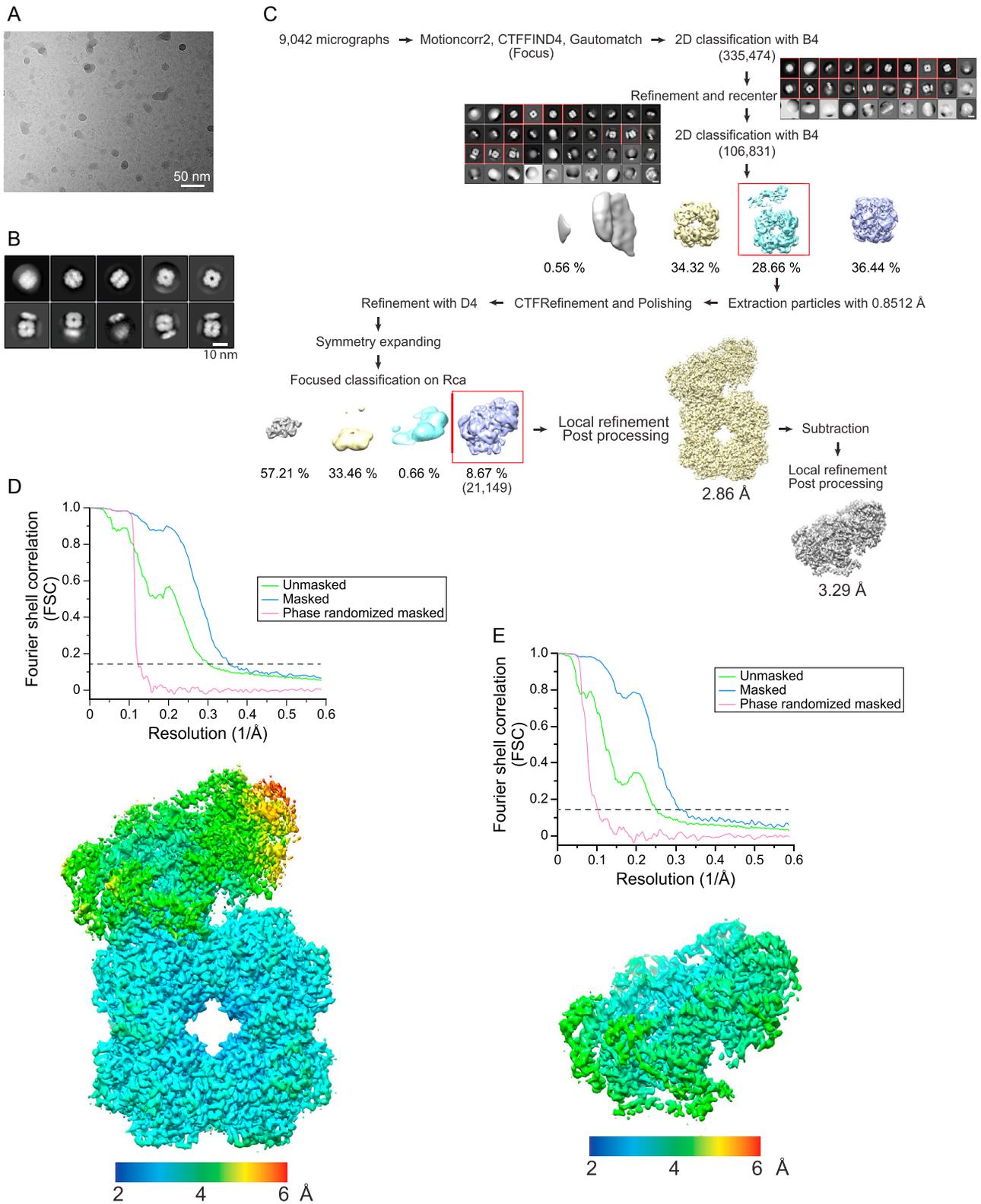
crystal lattice. The two crystallographically independent chains of *NosRca* Δ C are shown as ribbons. Subdomains and chain termini are indicated. The dashed line indicates the subdomain boundary in chain B.

(B) Superposition of chain A with chain B in the asymmetric unit of *NosRca* Δ C. Protein chains are represented as C α traces.

(C) Superposition of chain A (no nucleotide bound) of the asymmetric unit of *NosRca* Δ C with *AtRca* (PDB: 4W5W) (Hasse et al., 2015). Chain A of *NosRca* Δ C and *AtRca* are shown as ribbons in red and teal, respectively. Chain termini are indicated.

(D) A conserved subunit-subunit interaction found in *NosRca* (left), *NtRca* (middle) and *AtRca* (right). The interaction is formed between the α/β subdomain in one subunit and the C-terminal half of helix $\alpha 8$, $\alpha 9$ (H9) and the connecting linker in the α -helical subdomain of the adjacent subunit. Adjacent subunits in the *NtRca* crystal structure are shown in dark and bright green, and in teal and cyan for *AtRca*. The location of helix H9 and chain termini are indicated. Note that the orientation of helix $\alpha 9$ (H9) with respect to the four-helix bundle in the α -helical subdomain differs between the structures; the long helix $\alpha 8$ and the long $\alpha 9$ - $\alpha 10$ linker act as a stalk that can twist and bend like a flexible joint.

(E) Superposition of the α -helical subdomain of chain A in *NosRca* with chain B (left), with *NtRca* (middle) and *AtRca* (right). The rmsd values for matching C α positions are 0.24 Å (57 C α positions), 0.58 Å (42 C α positions) and 0.55 Å (47 C α positions), respectively. The locations of helix $\alpha 9$ (H9) and the flexible joint region are indicated. Chain termini are indicated.



(legend on next page)

Figure S3. Cryo-EM Single-Particle Reconstruction of *NosRcaΔC*:Rubisco Complex, Related to Figures 3 and 7, Table S3, and Video S1

(A) A representative micrograph of *NosRcaΔC*:Rubisco complexes.

(B) 2D class averages of complexes in (A).

(C) The single-particle data processing workflow for the *NosRcaΔC*:Rubisco complex. Particle numbers are in parentheses. B4, 4x4 pixel-binned image. See [STAR Methods](#) for details. Scale bars, 10 nm.

(D and E) Gold-standard FSC curves and local resolution maps of the *NosRcaΔC*:Rubisco reconstruction (D; EMD-11029) and the *NosRcaΔC* local map (E; EMD-11575). The resolutions are ~2.86 Å and ~3.29 Å, respectively, at the FSC cutoff of 0.143 for the masked and B-factor sharpened curves. The color gradient from blue to red indicates local resolution from 2.0 to 6.0 Å.

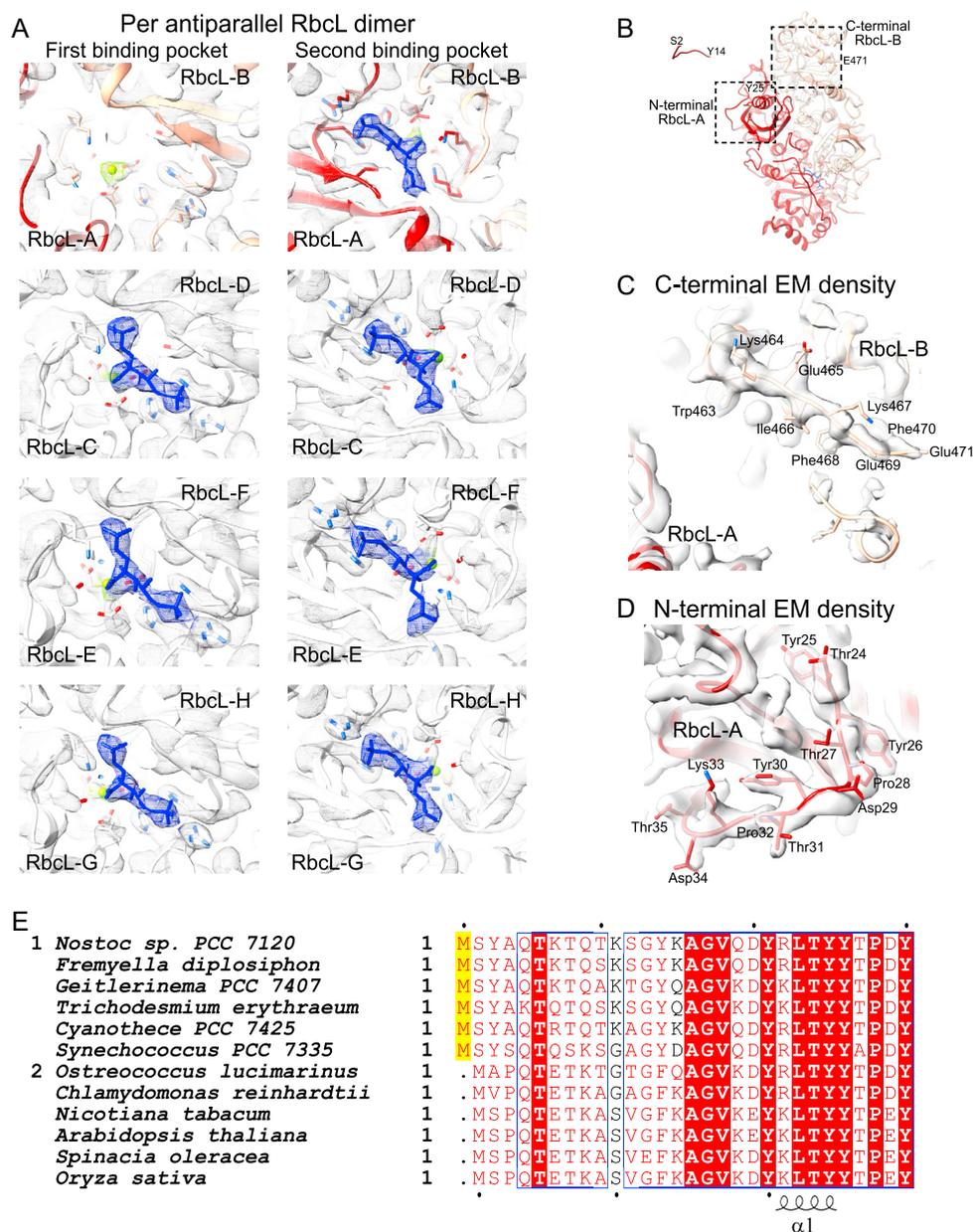


Figure S4. Key Structural Features of Inhibited NosRubisco in Complex with NosRcaΔC, Related to Figures 4, 5, and 7; Data S2; and Videos S1 and S2

(A) Close-up of the cryo-EM density of the eight substrate binding pockets of NosRubisco. Each Rbcl anti-parallel dimer has two binding pockets. The density is represented as a transparent iso-contour surface. CABP density is shown as a meshwork in blue. Protein is represented by ribbons with ligand-interacting side chains and CABP in stick representation. Mg^{2+} ions are shown as green spheres.

(B) An overview of the Rbcl-A/Rbcl-B dimer that is engaged by NosRcaΔC. Rbcl-A and Rbcl-B are shown in red and peach, respectively. The N- and C-terminal residues, as well as the Rbcl-A N-terminal peptide bound in the hexamer pore of NosRcaΔC are indicated.

(C and D) Zoomed-in views of the N- and C-terminal regions, indicated by dotted boxes in (B). The EM densities of the C-terminal residues of Rbcl-B (C) and the N-terminal residues of Rbcl-A (D) are well-resolved in the cryo-EM density map of the NosRcaΔC:Rubisco complex.

(E) Alignment of the N-terminal sequences of Rbcl from selected cyanobacterial and eukaryotic form 1B Rubisco proteins. Similar residues are shown in red and identical residues in white on a red background. Blue frames indicate homologous regions. Yellow background indicates systematic differences between cyanobacterial and eukaryotic sequences. The secondary structure elements of Rbcl from *Oryza sativa* (PDB: 1WDD) (Matsumura et al., 2012) are indicated below the sequences. The Uniprot and Genebank accession codes for the sequences are: P00879, *Nostoc* PCC 7120; NZ JH930358.1, *Fremyella diplosiphon*; K9SCF1, *Geitlerinema* PCC 7407; Q10WH6, *Trichodesmium erythraeum*; B8HQS5, *Cyanothece* PCC 7425; B4WP00, *Synechococcus* PCC 7335; Q0P3J3, *Ostreococcus lucimarinus*; P00877, *Chlamydomonas reinhardtii*; P00876, *Nicotiana tabacum*; O03042, *Arabidopsis thaliana*; P00875, *Spinacia oleracea*; P0C510, *Oryza sativa*.

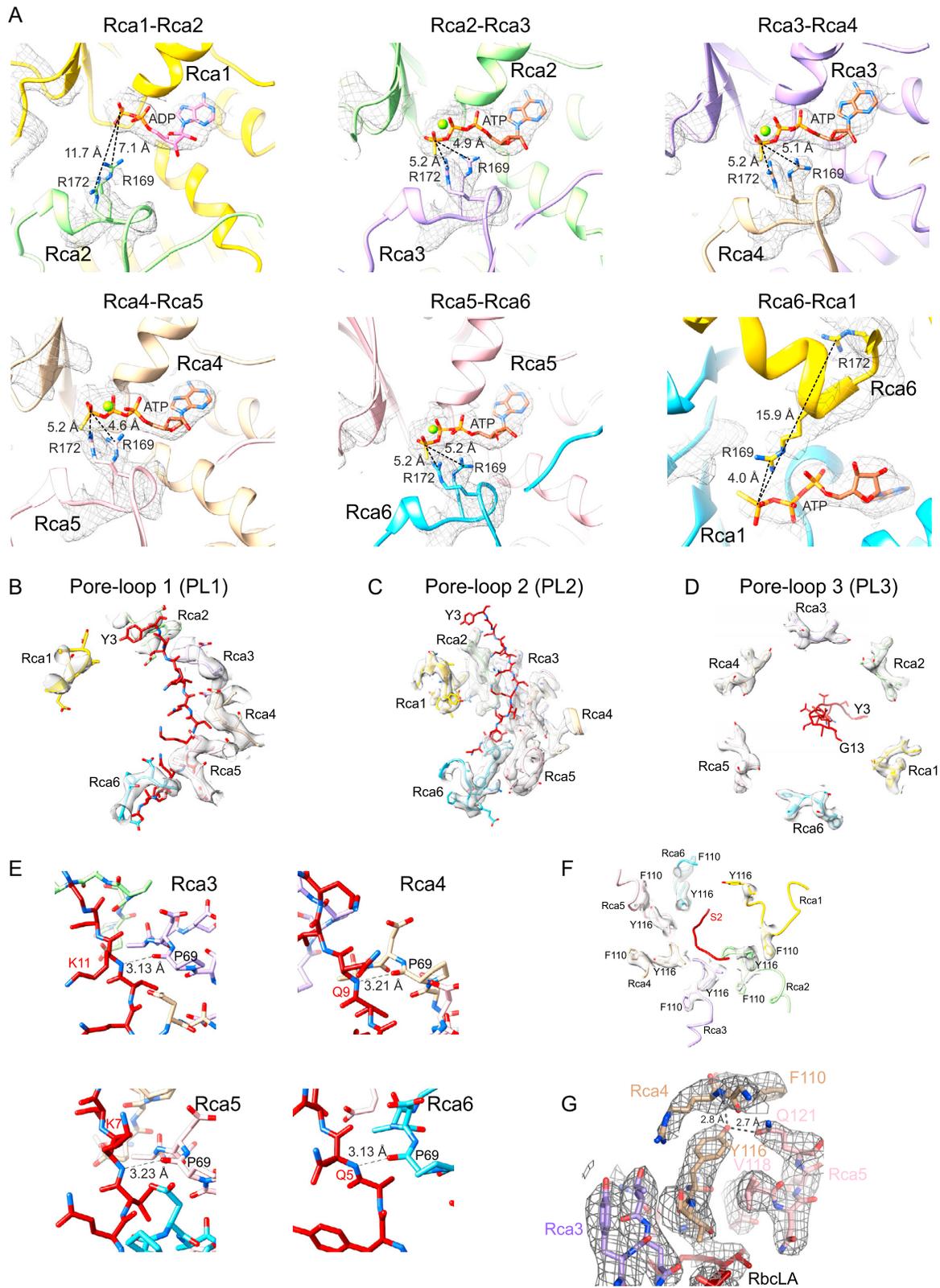


Figure S5. Key Structural Features in *NosRca*ΔC in the Complex with *NosRubisco*, Related to Figure 4 and Data S2

(A) Zoomed-in views of the nucleotide binding pockets formed between adjacent *NosRca* subunits. Distances of the carbon atom in the guanidino group of Arg169 and Arg172 to the last phosphorus atom of the nucleotide are indicated. Cryo-EM density is shown as a meshwork. The protein is shown in ribbon representation, nucleotide in stick representation and Mg^{2+} ion as a green sphere.

(B–D) Zoomed-in views of pore-loops PL1 (B), PL2 (C) and PL3 (D) in relationship to the bound N-terminal peptide of RbcL-A. Backbone in ribbon representation and side-chains in stick representation.

(E) Putative hydrogen bonds formed by the staggered pore-loops PL1 of Rca3 to Rca6 (Pro69) with the backbone of the RbcL N-terminal peptide at residues Lys11, Gln9, Lys7 and Gln5, respectively. Hydrogen bonds are indicated by dashed lines.

(F) Cryo-EM density of the aromatic residues Tyr116 and Phe110 in pore-loops PL2 of Rca1 to Rca6.

(G) Inter- and intramolecular contacts of the side-chain of Tyr116 of Rca4. Similar interactions were observed in all ATP-bound subunits, Rca2 to Rca6. The hydroxyl group of Tyr116 likely forms hydrogen bonds with the backbone of Phe110 in the same subunit, and to Gln121 in the subsequent subunit. Intermolecular van der Waals contacts are observed with Val118. Hydrogen bonds are indicated by dashed lines.

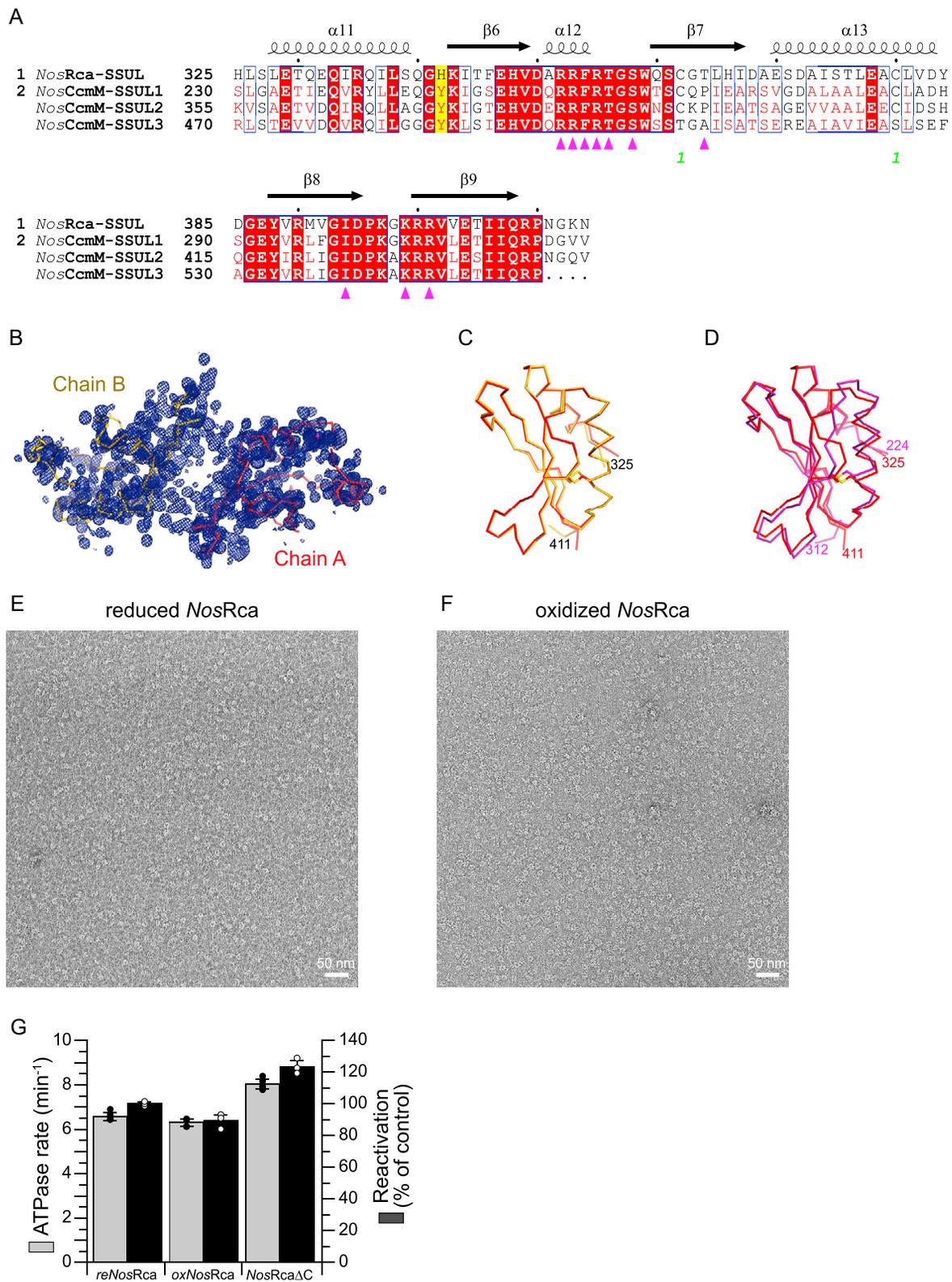


Figure S6. Role of the SSUL Domains of *NosRca* in Rubisco Binding, Related to Figures 6 and 7

(A) Sequence alignment of the SSUL domain of *NosRca* (*NosSSUL*) and the three SSUL modules of the scaffolding protein *CcmM* from *Nostoc* sp. PCC 7120. Secondary structure elements for *NosSSUL* are indicated above the sequence. Similar residues are shown in red and identical residues in white on a red

(legend continued on next page)

background. Blue frames indicate homologous regions. Yellow background indicates systematic differences between *NosSSUL* and the three SSUL modules of *NosCcmM*. Triangles (magenta) indicate contact residues with Rubisco. The green characters '1' indicate the position of the disulfide bridge in the crystal structure of *NosSSUL*. The Uniprot accession codes for the sequences are: P58555, *NosSSUL* domain from *Nostoc* sp. PCC 7120; Q8YYI3, *CcmM* from *Nostoc* sp. PCC 7120.

(B) Unbiased experimental electron density for the asymmetric unit in crystals of the *NosSSUL* domain at 1.4 Å resolution. The density after MAD phasing and density modification contoured at 1.5 σ is shown as a blue meshwork. The asymmetric unit contained two nearly identical copies (rmsd 0.43 Å for all $C\alpha$ positions) (for data collection and refinement statistics, see [Table S2](#)). Chains A and B of the *NosSSUL* domain are shown as $C\alpha$ traces in red and yellow, respectively.

(C) Superposition of chain A with chain B in the asymmetric unit of the *NosSSUL* crystal lattice. The rmsd for matching $C\alpha$ positions is 0.27 Å. Chain termini are indicated.

(D) Superposition of the *NosSSUL* domain with the SSUL1 domain of *CcmM* from *Synechococcus elongatus* PCC 7942 (SeSSUL; PDB: 6HBA). The SeSSUL is shown in purple. The rmsd for matching $C\alpha$ positions is 1.0 Å (82 matching $C\alpha$ positions).

(E and F) Negative-stain electron micrographs of reduced (A) and oxidized (B) *NosRca* (125 nM), demonstrating the presence of ring-complexes even at this low concentration.

(G) ATPase and reactivation rates of *reNosRca*, *oxNosRca* and *NosRca* Δ C. ATPase rates were measured in the absence of Rubisco at 20 mM KCl (see [STAR Methods](#)). CO₂ fixation was measured for 8 min and set to 100% for *reNosRca*. Error bars represent SD of at least three independent experiments.

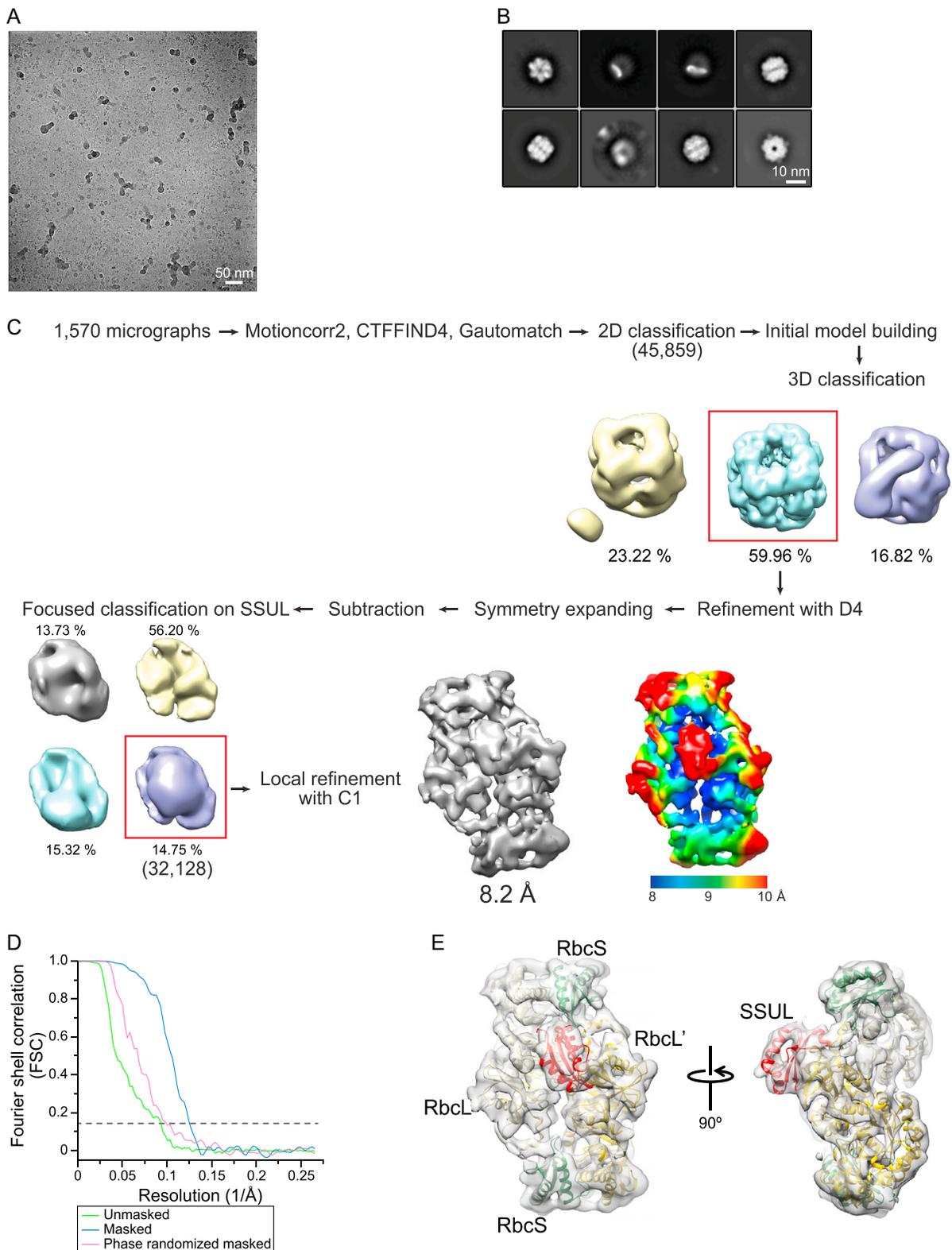


Figure S7. Cryo-EM Single-Particle Reconstruction of NosRca:Rubisco Complex, Related to Figures 6 and 7

(A) A representative micrograph of NosRca:Rubisco complexes.

(B) 2D class averages of complexes in (A).

(legend continued on next page)

(C) The single-particle data processing workflow for *NosRca:Rubisco* complex and local resolution map. The color gradient from blue to red indicates local resolution from 8.0 to 10 Å.

(D) Gold-standard FSC curves of the *NosRca:Rubisco* reconstruction. The resolution is ~8.2 Å at the FSC cutoff of 0.143 for the masked and B-factor sharpened curve.

(E) Density map of 3D reconstruction from cryo-EM and single-particle analysis of 2RbcL-2RbcS-SSUL units at ~8.2 Å resolution. The structural model in ribbon representation is docked into the cryo-EM density. The RbcL subunits in gold, RbcS in green and the bound SSUL domain in red.

4.3 Scaffolding protein CcmM directs multiprotein phase separation in β -carboxysome biogenesis (in revision)

1

2

3 **Scaffolding Protein CcmM Directs Multiprotein Phase**

4 **Separation in β -Carboxysome Biogenesis**

5

6 Kun Zang^{1#}, Huping Wang^{1#}, F. Ulrich Hartl¹ and Manajit Hayer-Hartl^{1,*}

7

8 ¹Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, Am Klopferspitz
9 18, 82152 Martinsried, Germany

10

11

12 #These authors contributed equally

13

14 *Correspondence: mhartl@biochem.mpg.de (M.H-H.)

15 **ABSTRACT**

16 Carboxysomes in cyanobacteria enclose the enzymes Rubisco and carbonic anhydrase to
17 optimize photosynthetic carbon fixation by Rubisco. Understanding carboxysome assembly has
18 implications in agricultural biotechnology. Here we used a combined structural and biochemical
19 approach to analyze the role of the trimeric scaffolding protein CcmM in the biogenesis of
20 carboxysomes in β -cyanobacteria. We find that the γ CAL domains of CcmM recruit the
21 tetrameric carbonic anhydrase, CcaA, by specifically binding the C-terminal tails of CcaA. The
22 SSUL modules of CcmM, besides their known function in recruiting Rubisco, also participate in
23 intermolecular interactions with the γ CAL domains, providing additional valency for network
24 formation. Moreover, the γ CAL domains mediate a head-to-head association of CcmM trimers.
25 These interactions cooperate to form an immobile multiprotein matrix. Our findings reveal the
26 mechanism by which CcmM functions as a central organizer of the pre-carboxysome,
27 sequestering the core components Rubisco and CcaA prior to carboxysome shell formation.

28

29 **Keywords:** Cyanobacteria; β -carboxysome; carbonic anhydrase; carbon concentration
30 mechanism; liquid-liquid phase separation; Rubisco; redox regulation; x-ray crystallography;
31 cryo-electron microscopy.

32 INTRODUCTION

33 Bacterial microcompartments (BMCs) are proteinaceous icosahedral structures that encapsulate a
34 segment of a metabolic pathway^{1,2}. Carboxysomes in cyanobacteria enclose the photosynthetic
35 key enzyme Rubisco (ribulose-1,5-bisphosphate carboxylase/oxygenase) together with carbonic
36 anhydrase (CA), generating a high CO₂ environment to optimize carbon fixation^{1,3,4}. The
37 implementation of a carboxysome-like CO₂ concentrating mechanism (CCM) in chloroplasts is
38 considered an important strategy for improving photosynthetic efficiency in crop plants⁵⁻¹⁴.
39 Successful building of carboxysomes will require a detailed understanding of their biogenesis.
40 Recent advances have shown that early in this process specialized scaffolding proteins initiate
41 phase separation of Rubisco into a condensate for subsequent encapsulation¹⁵⁻¹⁸. However, the
42 mechanisms underlying the sequestration of CA and other components together with Rubisco are
43 not yet understood.

44 Form I Rubisco, a complex of eight large (RbcL) and eight small (RbcS) subunits,
45 evolved in an atmosphere rich in CO₂. The drop in atmospheric CO₂ about 500 million years ago
46 generated the evolutionary pressure for the development of the CCM in photosynthetic
47 bacteria¹⁹. The proteinaceous shell of carboxysomes allows the entry of dissolved CO₂ in the
48 form of HCO₃⁻, which is converted to CO₂ by CA inside the carboxysome²⁰ (Fig. 1a). Since the
49 shell prevents CO₂ from diffusing out²¹, high levels of CO₂ are generated in the vicinity of
50 Rubisco^{19,22}, avoiding the competing side-reaction with oxygen (photorespiration)²³.

51 Carboxysomes come in two forms, α and β , which differ in their components and are likely the
52 result of convergent evolution^{3,22,24}.

53 The protein CcmM has been implicated as a central organizing scaffold in β -
54 carboxysome biogenesis²⁵⁻²⁷. CcmM in *Synechococcus elongatus* PCC 7942 (*Se7942*) is a

55 homotrimer of ~58 kDa subunits (*SeM58*). The M58 subunit consists of a γ CA domain at its N-
56 terminus followed by three Rubisco small subunit-like (SSUL) modules connected by flexible
57 linkers. A smaller, monomeric isoform (*SeM35*; ~35 kDa), comprising only the SSUL modules,
58 is generated from an internal ribosome binding site²⁸. CA activity is provided either by the γ CA
59 domain and/or by a separate β CA protein, CcaA (~30 kDa in *Se7942*)^{20,29}. The γ CA domain
60 mediates M58 trimer formation and is ~35% identical to canonical γ -class CA proteins^{30,31}. In
61 some β -cyanobacteria the γ CA domain has lost the functional motifs for CA activity^{26,30,32}, and is
62 referred to as γ CA-like (γ CAL). The β CAs, on the other hand, function as dimers or higher-order
63 oligomers^{25,33-37}. Deletion of CcaA results in loss of carbon concentrating function with growth
64 being dependent on high CO₂ (5%)^{38,39}. CcaA is redox-regulated and active only in the oxidizing
65 environment of the carboxysome⁴⁰. Recruitment of CcaA to carboxysomes is mediated by the
66 γ CAL domain of CcmM^{25,26} by a mechanism that remains to be elucidated.

67 We recently reported that in the initial stage of β -carboxysome biogenesis, the SSUL
68 modules of M35 bind to Rubisco to induce condensate formation¹⁵. The M35-Rubisco
69 interaction is redox-regulated by disulfide bond formation in the first two SSUL modules, with
70 higher affinity in the reducing cytosol and a lower affinity in the oxidizing carboxysome¹⁵. Here
71 we used a combined structural and biochemical approach to understand the interactions of M58
72 with CcaA and Rubisco. Our results reveal multiple, interwoven demixing and co-assembly
73 reactions with M58 functioning as a central organizer of the pre-carboxysome. The γ CAL
74 domains of trimeric M58 recruit CcaA by binding the C-terminal peptide sequence of CcaA in a
75 specific pocket. Moreover, we find that the high local concentration of SSUL modules in trimeric
76 M58 provides enhanced avidity for Rubisco compared to monomeric M35. This effect is
77 amplified by a head-to-head association of M58 trimers via their γ CAL domains. Additionally,

78 the SSUL modules engage in dynamic electrostatic interactions with the γ CAL domains,
79 facilitating M58 homo-demixing. These interactions cooperate to facilitate the efficient
80 multiprotein co-assembly of CcmM (M58 and M35), CcaA and Rubisco for encapsulation into
81 β -carboxysomes.

82

83 RESULTS

84 CcmM-CcaA condensate formation

85 To obtain mechanistic insight into the interaction of CcmM and CcaA, we recombinantly
86 expressed and purified CcmM (M58 and M35) and the CcaA of *Se7942* (Supplementary Fig. 1a-
87 c). Size-exclusion chromatography coupled to light scattering (SEC-MALS) confirmed that M58
88 is a trimer in solution, while CcaA behaved as a tetramer (Supplementary Fig. 1d,e and
89 Supplementary Table 1). To investigate whether the interaction between M58 and CcaA results
90 in condensate formation, we performed turbidity assays. While no turbidity was detected for
91 either protein alone under reducing or oxidizing conditions (Fig. 1b,c), a strong turbidity signal
92 was observed when M58 and CcaA were combined (Fig. 1b-d). Turbidity developed at a similar
93 rate independent of the redox state of the M58 SSUL modules ($t_{1/2} \sim 0.6$ min at 100 mM KCl)
94 with an apparent affinity (K_D^{app}) of CcaA for M58 of ~ 0.2 μ M (Fig. 1d). This is in contrast to the
95 redox-dependence of the interaction of M35 with Rubisco, where the interaction is 5-fold weaker
96 under oxidizing conditions¹⁵. The M58-CcaA interaction was impaired by high salt (Fig. 1b,c)
97 and thus involves electrostatic forces, similar to condensate formation between M35 and
98 Rubisco¹⁵. Microscopy of M58 or CcaA labeled N-terminally with fluorophores Alexa532
99 (M58_{AF5}) and Alexa405 (CcaA_{AF4}), and mixed 1:10 with the respective unlabeled protein,
100 showed a diffuse distribution (Fig. 1e,f). When the two proteins were combined at equimolar

101 concentration (0.25 μM), co-assembly into fluorescent condensates with an average Feret's
102 diameter of $\sim 1.0 \mu\text{m}$ (Supplementary Fig. 2a) was observed (Fig. 1e,f). Fluorescence recovery
103 after photobleaching (FRAP) experiments showed no recovery of the fluorescence signal in both
104 channels, indicating a strong association between M58 and CcaA (Supplementary Fig. 2b).

105 In summary, condensate formation with M58 provides a plausible mechanism for CcaA
106 sequestration and recruitment into carboxysomes.

107

108 **CcaA engages the γCAL domain of M58**

109 The CcaA protein consists of a N-terminal βCA domain followed by a hydrophilic and
110 intrinsically disordered C-terminal sequence of 60–70 residues^{3,20}. This sequence contains two
111 conserved regions of ~ 15 residues, C1 and C2, separated by a poorly conserved, hydrophilic
112 sequence of ~ 40 -50 amino acids that is predicted to be unstructured^{33,41} (Fig. 2a). C1 has been
113 shown to be required for oligomerization and enzymatic activity of CcaA⁴¹, while the function of
114 C2 has remained unclear^{27,33}. To understand the function of C2, we generated a mutant of CcaA
115 lacking the last 15 residues (CcaA ΔC2) (Fig. 2a, Supplementary Fig. 1a and Supplementary
116 Table 1). Notably, no turbidity signal was observed upon mixing CcaA ΔC2 and M58 (Fig. 2b),
117 and no demixing into fluorescent condensates was detected (Supplementary Fig. 2c). These
118 results suggested that the C2 sequence of CcaA mediates the interaction with the γCAL domains
119 of M58²⁶ for assembly into carboxysomes.

120 To confirm this interaction, we recombinantly expressed and purified the γCAL domain
121 of M58 (residues 1-198) (Fig. 2c, Supplementary Fig. 1a and Supplementary Table 1). As
122 expected, no condensate formation of CcaA with M35, which lacks the γCAL domain, was
123 observed (Fig. 2c,d). CcaA (0.5 μM) also failed to develop turbidity when combined with the

124 trimeric γ CAL (0.75 μ M) (Fig. 2d). However, upon increasing the concentration of CcaA and
125 γ CAL by 10-fold, a clear absorbance signal developed with slow kinetics ($t_{1/2}$ of \sim 5 min) (Fig.
126 2e), while no interaction was observed with M35 even at the higher concentration (22.5 μ M)
127 (Supplementary Fig. 2d). Interestingly, unlike the M58-CcaA interaction (Fig. 1b,c), the
128 interaction of CcaA with the γ CAL domain was somewhat enhanced at higher salt (Fig. 2e),
129 suggesting a contribution of hydrophobic forces to complex formation. Fluorescence microscopy
130 revealed small condensates on a background of diffusely distributed CcaA and γ CAL
131 (Supplementary Fig. 2e).

132 We next tested whether the C2 region of CcaA was sufficient to mediate binding to
133 γ CAL. To address this possibility, we attached the 15 residue C2 sequence
134 (LAPEQQRIYRGNAS) to the C-terminus of enhanced GFP via a short flexible linker
135 (GSGGS) ($E_{GFP}C2_{15}$) (Fig. 2f). No complex formation of $E_{GFP}C2_{15}$ with γ CAL was detected
136 upon analysis by native-PAGE (Fig. 2f). However, we noticed the presence of a tryptophan
137 residue just N-terminal of the C2 region, which would increase the hydrophobicity of the
138 sequence. Indeed, the GFP construct containing the last 17 residues of CcaA
139 (GWLAPEQQRIYRGNAS) ($E_{GFP}C2_{17}$) readily bound to γ CAL (Fig. 2f).

140 In summary, CcaA interacts with γ CAL via its C-terminal C2 sequence, and this
141 interaction is critical for the recruitment of CcaA by M58 into the pre-carboxysome.

142

143 **Structure of the C-terminal peptide of CcaA bound to the γ CAL domain**

144 Next to obtain structural insight into the interaction of the C2 sequence with $Se\gamma$ CAL (residues 1-
145 198), we analyzed the γ CAL-C2₁₇ complex by X-ray crystallography. However, crystals of the
146 complex diffracted to only \sim 3.5 Å resolution, prohibiting structure solution. A structural model

147 of γ CAL, based on the crystal structure of the γ CA domain of CcmM from
148 *Thermosynechococcus elongatus* BP-1 (residues 1-209, PDB: 3KWC)³⁰ (Supplementary Fig. 3a),
149 suggested that helix α 2 of *Se* γ CAL contains a break at residues I179 to S182, which might have
150 interfered with the formation of well-diffracting crystals. This position corresponds to the
151 resolved C-terminus of a shorter crystallized form of *Te* γ CA (residues 1-193; PDB: 3KWD)
152 (Supplementary Fig. 3a,b). Indeed, *Se* γ CAL containing residues 1-181 (Supplementary Figs. 1a
153 and 3a, and Supplementary Table 1) produced well-diffracting crystals, allowing us to solve the
154 structure of *Se* γ CAL by molecular replacement (PDB: 3KWC) to a resolution of 1.67 Å (Fig. 3a,
155 Supplementary Fig. 3c and Supplementary Table 2). The overall structure of the γ CAL protomer
156 is highly similar to that of *Te* γ CA (PDB: 3KWD) ($C\alpha$ r.m.s.d. 0.48 Å) (Supplementary Fig. 3d):
157 it consists of an N-terminal seven-turn left-handed β -helix, followed by a short helix α 1, a long
158 linker and part of helix α 2, which packs along one face of the β -helix (Fig. 3a and
159 Supplementary Fig. 3c,d). In the asymmetric unit of the crystal, γ CAL(1-181) is a protomer,
160 while it exists as a trimer in solution (Supplementary Table 1). Indeed, analysis using PISA⁴²
161 indicated an extensive interface between subunits (4180 Å² buried at the subunit interface from a
162 total accessible surface of 18860 Å²) (Supplementary Fig. 3e).

163 We also solved the structure of γ CAL(1-181) in complex with the C2₁₇ peptide of CcaA
164 at a resolution of 1.63 Å (Fig. 3b, Supplementary Fig. 3f,g and Supplementary Table 2). The
165 asymmetric unit of the crystal contained the protomer of γ CAL with one peptide bound (Fig. 3b
166 and Supplementary Fig. 3f), suggesting that all binding sites of the trimeric γ CAL are occupied
167 by peptide. Analysis by isothermal titration calorimetry (ITC) using EGFP-C2₁₇ revealed a binding
168 affinity (K_D) of ~2 μ M at a stoichiometry of ~2.7 per trimeric γ CAL (Supplementary Fig. 3h).
169 The peptide is bound as a two-turn α -helix (residues PEQQRIY) in a pocket beneath the

170 protruding β 10- β 11 loop, making extensive interactions with residues in β 11, β 17, the β 19- β 20
171 loop and α 1 of the γ CAL protomer. The hydrophobic interactions include π -stacking between the
172 indole ring of residue W257 in the C2 peptide with the benzene ring of F123 (β 17) in γ CAL (Fig.
173 3c). In addition, the hydrophobic residue Y267 of the C2 peptide contacts the γ CAL residue L81
174 (β 11) (Fig. 3c). Two electrostatic interactions (salt bridges) are formed by the guanidinium group
175 of the peptide residue R265 with E143 (β 20) and D163 (α 1) of γ CAL (Fig. 3c). The side chains
176 of peptide residues Q262 and Q263 form hydrogen bonds with the backbone of D141 (β 19- β 20
177 loop) and the side chain of N124 (β 17) in γ CAL, respectively. The side chain of peptide residue
178 Q262 also forms a hydrogen bond with the side chain of γ CAL residue Q159 (α 1). Moreover, the
179 guanidinium group of R126 (β 17) in γ CAL forms hydrogen bonds with the backbone of I266 and
180 Y267 in the C2 peptide, and another hydrogen bond is formed between the backbone of peptide
181 residue Y267 and L81 (β 11) of γ CAL (Fig. 3c).

182 In summary, each protomer of the γ CAL trimer interacts with one protomer of a CcaA
183 tetramer by binding the C-terminal C2 sequence of CcaA, resulting in multivalency as the basis
184 for condensate formation. The binding site of C2 is highly conserved in the γ CA/ γ CAL domains
185 of CcmM proteins.

186

187 **Contribution of SSUL modules to the M58-CcaA interaction**

188 The interaction between M58 and CcaA involves two components: (i) a salt sensitive interaction
189 between full-length M58 and CcaA, and (ii) the mainly hydrophobic interaction between the
190 γ CAL domain and the C2 peptide of CcaA. To understand the salt-sensitive component in more
191 detail, we investigated the interaction of CcaA with C-terminal truncation mutants of M58,
192 containing either two (γ CAL-2S) or one (γ CAL-1S) SSUL modules (Fig. 4a and Supplementary

193 Table 1). Condensate formation with CcaA, monitored by turbidity assay, was only mildly
194 impaired with γ CAL-2S, but was strongly reduced with γ CAL-1S (Fig. 4b). This finding
195 indicated that the SSUL modules provide additional, critical valency for condensate formation.
196 Since no binding of CcaA to M35 was observed (Fig. 2d and Supplementary Fig. 2d), this raised
197 the question of how the SSUL modules contribute CcaA-M58 complex formation. Might the
198 SSUL modules mediate homo-oligomeric interactions between M58 trimers?

199 To address this possibility, we next analyzed whether M58 can undergo condensate
200 formation on its own. Reasoning that interactions involving SSUL modules might be more salt
201 sensitive¹⁵, we conducted this analysis at a reduced salt concentration (50 mM KCl).
202 Interestingly, we observed homo-demixing of both reduced and oxidized M58 (M58_{red} and
203 M58_{ox}, respectively), with M58_{red} requiring somewhat higher concentrations (Fig. 4c,d and
204 Supplementary Fig. 4a). Fluorescence microscopy showed that condensate formation with M58_{ox}
205 was enhanced compared to M58_{red} (Fig. 4d and Supplementary Fig. 4b). As expected, homo-
206 demixing was highly salt-sensitive (Supplementary Fig. 4c,d). Notably, mutation of the disulfide
207 forming cysteines in SSUL1 and SSUL2 to serine disrupted homo-demixing (Supplementary Fig.
208 4e), consistent with the redox-dependence of condensate formation.

209 We next asked whether the SSUL modules are required for M58 homo-demixing.
210 Removal of one SSUL module from M58 (γ CAL-2S) had only a minor effect, while removal of
211 two SSUL modules (γ CAL-1S) completely abolished M58_{ox} homo-demixing (Fig. 4e),
212 suggesting that the SSUL modules play a role in mediating the interaction between M58 trimers.
213 Since M35 on its own did not undergo phase separation (Supplementary Fig. 2d), such
214 interactions would have to be specific for trimeric M58. To test this, we engineered a M35 trimer
215 by fusing a trimeric coiled-coil sequence⁴³ to the M35 N-terminus (CC_{TRI}M35) (Supplementary

216 Fig. 4f and Supplementary Table 1). However, no turbidity signal was detectable with
217 CC_{TRI}M35, even at high concentrations and low salt (Supplementary Fig. 4f), pointing to an
218 interaction of the SSUL modules with the γ CAL domains in M58, and not between SSUL
219 modules.

220 In summary, the SSUL modules contribute to both M58-CcaA condensate formation and
221 M58 homo-demixing, apparently by interacting with the γ CAL domains.

222

223 **Charge-charge interactions between SSUL and γ CAL**

224 Both the SSUL and γ CAL domains expose multiple positively and negatively charged residues.
225 An analysis of 500 SSUL and 150 γ CA/ γ CAL sequences revealed high conservation of \sim 19 and
226 \sim 20 charged residues, respectively (Supplementary Fig. 5a-d). Individual mutations of the
227 arginine residues 251, 252 or 298 in SSUL1, or R367 in SSUL2 or R481 in SSUL3 to aspartate
228 essentially abolished homo-demixing of M58_{ox}, as monitored by turbidity assays (Fig. 4f,
229 Supplementary Fig. 1b,c and Supplementary Table 1), indicating that all three SSUL modules are
230 involved. In contrast, individual mutations of the conserved negatively charged residues E246,
231 D249, D294 and E303 in SSUL1 to lysine resulted in enhanced turbidity, except for the mutant
232 E286K, which behaved like wild-type (Fig. 4f). Notably, all these charged residues are located
233 on one face of SSUL (Supplementary Fig. 5b), suggesting that this surface may be involved in
234 the interaction with γ CAL. Interestingly, this same surface is also involved in the scaffolding
235 function of the SSUL modules with Rubisco¹⁵.

236 To identify the surface of the γ CAL domain that may interact with SSUL, we individually
237 mutated the positively charged γ CAL residues R37, R43, K62, R79, R95 or R126 to aspartate,
238 and the negatively charged residues E17, D21, D35, E76 or D112 to lysine (Supplementary Fig.

239 1b,c). Among these residues, R79, R126, E17, D35, E76 and D112 are relatively conserved
240 (Supplementary Fig. 5c). Note that the trimeric state of M58 was maintained for all mutants
241 (Supplementary Table 1). Interestingly, the effect of these charge mutations was reversed
242 compared to the mutational analysis of the SSUL modules (Fig. 4f,g): conversion of positive
243 charges to negative enhanced M58_{ox} homo-demixing (except for mutant R43D), while changing
244 negative charges to positive reduced condensate formation (except for mutant D21K) (Fig. 4g
245 and Supplementary Fig. 5d). The residues that exert an effect are located on the surface of the
246 γ CAL domain facing the solvent and are not involved in γ CAL trimer formation (Supplementary
247 Fig. 5c,d). In contrast, residues D21K and R43, mutation of which had no effect, are at the edge
248 of the putative binding surface and are not highly conserved (Fig. 4g and Supplementary Fig.
249 5c,d). Notably, the binding region on γ CAL for SSUL does not overlap with the binding site for
250 the C2 peptide. Indeed, SSUL and CcaA binding are both required for efficient formation of the
251 M58-CcaA condensate (Fig. 4b)

252 In summary, the overall interaction between SSUL and γ CAL involves a complex
253 interplay of attractive and repulsive forces, consistent with single charge reversal mutations
254 having reducing or enhancing effects (Fig. 4f,g). The binding region for SSUL on γ CAL is
255 independent of that for the C2 peptide of CcaA.

256

257 **Head-to-head association of γ CAL trimers**

258 To understand the structural basis of the intermolecular interactions of M58, we next performed
259 cryo-electron microscopy (cryo-EM) and single-particle analysis of M58_{ox}. Reference-free 2D
260 class averages revealed a class of barrel-shaped complexes with dimensions of $\sim 5 \times 8.2$ nm,
261 consistent with two-stacked γ CAL trimers in side view (Supplementary Fig. 6a,b). Notably, a

262 head-to-head association of γ CA trimeric domains with very similar dimensions is present in the
263 asymmetric unit of the *Te* γ CA crystal (CcmM209; PDB: 3KWC)³⁰. Interestingly, we also
264 observed this head-to-head interaction in the molecular packing of both the γ CAL and γ CAL-
265 C2₁₇ crystals. A 3D classification without imposed symmetry of the 2D class averages resulted in
266 an EM-density map of ~ 3.6 Å resolution (Fig. 5a, Supplementary Fig. 6 and Supplementary
267 Table 3).

268 In the cryo-EM density map there was significant information loss in side views, due to
269 the end view orientation being preferred (Supplementary Fig. 6b). In contrast, in the end view of
270 the density map the three seven-turn β -helices and many bulky side-chains were well-resolved
271 (Fig. 5a and Supplementary Fig. 6e), thus allowing docking of the stacked γ CAL trimers from
272 the crystallographic model (Supplementary Fig. 3e). Additional densities are seen to protrude
273 from the edges of the complex (Fig. 5a), likely representing SSUL modules interacting either
274 intra- or intermolecularly. Note that while SSUL2 and SSUL3 may function preferentially to
275 form intermolecular contacts, our mutational analysis showed that all three modules participate
276 in M58 homo-demixing (Fig. 4f). The putative SSUL densities are smaller than the size of the
277 SSUL module and are of low resolution, suggesting a dynamic interaction, which precluded
278 docking and determination of the orientation of bound SSUL. Consistent with the mutational
279 analysis (Fig. 4f,g), it would appear that the SSUL binding site is located to the region above the
280 β 10- β 11 loop of γ CAL (Fig. 3a). This region is distinct from the interaction site of the C2
281 sequence of CcaA, which binds in a pocket below this loop (Fig. 3).

282 Both the γ CAL and γ CAL-C2₁₇ crystals revealed the presence of two protomer-protomer
283 salt-bridges across the dimer interface formed by the conserved residues R164 and D172 (Fig.
284 5b). Thus, a total of six salt-bridges stabilize the dimer-of-trimers. Note that in solution, γ CAL

285 nevertheless behaved as a trimer (Supplementary Table 1), suggesting that the head-to-head
286 association is dynamic, and occurs only at the high protein concentrations within the condensate.
287 To investigate the functional significance of this interaction, we disrupted the salt-bridges by
288 either mutating R164 to aspartate or D172 to lysine. Strikingly, both the R164D and D172K
289 mutants strongly reduced homo-demixing of M58 (Fig. 5c), indicating that the dimer-of-trimer
290 formation via the γ CAL domains provides critical valency, presumably by increasing the local
291 concentration of the SSUL modules. This increase in local density would enhance the avidity of
292 the SSUL modules for γ CAL domains of adjacent M58 complexes, thereby providing the
293 multivalency required for condensate formation. In contrast, disrupting the M58 head-to-head
294 dimer did not affect the interaction of CcaA with M58 via the γ CAL domains (Fig. 5d).

295 In summary, distinct interactions cooperate to ensure recruitment of CcaA into
296 carboxysomes: (i) the extreme C-terminal sequence (C2) of CcaA binds the γ CAL domain of
297 M58 driven by a combination of hydrophobic and electrostatic interactions; (ii) the SSUL
298 modules of M58 engage the γ CAL domains of adjacent M58 trimers via dynamic multivalent
299 electrostatic interactions. The C2 peptide and SSUL module bind to distinct regions on γ CAL;
300 (iii) M58 also undergoes homo-demixing mediated by both intermolecular γ CAL-SSUL
301 interactions and a M58 head-to-head association via the γ CAL trimers.

302

303 **M58 binds Rubisco with high affinity**

304 The SSUL modules of M35 have recently been shown to link the Rubisco holoenzyme¹⁵. To
305 investigate how the trimeric M58 interacts with Rubisco, we first determined the apparent
306 affinities of M58_{red} and M58_{ox} for Rubisco. M58 in both redox states displayed essentially
307 identical apparent affinities (K_D^{app}) of $\sim 0.07 \mu\text{M}$ for Rubisco at 50 mM KCl (Fig. 6a). This

308 interaction was ~3 times stronger than that of M35 with Rubisco (K_D^{app} of ~0.2 μM)¹⁵,
309 presumably due to the increased local concentration of SSUL modules in the M58 trimer. We
310 confirmed this using CC_{TRIM35}, which showed a similar apparent binding affinity for Rubisco
311 (Fig. 6a). The trimeric γCAL alone failed to interact with Rubisco, as expected (Fig. 6b,c).
312 Moreover, the interaction of M58 with Rubisco displayed salt resistance (up to 300 mM KCl for
313 M58_{red} and 200 mM KCl for M58_{ox}) (Supplementary Fig. 7a,b), in contrast to the salt sensitive
314 interaction of M35 with Rubisco¹⁵ (Fig. 6b,c). Notably, CC_{TRIM35} mimicked the salt resistance
315 observed with M58 (Supplementary Fig. 7c,d), further demonstrating that this property is due to
316 the high local concentration of SSUL modules in the trimer.

317 Cryo-EM analysis confirmed that the SSUL modules of M58 bind Rubisco in a groove
318 formed between two RbcL subunits and the adjacent RbcS¹⁵ (Supplementary Fig. 7e-h and
319 Supplementary Table 3). Interestingly, unlike M35, M58 also bound the RbcL₈ core complex of
320 Rubisco, albeit with a ~4-fold lower affinity (Supplementary Fig. 8a-c). As revealed by cryo-EM
321 analysis, the SSUL module occupied the same binding site on RbcL₈ as in the holoenzyme and
322 did not use the RbcS binding region (Supplementary Fig. 8d-g and Supplementary Table 3),
323 contrary to a recent suggestion⁴⁴. Thus, the high local SSUL concentration on M58 can
324 compensate for the missing interaction with RbcS¹⁵. However, the interaction of M58 with
325 RbcL₈ is unlikely to be significant *in vivo*, as Rubisco assembly factors, such as RbcX and/or
326 Raf1⁴⁵⁻⁴⁷, bind the RbcL₈ core and are only displaced upon RbcS binding^{15,23}. Addition of Raf1
327 could completely suppress the binding of M58 to RbcL₈ (Supplementary Fig. 8h), thus ensuring
328 that only the Rubisco holoenzyme would be recruited into carboxysomes.

329 In summary, the high local concentration of SSUL modules allows trimeric M58 to bind
330 Rubisco with higher affinity than M35 and in a redox-independent manner.

331

332 **Efficient co-assembly of M58, M35, CcaA and Rubisco**

333 Carboxysome biogenesis requires the scaffolding proteins M58 and M35 to sequester Rubisco
334 and CcaA under the reducing conditions of the cytosol. We next analyzed whether the complex
335 interactions of these proteins described above allow their co-assembly into a distinct condensate.
336 Combining Rubisco (N-terminally labeled with Alexa647, Rubisco_{AF6}, and mixed 1:10 with the
337 unlabeled protein) and M58 (M58_{AF5}:M58 1:10) at equimolar concentration resulted in
338 colocalization of both proteins within droplet shaped condensates (Fig. 7a), while no demixing
339 was observed with Rubisco and CcaA (1:10 of CcaA_{AF4}:CcaA) (Fig. 7b). As shown above, M58
340 also undergoes condensate formation with CcaA (Fig. 1e,f). We next investigated whether
341 Rubisco, M58 and CcaA can co-assemble. Assuming that CcaA could be substoichiometric to
342 M58 in the carboxysome²⁵, we first performed a sedimentation assay of the three proteins,
343 keeping the concentration of Rubisco and M58 constant (0.25 μ M each) and varying the
344 concentration of CcaA. We found that at an equimolar ratio, all three proteins were recovered in
345 the pellet fraction, with only a small amount of CcaA remaining in the supernatant
346 (Supplementary Fig. 9a). Using this condition, fluorescence microscopy demonstrated the
347 colocalization of all three proteins (differentially labeled) in a phase separated condensate (Fig.
348 7c).

349 M35 is present in carboxysomes in excess over M58²⁵. Upon addition of excess M35 (2
350 μ M) to the co-assembly reaction of Rubisco/M58/CcaA (0.25/0.25/0.125 μ M, respectively), all
351 four proteins were recovered in the pellet fraction upon sedimentation, with ~50% of M35
352 remaining in the supernatant (Supplementary Fig. 9b). Fluorescence microscopy using three
353 fluorophores to label either Rubisco, M35 and M58 or Rubisco, M58 and CcaA showed that all

354 four proteins efficiently colocalized (Fig. 7d,e). When M58 was replaced with CC_{TRI}M35, CcaA
355 no longer phase separated and was diffusely distributed (Fig. 7f). The average Feret's diameter
356 of the condensates varied from ~1.0 to ~2.5 μm dependent on total protein concentration
357 (Supplementary Table 4). FRAP experiments showed that M35 was relatively mobile in the
358 condensate with Rubisco (Fig. 7g), while the Rubisco was immobile (Fig. 7h). M35 mobility was
359 somewhat reduced in the condensate of the four proteins (Fig. 7i). In contrast, M58 was
360 immobile both with Rubisco or in the four-protein condensate (Fig. 7g-i), displaying the same
361 property as upon assembly with CcaA alone (Supplementary Fig. 2b).

362 In summary, the scaffolding proteins M58 and M35, differing in binding affinity and
363 dynamics, ensure the efficient sequestration of Rubisco and CcaA for co-packaging into
364 carboxysomes. M35 is the only component in the four-protein pre-carboxysome condensate that
365 shows some mobility.

366

367 **Discussion**

368 β -Carboxysome biogenesis involves the sequestration of Rubisco together with the carbonic
369 anhydrase CcaA, followed by shell formation, via an 'inside-out' assembly process⁴⁸. Our
370 biochemical and structural analysis elucidated how the scaffolding protein CcmM functions as a
371 central organizer in recruiting Rubisco and CcaA into the pre-carboxysome core. CcmM
372 orchestrates multiple, interwoven co-assembly reactions via its γ CAL and SSUL domains,
373 resulting in the formation of an essentially immobile protein mesh. Once captured under
374 reducing conditions, constituent carboxysome components cannot escape, facilitating efficient
375 encapsulation.

376 The ~21 kDa β -helical γ CAL domain of M58 (CcmM) is remarkably versatile and
377 participates in network formation through multiple distinct interactions (Fig. 8a,b). Each
378 protomer of the trimeric γ CAL binds the 17 residue C2 peptide of one protomer of tetrameric
379 CcaA with ~2 μ M affinity. The crystal structure identified multiple hydrophobic (salt stabilized)
380 contacts, including critical π -stacking of W257 in C2 with F123 in γ CAL. However, the γ CAL-
381 C2 interactions alone are inefficient in mediating M58-CcaA condensate formation. Additional
382 interactions between the SSUL modules and γ CAL are required. Both the SSUL modules and the
383 γ CAL domains expose multiple conserved positively and negatively charged residues, with the
384 overall interaction involving a complex interplay of attractive and repulsive forces. The
385 multivalency of the network is enhanced further by the ability of γ CAL trimers to form head-to-
386 head dimers.

387 Recruitment of Rubisco for carboxysome biogenesis is solely mediated by the SSUL
388 modules of M58 and M35, which bind in a groove at the interface between anti-parallel RbcL
389 dimers¹⁵. The trimeric M58 has a substantially higher affinity for Rubisco than M35, due to the
390 presence of 9 SSUL modules per M58 compared to only 3 SSUL modules per M35. What then is
391 the role of M35, which is present in excess over M58²⁵, and why are both proteins essential for
392 carboxysome biogenesis²⁸? We suggest that the differential redox-regulation of M35 and M58 is
393 important in converting the immobile pre-carboxysome condensate, required for initial capture of
394 Rubisco and CcaA, into a more dynamic state in the oxidizing interior of the carboxysome. This
395 redox regulation is mediated by disulfide bond formation in the SSUL modules, which is critical
396 for carboxysome biogenesis and the CCM *in vivo*¹⁵. Oxidation favors the interaction of the SSUL
397 modules with the γ CAL domains, thereby enhancing M58 homo-demixing. Indeed, under
398 oxidizing conditions, preformed M58_{ox} condensates were maintained within more dilute and

399 enlarged M58_{ox}-Rubisco droplets (Supplementary Fig. 10). The restructuring of the pre-
400 carboxysome condensate upon oxidation may promote the formation of channels around the
401 Rubisco lattice that can be navigated by other carboxysomal proteins and metabolites. This
402 would also facilitate the metabolic repair of Rubisco by Rubisco activase¹⁸.

403 In summary, our findings suggest the following model for the inside-out assembly of β -
404 carboxysomes: In the reducing cytosol, M58, in cooperation with M35, concentrates Rubisco and
405 CcaA into an immobile matrix that serves to effectively capture the core components for
406 encapsulation. In the oxidizing environment of the carboxysome, redox regulation of the SSUL
407 modules favors homo-demixing of M58, and renders the interaction of M35 with Rubisco more
408 dynamic. This transition is required for CCM function.

409

410 **Acknowledgements**

411 Technical assistance by R. Lange and S. Gaertner is gratefully acknowledged. We thank X. Yan
412 for performing some preliminary experiments and advice on droplet size analysis. A. Bracher is
413 acknowledged for help with X-ray data collection and crystal structure determination, J. Basquin
414 (MPIB crystallization facility) for help with X-ray data collection of the *SecY*CAL-C2₁₇ crystals,
415 C. Basquin (Department of Structural Cell Biology) for assisting with ITC measurements, and
416 the MPIB imaging facility for support. We are grateful for support by the staff at beamline ID23-
417 2 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France and at beamline
418 X06SA at the Swiss Synchrotron Light Source (SLS) in Villigen, Switzerland. We thank B.
419 Daniel and S. Tillman of the MPIB cryo-EM facility for help with EM data collection, and J.R.
420 Prabu for maintenance and development of the computational infrastructure.

421

422 **Author contributions**

423 M.H.-H. conceived and supervised the study. K.Z., H.W., F.U.H. and M.H.-H designed the
424 experiments. K.Z. and H.W. performed the structural and biochemical analysis. The paper was
425 written by M.H.-H and F.U.H., with contributions from K.Z. and H.W.

426

427 **Competing interests**

428 The authors declare no competing interests.

429 **FIGURE LEGENDS**

430 **Fig. 1 | Condensate formation of M58 and CcaA.** **a**, CO₂ concentrating mechanism (CCM) of
431 β -cyanobacteria. See Introduction for details. 3PGA, 3-phosphoglycerate; RuBP, ribulose-1,5-
432 bisphosphate, CA, carbonic anhydrase. **b-c**, Salt dependence of condensate formation of CcaA
433 and M58 from β -cyanobacterial species *Se7942*. Purified CcaA (0.5 μ M) and M58 (0.5 μ M)
434 were incubated in buffer containing 50-400 mM KCl under reducing (**b**) or oxidizing conditions
435 (**c**) at 25 °C. Condensate formation was monitored by turbidity assay at 340 nm. The proteins
436 alone were analyzed as control. Data are averages \pm SD (n=3). **d**, M58-CcaA condensate
437 formation dependent on the concentration of CcaA. Turbidity was measured as above at 0.5 μ M
438 of reduced M58 (M58_{red}) or oxidized M58 (M58_{ox}) with 0-1.25 μ M CcaA. The absorbance
439 values reached after 10 min incubation are plotted. Averages \pm SD are shown (n=3). **e-f**,
440 Condensate formation analyzed by fluorescence microscopy under reducing (**e**) and oxidizing (**f**)
441 conditions. CcaA (0.25 μ M) was mixed with either 0.25 μ M M58_{red} or M58_{ox}. M58 and CcaA
442 were N-terminally labeled with Alexa532 and Alexa405, respectively (M58_{AF5}; CcaA_{AF4}), and
443 used as 1:10 mixtures with unlabeled protein. Representative data are shown (n=3).

444
445 **Fig. 2 | Structural requirements for M58-CcaA condensate formation.** **a-b**, Condensate
446 formation depends on the C-terminal C2 peptide sequence of CcaA. (**a**) Domain structure of
447 CcaA and CcaA Δ C2, lacking the C-terminal 15 residues of CcaA. Amino acid numbers are
448 indicated. (**b**) Turbidity assays as in Fig. 1b,c at 100 mM KCl with 0.5 μ M CcaA/CcaA Δ C2 and
449 0.5 μ M M58_{red} or M58_{ox}. Data are averages \pm SD (n=3). **c**, Domain structure of M58, γ CAL and
450 M35 constructs. Amino acid numbers are indicated. The SSUL modules of M58 and M35 are
451 numbered SSUL1 to SSUL3 from N- to C-terminus. **d**, Condensate formation requires the γ CAL

452 domains and the SSUL modules of M58. Turbidity assays with the combinations of proteins
453 indicated were performed at 100 mM KCl using 0.75 μ M γ CAL, 0.5 μ M CcaA, 0.5 μ M
454 M58_{red}/M58_{ox} and 2.25 μ M M35. Note that the relative concentrations of M58 and M35 were
455 adjusted to maintain the SSUL modules at a similar concentration. Data are averages \pm SD (n=3).
456 **e**, Condensate formation of γ CAL and CcaA requires high protein concentrations and presence of
457 salt. Turbidity assays with the combinations of proteins indicated were performed at 50-200 mM
458 KCl using 5.0 μ M CcaA and 7.5 μ M γ CAL. Data are averages \pm SD (n=3). **f**, The C2₁₇ sequence
459 of CcaA is sufficient for CcaA binding to γ CAL. Left: eGFP-C2 constructs containing either the
460 C-terminal 15 or 17 residues of CcaA fused to eGFP. Right: The purified proteins (22.5 μ M)
461 were incubated with γ CAL (7.5 μ M) (100 mM KCl) for 15 min at 25 °C, followed by analysis of
462 complex formation by native-PAGE and Coomassie staining. Representative data are shown
463 (n=3).

464

465 **Fig 3 | Structure of the γ CAL-C2 complex.** **a**, Structural model of the *Se7942* γ CAL domain
466 (residues 1-181) at 1.67 Å resolution in ribbon representation. Secondary structural elements are
467 indicated. **b**, Structural model of the γ CAL(1-181)-C2₁₇ complex at 1.63 Å resolution. The γ CAL
468 domain (blue) is shown in transparent surface representation and the bound helical C2₁₇ peptide
469 (pink) in ribbon representation. **c**, Details of the interaction of C2 peptide with γ CAL. Proteins
470 are colored as in (b) and shown in ribbon representation with critical amino acid residues in stick
471 representation. Secondary structural elements and residue numbers are indicated.

472

473 **Fig. 4 | Intermolecular association between SSUL modules and γ CAL domains of M58.** **a**,
474 Schematic representation of wild-type M58 and truncation mutants containing two or one SSUL

475 module (γ CAL-2S and γ CAL-1S, respectively). Residue numbers are indicated. **b**, Contribution
476 of SSUL modules to M58-CcaA condensate formation. Turbidity was monitored over time after
477 mixing CcaA (0.5 μ M) with the reduced and oxidized wild-type M58 or truncated M58
478 constructs (0.5 μ M) indicated at 100 mM KCl. Data are averages \pm SD (n=3). **c**, Concentration
479 dependence of M58 homo-condensate formation. Turbidity assays were performed upon dilution
480 of M58_{ox} from high salt (500 mM KCl) to reach final protein concentrations of 0.5-3.0 μ M and a
481 salt concentration of 50 mM KCl. M58_{red} (5 and 6 μ M) was also analyzed. Data are averages \pm
482 SD (n=3). **d**, M58 homo-condensate formation is strongly enhanced under oxidizing conditions.
483 M58_{ox} and M58_{red} (2.5 μ M) homo-demixing was analyzed by fluorescence microscopy as in Fig.
484 1e,f. M58 was N-terminally labeled with fluorophore Alexa532 (M58_{AF5}) and used as a 1:10
485 mixture with unlabeled protein. Representative data are shown (n=3). **e**, M58_{ox} homo-condensate
486 formation is mediated by the SSUL modules of M58. Turbidity assays were performed for
487 reactions containing M58_{ox}, γ CAL-2S_{ox} and γ CAL-1S_{ox} (3 μ M each at 50 mM KCl). Data are
488 averages \pm SD (n=3). **f**, Mutation of charged residues in SSUL1, SSUL2 or SSUL3 inhibit or
489 enhance M58_{ox} homo-condensate formation, as measured by turbidity assay. Mutations of
490 negative to positive (D/E to K; blue arrow) enhance the interaction, while mutations of positive
491 to negative (R to D; red arrow) are inhibitory. Numbers of mutated residues are indicated.
492 Protein concentration was 3 μ M at 50 mM KCl. Representative data are shown (n=2). **g**, Point
493 mutations of charged residues in the γ CAL domain reduce or enhance M58_{ox} homo-condensate
494 formation, as measured by turbidity assay. Mutations of positive to negative (K/R to D; red
495 arrow) enhanced the interaction and mutations of negative to positive (D/E to K; blue arrow)
496 were inhibitory. Numbers of mutated residues are indicated. Protein concentration was 3.0 μ M at
497 50 mM KCl. Representative data are shown (n=2).

498

499 **Fig. 5 | Head-to-head association of M58 trimers.** **a**, The structural model of the dimer of
500 γ CAL trimers from the crystal structure was fitted into the barrel-shaped cryo-EM density map
501 reconstructed from the M58_{ox} condensate. The structure is shown in ribbon representation.
502 Additional densities at the corners probably represent SSUL modules that associate in a dynamic
503 manner and are incompletely resolved. The cryo-EM density map was low-pass filtered to 5 Å to
504 show these additional densities. **b**, Detail of the structural model from the crystal structure of
505 *Se7942* γ CAL(1-181) showing the two salt bridges connecting protomers of opposing trimers.
506 The side chains of residues R164 and D172 forming critical salt bridges are shown in stick
507 representation. **c**, Head-to-head association of M58 trimers is required for efficient M58_{ox} homo-
508 condensate formation. M58_{ox} and mutants M58_{ox}R164D and M58_{ox}D172K, disrupting the salt
509 bridges between M58 trimers, were analyzed by turbidity assay as in Fig. 4c at 3 μ M final
510 protein concentration (50 mM KCl). Data are averages \pm SD (n=3). **d**, Head-to-head association
511 of M58 trimers is not required for M58-CcaA condensate formation. M58 and mutants
512 M58R164D and M58D172 (0.5 μ M) were incubated with CcaA (0.5 μ M) under reducing and
513 oxidizing conditions and condensate formation analyzed by turbidity assay at 100 mM KCl. Data
514 are averages \pm SD (n=3).

515

516 **Fig. 6 | Trimeric M58 binds Rubisco with high affinity.** **a**, CC_{TRI}M35 mimics the high affinity
517 of M58 for Rubisco, independent of redox state. Condensate formation of 0.25 μ M Rubisco
518 (RbcL₈S₈) with increasing concentrations of M58_{red}, M58_{ox}, CC_{TRI}M35_{red} and CC_{TRI}M35_{ox} was
519 analyzed by turbidity at 50 mM KCl. Averages \pm SD are shown (n=3). **b-c**, M58-Rubisco
520 condensate formation is mediated by the SSUL modules of M58, not by the γ CAL domains.

521 Condensate formation of 0.25 μ M RbcL_{8S8} with M58 (0.25 μ M) or CC_{TRI}M35 (0.25 μ M) or
522 M35 (0.75 μ M) was analyzed by turbidity assay under reducing (**b**) and oxidizing conditions (**c**)
523 at 50 mM and 100 mM KCl. Averages \pm SD are shown (n=3).

524

525 **Fig. 7 | Co-assembly of Rubisco, M58, M35 and CcaA. a-b**, Condensate formation of Rubisco
526 with M58_{red} (**a**) and absence of interaction of Rubisco with CcaA (**b**), as analyzed by
527 fluorescence microscopy. M58, CcaA and Rubisco were N-terminally labeled with Alexa532,
528 Alexa405 and Alexa647, respectively (M58_{AF5}; CcaA_{AF4}; Rb_{CAF6}), and used as 1:10 mixtures
529 with unlabeled protein. Protein concentrations are indicated. Representative data are shown
530 (n=2). **c**, Co-assembly into condensates of Rubisco, M58_{red} and CcaA at the concentrations
531 indicated (100 mM KCl). Proteins were labeled as in (a) and (b). Representative data are shown
532 (n=2). **d**, Co-assembly into condensates of fluorescence labeled Rubisco, M35_{red}, and M58_{red}
533 with unlabeled CcaA at the concentrations indicated (100 mM KCl). M35 was N-terminally
534 labeled with Alexa405 (M35_{AF4}). Representative data are shown (n=2). **e**, Co-assembly of
535 fluorescence labeled Rubisco, M58_{red} and CcaA with unlabeled M35_{red} at the concentrations
536 indicated (100 mM KCl). Representative data are shown (n=2). **f**, Co-assembly of fluorescence
537 labeled Rubisco, CC_{TRI}M35_{red} and CcaA with unlabeled M35_{red}. CC_{TRI}M35 was N-terminally
538 labeled with Alexa532 (CC_{TRI}M35_{AF5}). CcaA does not enter the condensate in the absence of
539 M58. Representative data are shown (n=2). **g**, Mobility of M35 and M58 in condensates with
540 Rubisco. Fluorescence recovery after photobleaching (FRAP) experiments are shown for
541 condensates formed by unlabeled Rubisco with either labeled M35_{red} or M58_{red} (0.5 μ M
542 Rubisco/2.0 μ M M35_{red} or 0.25 μ M M58_{red}; 100 mM KCl). Pre-bleach fluorescence is set to 1.
543 Averages \pm SD from n=20 droplets. **h**, Mobility of Rubisco in condensates with M35 or M58.

544 FRAP experiments are shown for condensates formed by labeled Rubisco with either unlabeled
545 M35_{red} or M58_{red}. Concentrations as in (g). Pre-bleach fluorescence is set to 1. Averages \pm SD
546 from n=20 droplets. **i**, Mobility of M35 or M58 in condensates of Rubisco, CcaA, M58 and M35.
547 FRAP experiments are shown for condensates formed by the four proteins. Either M35 or M58
548 was labeled (0.5 μ M Rubisco/2 μ M M35_{red}/0.25 μ M M58_{red}/0.25 μ M CcaA; 100 mM KCl). Pre-
549 bleach fluorescence is set to 1. Averages \pm SD from n=20 droplets.

550

551 **Fig. 8 | Role of M58 as central organizer of the pre-carboxysome condensate. a**, Versatility
552 of M58-mediated protein interactions in multi-protein phase separation. (i) Interaction of trimeric
553 M58 (blue) with Rubisco (green) mediated by the SSUL modules of M58 (yellow) present at
554 high local concentration. (ii) Interaction of the trimeric γ CAL domains (blue) of M58 with the
555 C2 sequence of tetrameric CcaA (orange). The C2 peptide is shown in pink. (iii) M58 homo-
556 condensate formation mediated by intermolecular interactions between SSUL modules and
557 γ CAL domains and by head-to-head association of γ CAL trimers. Dotted lines indicate linker
558 sequences between protein domains interacting with nearby molecules. Effects of redox state of
559 the SSUL modules on the interactions shown in (i)-(iii) are indicated. **b**, Composite structural
560 model of head-to-head associated γ CAL trimers in surface representation in side view (left) and
561 end view (right). Bound C2₁₇ peptides of CcaA are shown in cartoon representation. The putative
562 binding surface of SSUL on the γ CAL domains is highlighted in yellow. One γ CAL domain in
563 the end view is transparent so as to show the bound C2₁₇ peptide.

564

565 **Methods**

566 **Strains**

567 *E. coli* DH5 α (ThermoFisher) cells were used for the amplification of plasmid DNA. Positive
568 clones were selected and cultivated in LB medium at 37 °C for 8 h. *E. coli* BL21 (DE3) (Agilent)
569 was used for recombinant protein expression (see Method Details).

570 The cyanobacterium *S. elongatus* PCC 7942 (*Se7942*) (Institut Pasteur Paris) was used to
571 obtain genomic DNA of *ccmM* and *ccaA*. *Se7942* was cultured in BG-11 medium at 30 °C and
572 50 rpm in continuous light.

573

574 **Plasmids**

575 The oligos used for amplification and generation of plasmids are listed in Supplementary Table
576 5.

577 **Genomic DNA**

578 *Se7942* was grown to high density and cells pelleted by centrifugation at 10,000 x g for 10 min.
579 The cell pellet was resuspended in 100 mL buffer (50 mM Tris-HCl pH 8.0/50 mM NaCl) and
580 cells lysed by 5 cycles of heating (3 min at 95°C) and snap freezing in liquid nitrogen. The lysate
581 was centrifuged (20,000 x g for 10 min) and 1 μ L of supernatant was used as template in PCR
582 reactions. The full-length *ccmM* gene was amplified using oligos #1/2 and the *ccaA* gene using
583 oligos #67/68 (Supplementary Table 5).

584 **Plasmids**

585 The pHUE vector for His6-ubiquitin (H₆Ub) fusion proteins^{49,50} was used to generate the
586 plasmids used in this study. Plasmids were assembled by PCR and using the Gibson assembly
587 cloning kit (NEB). The plasmids generated or used in this study are listed in Supplementary
588 Table 6.

589 pHUE-*SeM58* was generated by amplification of the full-length *ccmM* gene from
590 genomic DNA (*Se7942*) and subsequent cloning into the pHUE vector. The shorter constructs

591 containing pHUE-*Se* γ CAL-2S(1-429), pHUE-*Se* γ CAL-1S(1-313), pHUE-*Se* γ CAL(1-198) and
592 pHUE-*Se* γ CAL(1-181) were prepared by PCR from pHUE-*Se*M58 and cloned into the pHUE
593 vector. Point mutations in pHUE-*Se*M58 were introduced by QuikChange mutagenesis (Agilent)
594 to generate the following constructs: pHUE-*Se*M58-E17K; pHUE-*Se*M58-D21K; pHUE-*Se*M58-
595 D35K; pHUE-*Se*M58-R37D; pHUE-*Se*M58-R43D; pHUE-*Se*M58-K62D; pHUE-*Se*M58-E76K;
596 pHUE-*Se*M58-R79D; pHUE-*Se*M58-R95D; pHUE-*Se*M58-D112K; pHUE-*Se*M58-R126D;
597 pHUE-*Se*M58-R164D; pHUE-*Se*M58-D172K; pHUE-*Se*M58-E246K; pHUE-*Se*M58-D249K;
598 pHUE-*Se*M58-R251D; pHUE-*Se*M58-R252D; pHUE-*Se*M58-E286K; pHUE-*Se*M58-D294K;
599 pHUE-*Se*M58-R298D; pHUE-*Se*M58-E303K; pHUE-*Se*M58-R367D and pHUE-*Se*M58-R481D.

600 pHUE-*Se*M58-C4S (C261S/C279S/C377S/C395S) was generated by replacing the M35
601 fragment of pHUE-*Se*M58 with M35-C4S from the plasmid pHUE-
602 *Syn6301_ccmM_M35_C261S/C279S/C377S/C395S*¹⁵ by PCR and subsequent Gibson
603 assembly.

604 To generate the construct pHUE-CC_{TRI}M35, the trimeric coiled-coil sequence
605 GEIAAIKQEIAAIKKEIAAIKQEIAAIKQGS⁴³ was inserted into the plasmid pHUE-
606 *Syn6301_ccmM_M35*¹⁵ between the C-terminus of ubiquitin and the N-terminus of M35 by PCR
607 and subsequent Gibson assembly.

608 pHUE-*Se*CcaA was generated by amplification of the *ccaA* gene from genomic DNA
609 (*Se*7942) and subsequent cloning into the pHUE vector. pHUE-*Se*CcaA Δ C2 or pHUE-*Se*CcaA-
610 C2₁₇ were generated by cloning the first 257 residues (1-257) or the last 17 residues (256-272) of
611 *Se*CcaA from pHUE-*Se*CcaA into pHUE vector by PCR and subsequent Gibson assembly,
612 respectively.

613 pHUE-EGFP was generated by amplification of EGFP from pEF-*gfp* (Addgene) with
614 GSGGS at the C-terminus and subsequent cloning into pHUE. pHUE-EGFP_{C215} or pHUE-
615 EGFP_{C217} were generated by replacing residues 1-257 or 1-255 of *SeCcaA* in pHUE-*SeCcaA* with
616 EGFP-GSGGS by PCR and subsequent Gibson assembly, respectively.

617

618 **Protein expression and purification**

619 *SeRubisco*^{45,51}, *SeRbcL8*^{45,51}, *SeM35*¹⁵ and *SeRaf1*⁴⁷ proteins were expressed and purified as
620 previously described. Protein concentrations were determined spectrophotometrically at 280 nm.

621 ***M58 and mutants***

622 M58 was expressed and purified from *E. coli* BL21 (DE3) cells harboring the pHUE-*SeM58*
623 plasmid. Briefly, cells were grown in LB medium at 37 °C with shaking at 130 rpm until OD₆₀₀
624 0.4-0.5. Cells were equilibrated to 18 °C (~ 1 h) and protein expression induced by addition of
625 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 14 h/120 rpm. Cells were harvested
626 and incubated in buffer A (50 mM Tris-HCl pH 8.0/500 mM NaCl/5% glycerol) containing 20
627 mM imidazole/1 g L⁻¹ lysozyme/2.5 U mL⁻¹/*SmDNAse*/complete protease inhibitor cocktail
628 (Roche) for 1 h prior to lysis using EmulsiFlex C5 (Avestin, Inc). After high speed centrifugation
629 (40 000 x g/40 min/4 °C), the supernatant was loaded on to a gravity Ni-NTA metal affinity
630 column (Qiagen), equilibrated and washed with 10 column volumes (CV) buffer A/20 mM
631 imidazole. The bound protein was eluted with buffer A/300 mM imidazole. The H₆Ub moiety
632 was cleaved using H₆-Usp2 overnight at 4 °C⁵⁰. The cleaved protein was buffer exchanged on a
633 HiPrep 26/10 desalting column (GE) to buffer A. The protein eluate was then applied to a Ni-
634 NTA column for removal of H₆-Usp2, the cleaved H₆Ub moiety and any uncleaved protein. The
635 flow through was concentrated to ~3 mL and applied to a size-exclusion column (HiLoad 16/60
636 Superdex 200; GE) equilibrated in buffer B (50 mM Tris-HCl pH 8.0/500 mM KCl/10%

637 glycerol). The protein containing fractions were concentrated by ultrafiltration using Vivaspin
638 MWCO 30000 (GE), aliquoted and flash frozen in liquid N₂. All M58 point mutant proteins were
639 expressed in *E. coli* BL21 (DE3) cells harboring the respective plasmids and purified as
640 described for wild-type M58.

641 M58-4C-S, γ CAL-2S, γ CAL-1S and CC_{TRI}M35 were expressed in *E. coli* BL21 (DE3)
642 cells harboring pHUE-*Se*M58-C4S(C261S/C279S/C377S/C395S), pHUE-*Se* γ CAL-2S(1-429),
643 pHUE-*Se* γ CAL-1S(1-313) or pHUE-CC_{TRI}M35, respectively. These proteins were expressed and
644 purified as described for M58.

645 Reduced M58 proteins were generated by adding 5 mM DTT to the purified proteins
646 prior to use. To generate the oxidized M58 proteins, the purified proteins were incubated prior to
647 use on ice with 2 mM H₂O₂ for 30 min, followed by buffer exchange on a PD MiniTrap G-10
648 column (GE) to buffer B to remove unreacted H₂O₂.

649 ***γ CAL(1-198) and γ CAL(1-181)***

650 γ CAL(1-198) and γ CAL(1-181) were expressed in *E. coli* BL21 (DE3) cells harboring pHUE-
651 *Se* γ CAL(1-198) and pHUE-*Se* γ CAL(1-181), respectively, essentially as described for M58,
652 except that buffer A contained 150 mM NaCl and buffer B 150 mM KCl. After size-exclusion
653 chromatography (HiLoad 16/60 Superdex 200; GE), the purified proteins were concentrated by
654 ultrafiltration using Vivaspin MWCO 10000 (GE), aliquoted and flash frozen in liquid N₂.

655 ***CcaA and CcaA Δ C2***

656 CcaA and CcaA Δ C2 were expressed in *E. coli* BL21 (DE3) cells from the plasmid pHUE-
657 *Se*CcaA and pHUE-*Se*CcaA Δ C2, respectively, and purified as described for γ CAL. The proteins
658 were concentrated by ultrafiltration using Vivaspin MWCO 30000 (GE), aliquoted and flash
659 frozen in liquid N₂.

660 ***E_{GFP}, E_{GFP}C2₁₅ and E_{GFP}C2₁₇***

661 *E. coli* BL21 (DE3) cells harboring pHUE-EGFP, pHUE-EGFP_{C215} or pHUE-EGFP_{C217} were used
662 to express EGFP, EGFP_{C215} and EGFP_{C217}, respectively, and purified as described for γ CAL. After
663 size-exclusion chromatography (HiLoad 16/60 Superdex 75; GE), the purified proteins were
664 concentrated by ultrafiltration using Vivaspin MWCO 10000 (GE), aliquoted and flash frozen in
665 liquid N₂.

666 ***γ CAL and C2₁₇ peptide for crystallography***

667 For crystallization trials, purified γ CAL(1-198) and γ CAL(1-181) were buffer exchanged to
668 buffer C (20 mM Tris-HCl pH 8.0/150 mM NaCl) using a PD MiniTrap G-10 column (GE). The
669 C₂₁₇ peptide (residues 256-272) of *SeCcaA* was produced by expressing the vector pHUE-
670 *SeCcaA-C2₁₇* containing H₆Ub tagged C₂₁₇, and purification by Ni-NTA column as described for
671 γ CAL. The eluate from the Ni-NTA column was concentrated and applied onto a size-exclusion
672 column (HiLoad 16/60 Superdex 75; GE) equilibrated in buffer C. The protein containing
673 fractions were collected and cleaved by H₆-Usp2 overnight at 4 °C. The cleaved protein was
674 applied to a Ni-NTA column for removal of H₆-Usp2, the cleaved H₆Ub moiety and any
675 uncleaved protein. The C₂₁₇ peptide in the flow-through was mixed with γ CAL(1-198) or
676 γ CAL(1-181) at a molar ratio of 6:1, and incubated for 1h at 4 °C to generate the complex. The
677 respective complexes were purified by size-exclusion chromatography (HiLoad 16/60 Superdex
678 75 equilibrated in buffer C). The protein containing fractions were concentrated by ultrafiltration
679 using Vivaspin MWCO 3000 (GE), aliquoted and flash frozen in liquid N₂. The presence of the
680 C₂₁₇ peptide in the complex was confirmed by mass spectrometry.

681

682 **Turbidity assay**

683 Measurements were performed at 25 °C in buffer (50 mM Tris-HCl pH 8.0, 10 mM Mg(OAc)₂)
684 containing different concentrations of KCl and in the presence or absence of 5 mM DTT as

685 indicated in figure legends. Reactions (100 μ l) containing proteins as stated in figure legends
686 were rapidly mixed by vortexing, and absorbance at 340 nm was monitored over time on a Jasco
687 V-560 spectrophotometer set to 25 $^{\circ}$ C.

688

689 **Condensate formation analyzed by fluorescence microscopy**

690 Proteins to be analyzed for phase separation by microscopy were fluorescently labeled at the N-
691 terminus. Rubisco holoenzyme was labeled with Alexa Fluor 647 NHS ester (ThermoFisher)
692 (RbcLS_{AF6}) according to manufacturer's instructions (~4.6 dye molecules bound per Rubisco
693 holoenzyme). M58_{ox}, CC_{TRIM35ox} and γ CAL(1-198) were labeled with the fluorophore Alexa
694 Fluor 532 NHS ester (ThermoFisher) (M58_{AF5}; CC_{TRIM35AF5}; γ CAL_{AF5}) (~2.0, ~2.1 and ~1.4
695 dye molecules bound per M58_{ox} trimer, CC_{TRIM35ox} trimer and γ CAL(1-198) trimer,
696 respectively), while M35_{ox}, CcaA and CcaA Δ C2 were labeled with the fluorophore Alexa Fluor
697 405 NHS ester (ThermoFisher) (M35_{AF4}; CcaA_{AF4}; CcaA Δ C2_{AF4}) (~0.8, ~1.2 and ~3.0 dye
698 molecules bound per M35_{ox}, CcaA tetramer and CcaA Δ C2 tetramer, respectively). Labeled
699 protein was mixed with unlabeled protein at a ratio of 1:10. Reactions (20 μ L) in (50 mM Tris
700 pH 8.0/10 mM Mg(OAc)₂) with 50 mM or 100 mM KCl in the presence or absence of 5 mM
701 DTT, and proteins at the concentrations stated in figure legends were combined. After incubation
702 for 5 min at 25 $^{\circ}$ C, reactions were transferred to an uncoated chambered coverslip (μ -Slide
703 angiogenesis; Ibidi) followed by incubation for another 5 min before analysis. Samples were
704 illuminated with a Lumencor SPECTRA X Light Engine at 398 nm, 558 nm and 640 nm for
705 fluorescence imaging. Images were recorded by focusing on the bottom of the plate using a Leica
706 Thunder Widefield 2 microscope with Leica DFC9000 GTC camera and a HC PL APO 63x/1.47
707 oil objective.

708 The software Fiji⁵² was used for analysis of size distribution of droplets. In brief, after
709 preprocessing of images with background subtraction and Gaussian blur, the MaxEntropy
710 method was applied to determine the threshold of segmentation.

711

712 **Fluorescence recovery after photobleaching**

713 FRAP experiments were carried out with a Leica TCS SP8 AOBS Confocal Laser Scanning
714 Microscope (HCX PL APO 63×/1.2 water objective, PMT detector). Rubisco holoenzyme was
715 labeled with Alexa Fluor 532 NHS ester (ThermoFisher) (Rubisco_{AF5}) (~3.6 dye molecules
716 bound per Rubisco holoenzyme) and other proteins were labeled as described above. Reactions
717 (20 μL) in buffer D (50 mM Tris pH 8.0/100 mM KCl/10 mM Mg(OAc)₂) containing 5 mM
718 DTT, and proteins at the concentrations stated in figure legends were combined. After incubation
719 for 5 min at 25 °C, reactions were transferred to an uncoated chambered coverslip (μ-Slide
720 angiogenesis; Ibidi) followed by incubation for another 15 min before analysis. Images before
721 and 10 min after photobleaching were recorded in a single focal plane at a 5-s time interval.
722 Bleaching was performed with a bleach point model using a 405 nm diode laser at 2% intensity
723 or 532 nm argon laser at 100% intensity in one repeat with a dwell time of 100 ms. The software
724 Fiji was used for image analysis⁵².

725

726 **Size-exclusion chromatography coupled to multi-angle static light scattering (SEC- 727 MALS)**

728 Purified proteins at 2 mg mL⁻¹ were analyzed using static and dynamic light scattering by auto-
729 injection of the sample onto a SEC column (5 μm, 4.6x300 mm column, Wyatt Technology,
730 product # WTC-030N5) at a flow rate of 0.2 or 0.25 mL min⁻¹ in buffer 50 mM Tris pH 8.0/150
731 mM KCl/10 mM MgCl₂ or 50 mM Tris pH 8.0/500 mM KCl/5 mM DTT at 25 °C. The column

732 was in line with the following detectors: a variable UV absorbance detector set at 280 nm
733 (Agilent 1100 series), the DAWN EOS MALS detector (Wyatt Technology, 690 nm laser) and
734 the Optilab rEXTM refractive index detector (Wyatt Technology, 690 nm laser)⁵³. Molecular
735 masses were calculated using the ASTRA software (Wyatt Technology) with the dn/dc value set
736 to 0.185 mL g⁻¹. Bovine serum albumin (ThermoFisher) was used as the calibration standard.

737

738 **Isothermal titration calorimetry (ITC)**

739 ITC measurements were carried out on a ITC200 calorimeter (Microcal, GE) at 20° C. After
740 dialysis into buffer D, E_{GFP}C₂₁₇ (365 μM) was loaded into the syringe and titrated into the
741 sample cell containing γCAL(1-198) (14 μM). The reference cell contained buffer D. For each
742 titration point, 10 μL E_{GFP}C₂₁₇ was injected at time intervals of 3 min. Titration data were
743 analyzed using the software Origin 7.0 and fitted with a one-site binding model.

744

745 **Crystallization and data collection**

746 Crystals of γCAL(1-181) and γCAL(1-181)-C₂₁₇ were grown by the sitting-drop vapor diffusion
747 method at 20 °C. Drops containing 0.6 μL of a 1:1 mixture of 10 mg mL⁻¹ protein in buffer C and
748 precipitant (0.1 M HEPES pH 7.5/25% PEG-3350) were equilibrated against 100 μL precipitant.
749 For cryo-mounting, crystals were transferred into cryo-buffer (0.1 M HEPES pH 7.5/25% PEG-
750 3350/10% glycerol), and subsequently cryo-cooled by dipping into liquid N₂.

751 ***Crystallographic data collection, structure solution and refinement***

752 The diffraction data of γCAL(1-181) crystals and γCAL(1-181)-C₂₁₇ crystals were collected at
753 beamline ID23-2 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France,
754 and beamline X06SA at the Swiss Synchrotron Light Source (SLS) in Villigen, Switzerland,

755 respectively, and processed with autoPROC (Global Phasing)⁵⁴ using XDS⁵⁵, POINTLESS⁵⁶ and
756 AIMLESS⁵⁷.

757 The structure of γ CAL(1-181) was solved to 1.67 Å resolution by molecular replacement
758 using the program MOLREP⁵⁸ with the γ CAL(1-209) domain of *Thermosynechococcus*
759 *elongatus* BP-1 (PDB: 3KWC) as a search template. The asymmetric unit contained one
760 γ CAL(1-181) protomer with residues 1–15 disordered. The model was edited manually using
761 Coot⁵⁹ as implemented in the CCP4i graphical user interface⁶⁰. REFMAC5 was used for model
762 refinement⁶¹. The model of γ CAL(1-181) contains 127 ordered water molecules, a presumably
763 ordered Cl⁻ ion and a Ni²⁺ atom from the Ni-NTA metal affinity column. The bound Ni²⁺ atom
764 was identified by X-ray fluorescence scanning.

765 The structure of γ CAL(1-181)-C2₁₇ was solved to 1.63 Å resolution by molecular
766 replacement using the γ CAL(1-181) model and refined as described above. The asymmetric unit
767 contained one γ CAL(1-181) protomer with residues 1-15 disordered, one bound C2₁₇ peptide
768 with residues 256-270 resolved (Supplementary Fig. 3g), one presumably ordered Cl⁻ ion, one
769 Ni²⁺ ion and 103 ordered water molecules.

770

771 **Structure analysis**

772 The quality of the structural models was analyzed with the program Molprobity⁶². The final
773 models of γ CAL(1-181) and γ CAL(1-181)-C2₁₇ exhibit reasonable stereochemistry with 98.2%
774 and 98.3% of residues, respectively, in the favored regions of the Ramachandran plot.

775 Coordinates were aligned with Lsqkab and Lsqman⁶³. Molecular interfaces were analyzed with
776 PISA⁴² and figures were created with PyMol (<http://www.pymol.org/>).

777

778 **Cryo-electron microscopy and reconstruction**

779 ***Sample preparation and data collection***

780 All cryo-grids were prepared with a Vitrobot Mark 4 (FEI). A volume of 3 μL of the sample was
781 applied to a glow-discharged grid (Quantifoil R2/1 300 mesh) at 25°C and 90% humidity, then
782 semi-automatically blotted and plunge-frozen into liquid ethane.

783 For the analysis of M58 head-to-head complexes 6 μM M58_{ox} was incubated in buffer
784 (50 mM Tris-HCL pH 8.0, 10 mM Mg(OAc)₂, 50 mM KCl) at 25 °C for 10 min and cryo-grids
785 prepared as described above. The cryo-grids were initially screened on a Glacios transmission
786 electron microscope (ThermoFisher), equipped with a K2 summit direct electron detector
787 (Gatan), operated at 200 keV. Selected grids were transferred to a Titan Krios 300 kV TEM
788 (FEI) equipped with GIF Quantum Energy Filters (Gatan), and a K3 direct detector (Gatan).
789 1,836 movies were automatically collected by SerialEM⁶⁴ using a pixel size of 0.4114 Å. The
790 total exposure time of 1.2 s was divided into 20 frames with an accumulated dose of 60 electrons
791 per Å² and a defocus range of -0.65 μm to -2.15 μm .

792 The M58_{red}-Rubisco and M58_{red}-RbcL₈ complexes were prepared by mixing Rubisco (6
793 μM) and M58_{red} (8 μM) or RbcL₈ (6.25 μM) and M58_{red} (16.7 μM) in buffer (50 mM Tris-HCL
794 pH 8.0, 10 mM Mg(OAc)₂, 50 mM KCl, 5 mM DTT), respectively, for 10 min at 25 °C and
795 cryo-grids prepared as described above. The cryo-grids were screened on a Glacios transmission
796 electron microscope (ThermoFisher), equipped with a K2 summit direct electron detector
797 (Gatan), operated at 200 keV. The selected grid on stage was used for data collection directly
798 with K2 summit. Exposure times of 12 s were divided into 40 frames with an accumulated dose
799 of 45 electrons per Å². 976 and 1,027 movies were automatically collected for M58_{red}-Rubisco
800 and M58_{red}-RbcL₈, respectively, by SerialEM⁶⁴ with a pixel size of 1.885 Å and a defocus range
801 of -0.7 μm to -4.5 μm .

802 **Image processing**

803 **(i) M58**

804 On-the-fly processing during data collection was performed with MotionCor2⁶⁵ and CTFFIND-
805 4.1⁶⁶, as implemented in the Focus software⁶⁷. Only micrographs with good signal quality, with
806 an estimated maximum resolution below 5 Å, were kept for further data processing with
807 RELION 3.1⁶⁸ (Supplementary Fig. 6). A total of 349,391 particles were auto-picked by
808 Gautomatch (<http://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch>) and extracted at a pixel size
809 of 1.65 Å (four-fold binned). The first round of 2D-classification was performed to exclude ice
810 contaminants and classes with no structural features. The side view classes suggested a head-to-
811 head stack of two γ CAL domains. Thus, the crystallographic model of the γ CAL dimer-of-
812 trimers (PDB: 7O4Z) was converted into an EM density map in mrc format with a low-pass filter
813 to 15 Å, which was used as a reference for 3D classification and refinement. The selected
814 particles were subjected to one round of refinement and new particles were extracted with
815 refined coordinate (re-center) at a pixel size of 0.82 Å. 3D-classification resulted in one major
816 class, containing 128,330 particles (Supplementary Fig. 6c). These particles were subjected to
817 CTF refinement and polishing to generate the final map at 3.57 Å resolution, determined by
818 gold-standard Fourier shell correlation (FSC) with a cutoff at 0.143 (Supplementary Fig. 6d).
819 The particle distribution plot suggested a lack of information in side view (Supplementary Fig.
820 6b). As a result, the side view of the reconstruction was stretched. Based on the information from
821 the well-resolved end view (Supplementary Fig. 6e), we docked the crystallographic model of
822 the γ CAL trimers into the EM density map using Chimera⁶⁹. The EM density map is deposited to
823 EMDB under the accession code EMD-12730.

824 **(ii) M58_{red}-Rubisco**

825 The raw movies of the dataset for the M58_{red}-Rubisco complex were first processed with
826 MotionCor2⁶⁵ with dose-weighting. CTFFIND-4.1⁶⁶ estimated the CTF parameters for each
827 micrograph. A total of 620,012 particles were picked by Gautomatch ([http://www.mrc-](http://www.mrc-imb.cam.ac.uk/kzhang/Gautomatch)
828 [imb.cam.ac.uk/kzhang/Gautomatch](http://www.mrc-imb.cam.ac.uk/kzhang/Gautomatch)) (Supplementary Fig. 7e-h). Particles were first extracted at a
829 pixel size of 7.54 Å (four-fold binned). One round of 2D-classification resulted in 507,604 clean
830 particles, with ice contaminants and classes with no structural features excluded (Supplementary
831 Fig. 7f,g). These particles were refined with a low resolution reference converted from the
832 crystal structure coordinates of the Rubisco holoenzyme (PDB: 1RBL), and extracted at a pixel
833 size of 3.77 Å. A single round of 3D classification with *D4* symmetry resulted in four classes
834 with no major differences. Thus, particles from all four classes were subjected to further
835 analysis. These particles were again extracted with full resolution at a pixel size of 1.885 Å. We
836 next followed the same symmetry-expansion procedure as previously published¹⁵, i.e. particles
837 were first aligned with *D4* symmetry, to account for multiple SSUL modules bound per Rubisco.
838 Each asymmetric unit was processed as an individual particle, which is achieved by the
839 symmetry-expanding command, `reion_particle_symmetry_expand`, and particle subtraction was
840 done based on a mask covering two RbcL, two RbcS and two SSUL. A focused classification
841 with a SSUL mask resulted in one class of particles with detailed SSUL features. A total of
842 698,820 particles from this class were selected and subjected to final local refinement, and the
843 post-processing job pushed the resolution of the EM density map to ~4 Å determined by gold-
844 standard FSC curve at 0.143 cutoff (Supplementary Fig. 7g,h). Two EM density maps were
845 deposited to EMDB under the accession code EMD-12731, one without sharpening applied and
846 the other map sharpened with DeepEMhancer (<https://doi.org/10.1101/2020.06.12.148296>).
847 **(iii) M58_{red}-RbcL₈**

848 The raw movies of the dataset for the M58_{red}-RbcL₈ complex were first processed with
849 MotionCor2⁶⁵ with dose-weighting. CTFFIND-4.1⁶⁶ and the CTF parameters for each
850 micrograph were estimated. 258,285 Particles were picked by Gautomatch ([http://www.mrc-](http://www.mrc-imb.cam.ac.uk/kzhang/Gautomatch)
851 [imb.cam.ac.uk/kzhang/Gautomatch](http://www.mrc-imb.cam.ac.uk/kzhang/Gautomatch)). Particles were first extracted at a pixel size of 7.54 Å (four-
852 fold binned). One round of 2D-classification excluded ice contaminations and classes with no
853 structural features, and resulted in 136,505 clean particles (Supplementary Fig. 8d-f). These
854 particles were refined with a low resolution reference converted from the RbcL₈ crystal structure
855 coordinates (PDB: 1RBL with RbcS subunits deleted), and extracted at a pixel size of 3.77 Å. A
856 single round of 3D classification identified a major class with detailed RbcL₈ features (92,899
857 particles) (Supplementary Fig. 8f). These particles were again extracted with full resolution at a
858 pixel size of 1.885 Å. We next followed the same symmetry-expansion as used for the image
859 processing for M58_{red}-Rubisco above. A focused classification with a SSUL mask resulted in a
860 single class (193,877 particles) with detailed SSUL features (Supplementary Fig. 8f). This class
861 was subjected to final local refinement yielding a map with ~8 Å resolution without post-
862 processing (Supplementary Fig. 8f). To exclude the bias due to focused classification on SSUL,
863 we performed another round of focused classification with a mask covering one RbcS subunit.
864 The classification resulted in one dominant class containing no EM density in the region where
865 RbcS is bound. The previously published model of RbcL₂-RbcS₂-SSUL (PDB: 6HBC)¹⁵ was
866 fitted into the experimental EM density map using Chimera⁶⁹. Two maps were deposited to
867 EMDDB under the accession code EMD-12732, one without sharpening applied and the other map
868 sharpened with DeepEMhancer (<https://doi.org/10.1101/2020.06.12.148296>). The resolution was
869 determined by gold-standard FSC curve at 0.143 cutoff (Supplementary Fig. 8g).

870

871 **Sequence alignment**

872 Conservation of protein sequences was analyzed using the ConSurf web server⁷⁰. Searching of
873 sequences homologous to *SeSSUL* (219-311) or *SeCcmM* (1-539) was performed against the
874 UniProt database. HMMER, 3 and 0.0001 were set for homolog search algorithm, number of
875 iterations and E-value cutoff, respectively. Multiple sequence alignment containing 500 SSUL or
876 150 γ CAL homologous sequences was built using MAFFT and submitted to WebLogo server⁷¹
877 to create the sequence logos.

878

879 **Data availability**

880 The crystallographic structure factors and models for *Se γ CAL*(1-181) and *Se γ CAL*(1-181)-C2₁₇
881 complex have been deposited to the PDB database under accession codes 7O4Z and 7O54,
882 respectively. The local electron density maps for *SeM58_{ox}*, *SeM58_{red}-SeRubisco* and *SeM58_{red}-*
883 *SeRbcL₈* are deposited under EMDB accession codes EMD-12730, EMD-12731 and EMD-
884 12732, respectively. Additional data supporting the findings of this manuscript are available
885 from the corresponding author on reasonable request.

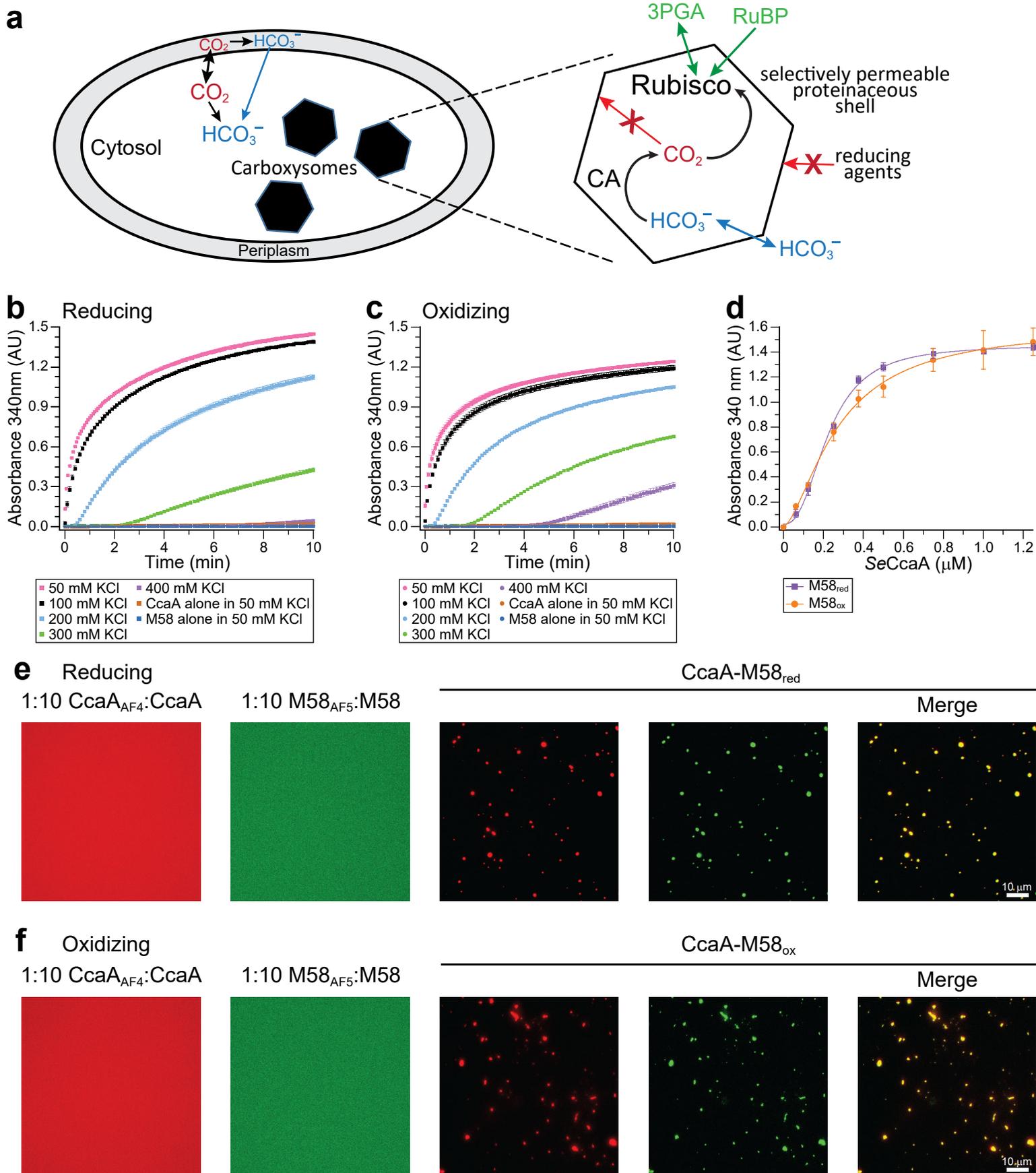
886 **References**

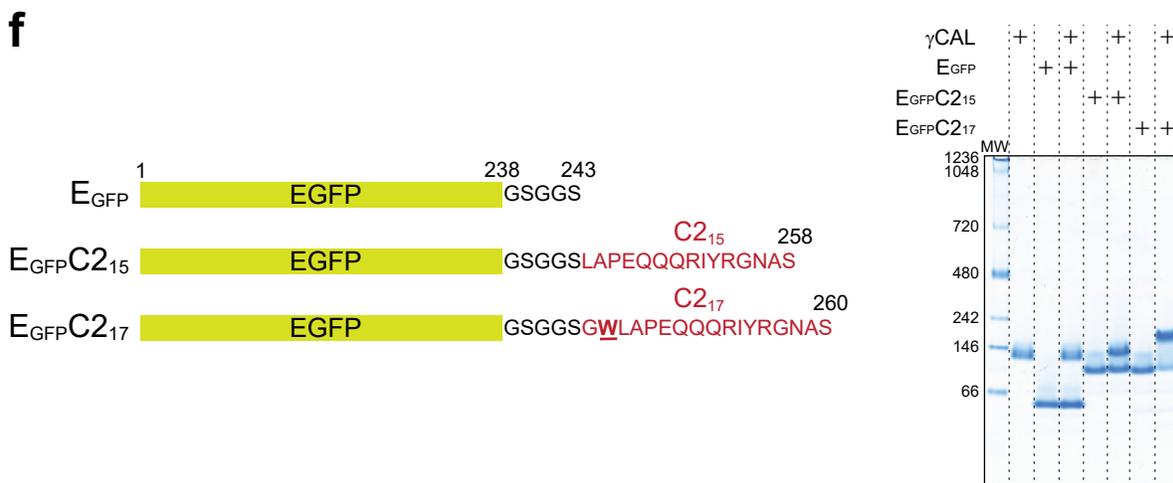
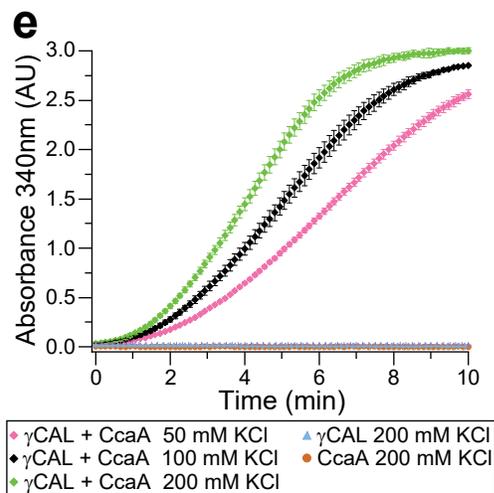
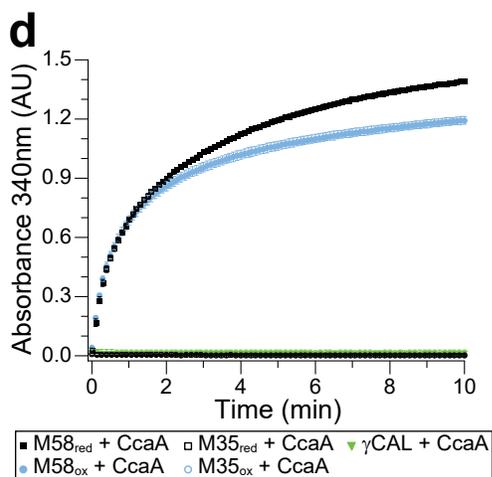
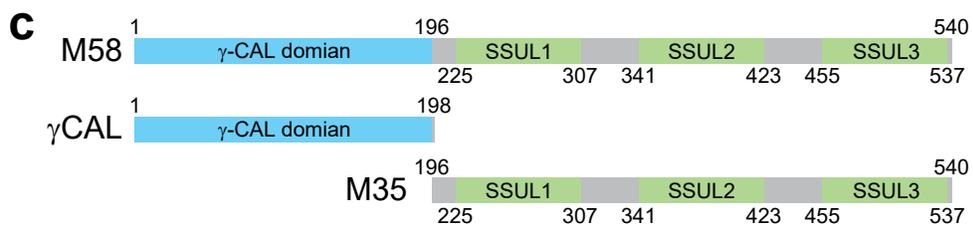
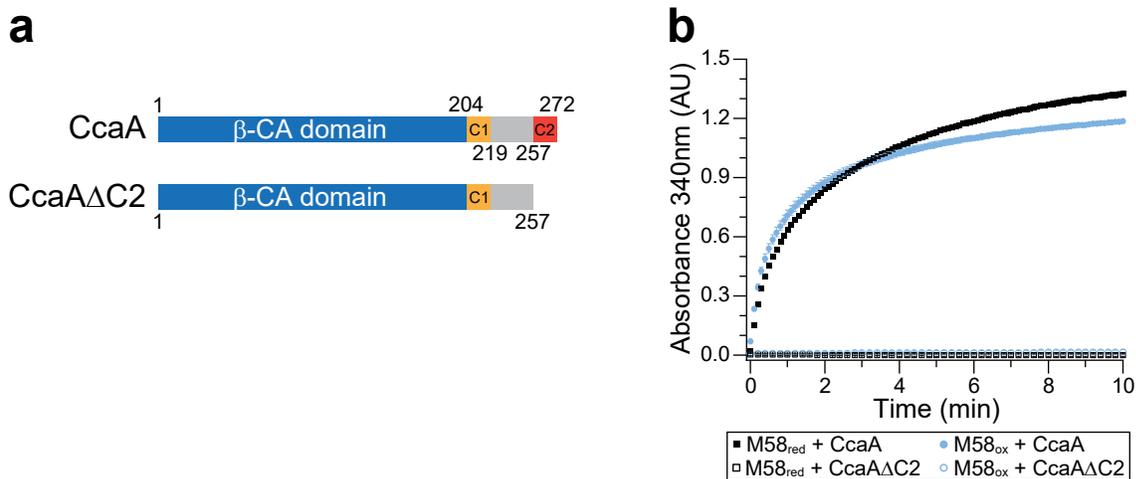
- 887 1 Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F. & Sutter, M. Bacterial microcompartments.
888 *Nat. Rev. Microbiol.* 16, 277-290 (2018).
- 889 2 Greening, C. & Lithgow, T. Formation and function of bacterial organelles. *Nat. Rev. Microbiol.*
890 (2020).
- 891 3 Espie, G. S. & Kimber, M. S. Carboxysomes: cyanobacterial Rubisco comes in small packages.
892 *Photosynthe. Res.* 109, 7-20 (2011).
- 893 4 Turmo, A., Gonzalez-Esquer, C. R. & Kerfeld, C. A. Carboxysomes: metabolic modules for CO₂
894 fixation. *FEMS Microbiol. Lett.* 364 (2017).
- 895 5 Hanson, M. R., Lin, M. T., Carmo-Silva, A. E. & Parry, M. A. Towards engineering carboxysomes
896 into C3 plants. *Plant J.* 87, 38-50 (2016).
- 897 6 Long, B. M., Rae, B. D., Rolland, V., Forster, B. & Price, G. D. Cyanobacterial CO₂-concentrating
898 mechanism components: function and prospects for plant metabolic engineering. *Curr. Opin. Plant*
899 *Biol.* 31, 1-8 (2016).
- 900 7 Rae, B. D. *et al.* Progress and challenges of engineering a biophysical CO₂-concentrating mechanism
901 into higher plants. *J. Exp. Bot.* 68, 3717-3737 (2017).
- 902 8 Giessen, T. W. & Silver, P. A. Engineering carbon fixation with artificial protein organelles. *Curr.*
903 *Opin. Biotechnol.* 46, 42-50 (2017).

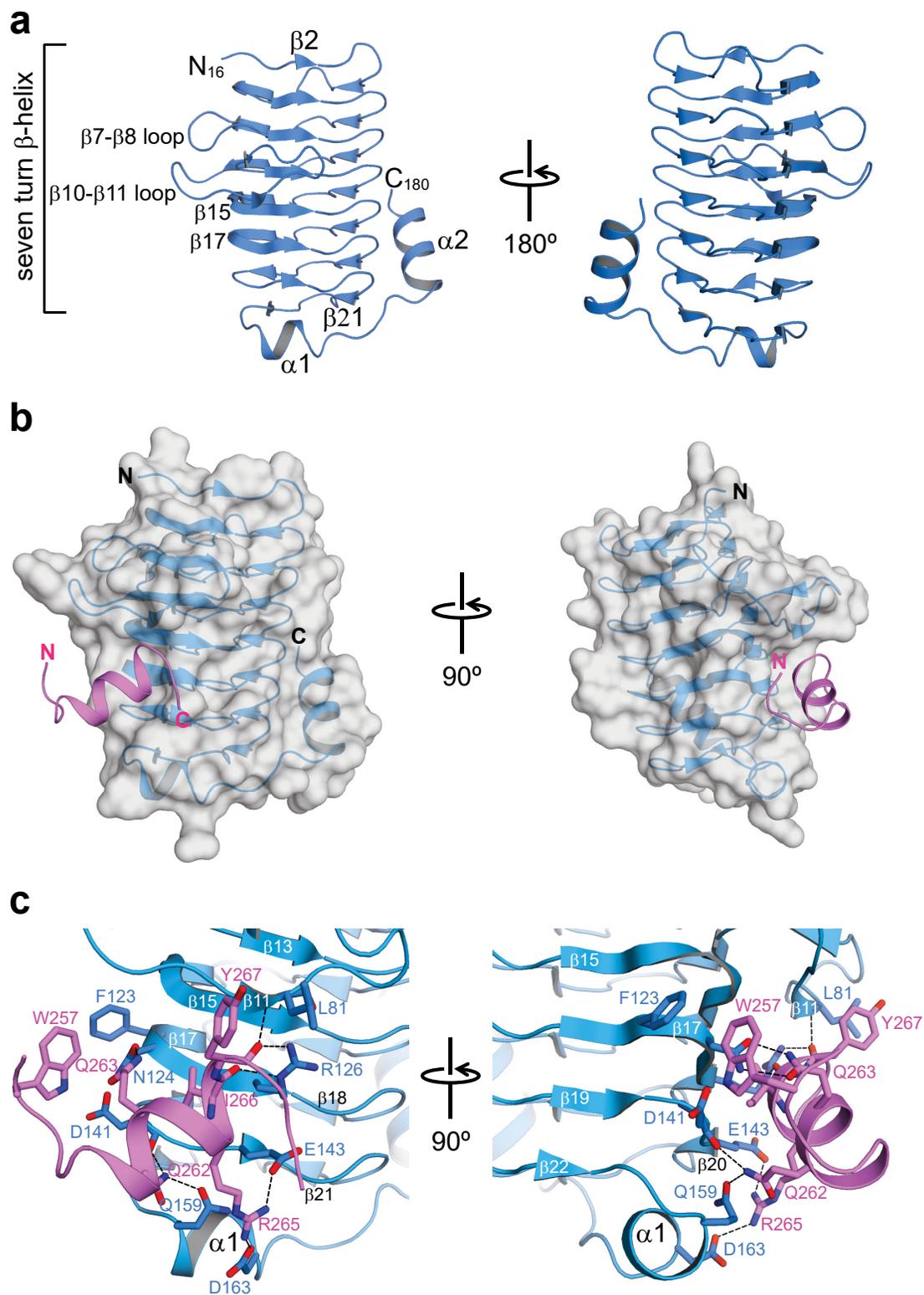
- 904 9 Kubis, A. & Bar-Even, A. Synthetic biology approaches for improving photosynthesis. *J. Exp. Bot.*
905 70, 1425-1433 (2019).
- 906 10 Kirst, H. & Kerfeld, C. A. Bacterial microcompartments: catalysis-enhancing metabolic modules for
907 next generation metabolic and biomedical engineering. *BMC Biol.* 17, 79 (2019).
- 908 11 Wunder, T. & Mueller-Cajar, O. Biomolecular condensates in photosynthesis and metabolism. *Curr.*
909 *Opin. Plant Biol.* 58, 1-7 (2020).
- 910 12 Hennacy, J. H. & Jonikas, M. C. Prospects for engineering biophysical CO₂ concentrating
911 mechanisms into land plants to enhance yields. *Annu. Rev. Plant Biol.* 71, 461-485 (2020).
- 912 13 Zhu, X. G., Ort, D. R., Parry, M. A. J. & von Caemmerer, S. A wish list for synthetic biology in
913 photosynthesis research. *J. Exp. Bot.* 71, 2219-2225 (2020).
- 914 14 Borden, J. S. & Savage, D. F. New discoveries expand possibilities for carboxysome engineering.
915 *Curr. Opin. Microbiol.* 61, 58-66 (2021).
- 916 15 Wang, H. *et al.* Rubisco condensate formation by CcmM in beta-carboxysome biogenesis. *Nature*
917 566, 131-135 (2019).
- 918 16 Oltrogge, L. M. *et al.* Multivalent interactions between CsoS2 and Rubisco mediate α -carboxysome
919 formation. *Nat. Struct. Mol. Biol.* 27, 281-287 (2020).
- 920 17 Lechno-Yossef, S. *et al.* Cyanobacterial carboxysomes contain an unique Rubisco-activase-like
921 protein. *New Phytol.* 225, 793-806 (2020).
- 922 18 Flecken, M. *et al.* Dual functions of a Rubisco activase in metabolic repair and recruitment to
923 carboxysomes. *Cell* 183, 457-473.e20 (2020).
- 924 19 Kaplan, A. On the cradle of CCM research: discovery, development, and challenges ahead. *J. Exp.*
925 *Bot.* 68, 3785-3796 (2017).
- 926 20 Kimber, M. S. in *Carbonic Anhydrase: Mechanism, Regulation, Links to Disease, and Industrial*
927 *Applications* (eds Susan C. Frost & Robert McKenna) 89-103 (Springer Netherlands, 2014).
- 928 21 Dou, Z. *et al.* CO₂ fixation kinetics of *Halothiobacillus neapolitanus* mutant carboxysomes lacking
929 carbonic anhydrase suggest the shell acts as a diffusional barrier for CO₂. *J. Biol. Chem.* 283, 10377-
930 10384 (2008).
- 931 22 Kerfeld, C. A. & Melnicki, M. R. Assembly, function and evolution of cyanobacterial carboxysomes.
932 *Curr. Opin. Plant Biol.* 31, 66-75 (2016).
- 933 23 Hayer-Hartl, M. & Hartl, F. U. Chaperone machineries of Rubisco - The most abundant enzyme.
934 *Trends Biochem. Sci.* 45, 748-763 (2020).
- 935 24 Rae, B. D., Long, B. M., Badger, M. R. & Price, G. D. Functions, compositions, and evolution of the
936 two types of carboxysomes: polyhedral microcompartments that facilitate CO₂ fixation in
937 cyanobacteria and some proteobacteria. *Microbiol. Mol. Biol. Rev.* 77, 357-379 (2013).
- 938 25 Long, B. M., Badger, M. R., Whitney, S. M. & Price, G. D. Analysis of carboxysomes from
939 *Synechococcus* PCC7942 reveals multiple Rubisco complexes with carboxysomal proteins CcmM
940 and CcaA. *J. Biol. Chem.* 282, 29323-29335 (2007).
- 941 26 Cot, S. S., So, A. K. & Espie, G. S. A multiprotein bicarbonate dehydration complex essential to
942 carboxysome function in cyanobacteria. *J. Bacteriol.* 190, 936-945 (2008).
- 943 27 Kinney, J. N., Salmeen, A., Cai, F. & Kerfeld, C. A. Elucidating essential role of conserved
944 carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly. *J.*
945 *Biol. Chem.* 287, 17729-17736 (2012).
- 946 28 Long, B. M., Tucker, L., Badger, M. R. & Price, G. D. Functional cyanobacterial beta-carboxysomes
947 have an absolute requirement for both long and short forms of the CcmM protein. *Plant Physiol.* 153,
948 285-293 (2010).
- 949 29 DiMario, R. J., Clayton, H., Mukherjee, A., Ludwig, M. & Moroney, J. V. Plant carbonic
950 anhydrases: Structures, locations, evolution, and physiological roles. *Mol. Plant* 10, 30-46 (2017).
- 951 30 Pena, K. L., Castel, S. E., de Araujo, C., Espie, G. S. & Kimber, M. S. Structural basis of the
952 oxidative activation of the carboxysomal gamma-carbonic anhydrase, CcmM. *Proc. Natl. Acad. Sci.*
953 *U.S.A.* 107, 2455-2460 (2010).

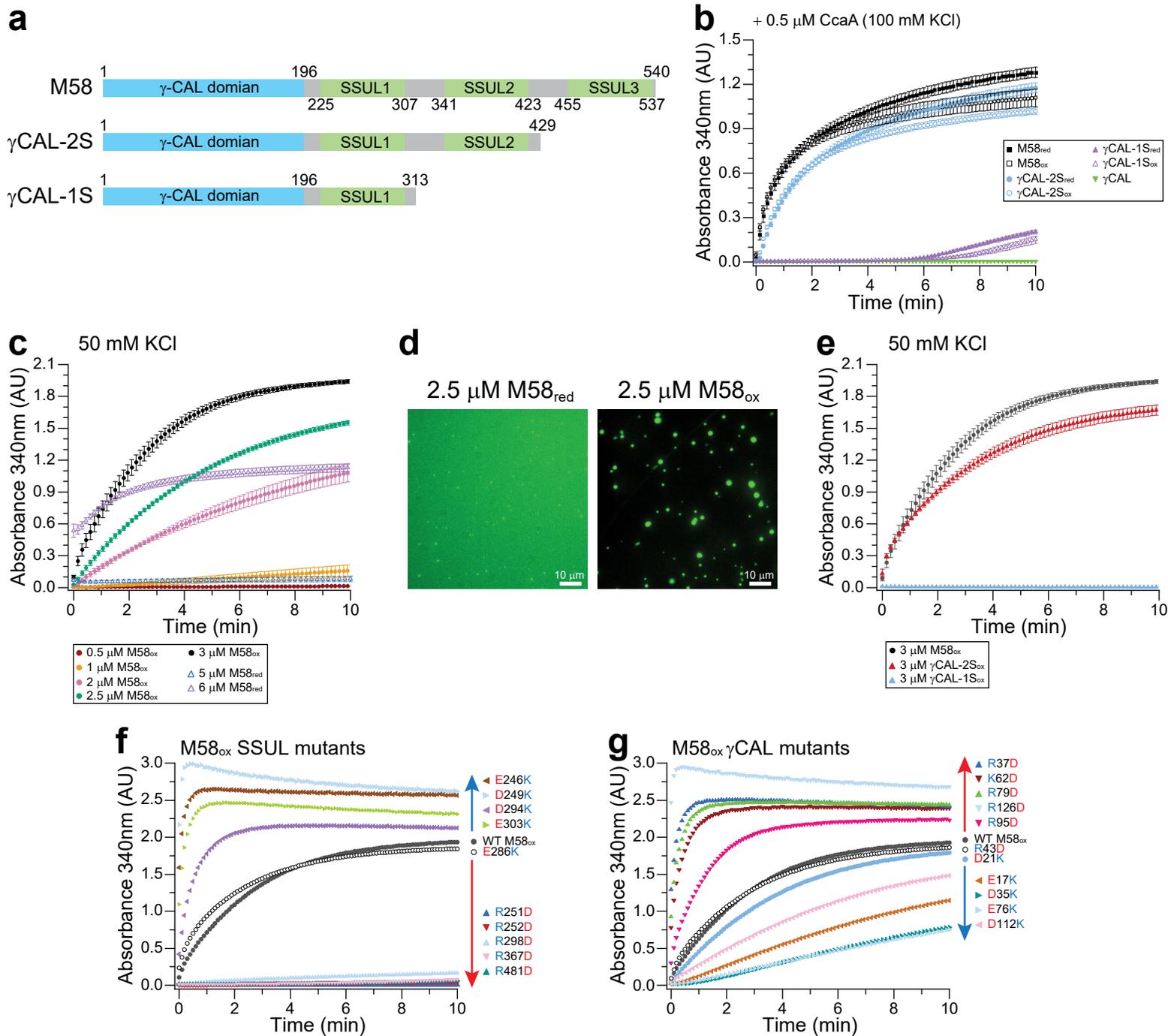
- 954 31 Kisker, C., Schindelin, H., Alber, B. E., Ferry, J. G. & Rees, D. C. A left-hand beta-helix revealed by
955 the crystal structure of a carbonic anhydrase from the archaeon *Methanosarcina thermophila*. *EMBO*
956 *J.* 15, 2323-2330 (1996).
- 957 32 So, A. K. & Espie, G. S. Cloning, characterization and expression of carbonic anhydrase from the
958 cyanobacterium *Synechocystis* PCC6803. *Plant Mol. Biol.* 37, 205-215 (1998).
- 959 33 McGurn, L. D. *et al.* The structure, kinetics and interactions of the beta-carboxysomal beta-carbonic
960 anhydrase, CcaA. *Biochem. J.* 473, 4559-4572 (2016).
- 961 34 Kimber, M. S., Coleman, J. R. & Pai, E. F. Beta-carbonic anhydrase from *Pisum sativum*:
962 crystallization and preliminary X-ray analysis. *Acta Crystallogr. D Biol. Crystallogr.* 56, 927-929
963 (2000).
- 964 35 Aggarwal, M., Chua, T. K., Pinard, M. A., Szebenyi, D. M. & McKenna, R. Carbon dioxide
965 "trapped" in a β -carbonic anhydrase. *Biochemistry* 54, 6631-6638 (2015).
- 966 36 Schlicker, C. *et al.* Structure and inhibition of the CO₂-sensing carbonic anhydrase Can2 from the
967 pathogenic fungus *Cryptococcus neoformans*. *J. Mol. Biol.* 385, 1207-1220 (2009).
- 968 37 Cronk, J. D., Endrizzi, J. A., Cronk, M. R., O'Neill, J. W. & Zhang, K. Y. J. Crystal structure of *E.*
969 *coli* β -carbonic anhydrase, an enzyme with an unusual pH-dependent activity. *Protein Sci.* 10, 911-
970 922 (2001).
- 971 38 So, A. K., John-McKay, M. & Espie, G. S. Characterization of a mutant lacking carboxysomal
972 carbonic anhydrase from the cyanobacterium *Synechocystis* PCC6803. *Planta* 214, 456-467 (2002).
- 973 39 Nishimura, T., Yamaguchi, O., Takatani, N., Maeda, S. & Omata, T. In vitro and in vivo analyses of
974 the role of the carboxysomal β -type carbonic anhydrase of the cyanobacterium *Synechococcus*
975 *elongatus* in carboxylation of ribulose-1,5-bisphosphate. *Photosynth. Res.* 121, 151-157 (2014).
- 976 40 Price, G. D., Coleman, J. R. & Badger, M. R. Association of carbonic anhydrase activity with
977 carboxysomes isolated from the cyanobacterium *Synechococcus* PCC7942. *Plant Physiol.* 100, 784-
978 793 (1992).
- 979 41 So, A. K. C., Cot, S. S. W. & Espie, G. S. Characterization of the C-terminal extension of
980 carboxysomal carbonic anhydrase from *Synechocystis* sp. PCC6803. *Funct. Plant Biol.* 29, 183-194
981 (2002).
- 982 42 Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol.*
983 *Biol.* 372, 774-797 (2007).
- 984 43 Fletcher, J. M. *et al.* A basis set of de novo coiled-coil peptide oligomers for rational protein design
985 and synthetic biology. *ACS Synth. Biol.* 1, 240-250 (2012).
- 986 44 Rohnke, B. A., Kerfeld, C. A. & Montgomery, B. L. Binding options for the small subunit-like
987 domain of cyanobacteria to Rubisco. *Front. Microbiol.* 11, 187 (2020).
- 988 45 Saschenbrecker, S. *et al.* Structure and function of RbcX, an assembly chaperone for hexadecameric
989 Rubisco. *Cell* 129, 1189-1200 (2007).
- 990 46 Bracher, A., Starling-Windhof, A., Hartl, F. U. & Hayer-Hartl, M. Crystal structure of a chaperone-
991 bound assembly intermediate of form I Rubisco. *Nat. Struct. Mol. Biol.* 18, 875-880 (2011).
- 992 47 Hauser, T. *et al.* Structure and mechanism of the Rubisco-assembly chaperone Raf1. *Nat. Struct.*
993 *Mol. Biol.* 22, 720-728 (2015).
- 994 48 Cameron, J. C., Wilson, S. C., Bernstein, S. L. & Kerfeld, C. A. Biogenesis of a bacterial organelle:
995 the carboxysome assembly pathway. *Cell* 155, 1131-1140 (2013).
- 996 49 Catanzariti, A. M., Soboleva, T. A., Jans, D. A., Board, P. G. & Baker, R. T. An efficient system for
997 high-level expression and easy purification of authentic recombinant proteins. *Protein Sci.* 13, 1331-
998 1339 (2004).
- 999 50 Baker, R. T. *et al.* in *Methods Enzymol.* Vol. 398 Ubiquitin and Protein Degradation, Part A 540-554
1000 (Academic Press, 2005).
- 1001 51 Liu, C. *et al.* Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature*
1002 463, 197-202 (2010).
- 1003 52 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676-
1004 682 (2012).

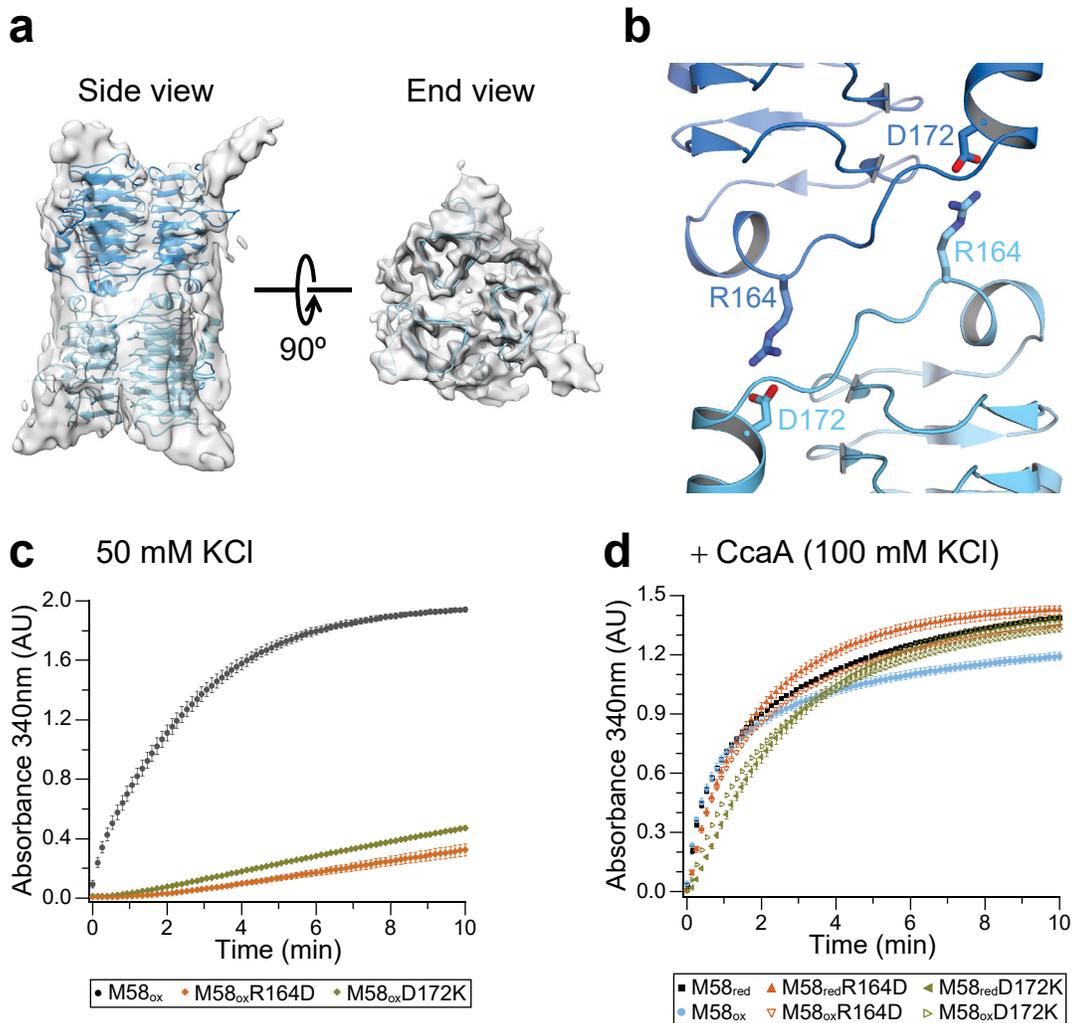
- 1005 53 Wyatt, P. J. Light scattering and the absolute characterization of macromolecules. *Anal. Chim. Acta*
1006 272, 1-40 (1993).
- 1007 54 Vonrhein, C. *et al.* Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D*
1008 *Biol. Crystallogr.* 67, 293-302 (2011).
- 1009 55 Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125-132 (2010).
- 1010 56 Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* 62, 72-82
1011 (2006).
- 1012 57 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta*
1013 *Crystallogr. D Biol. Crystallogr.* 69, 1204-1214 (2013).
- 1014 58 Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D Biol.*
1015 *Crystallogr.* 66, 22-25 (2010).
- 1016 59 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D*
1017 *Biol. Crystallogr.* 60, 2126-2132 (2004).
- 1018 60 Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. A graphical user interface to the CCP4
1019 program suite. *Acta Crystallogr. D Biol. Crystallogr.* 59, 1131-1137 (2003).
- 1020 61 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta*
1021 *Crystallogr. D Biol. Crystallogr.* 67, 355-367 (2011).
- 1022 62 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography.
1023 *Acta Crystallogr. D Biol. Crystallogr.* 66, 12-21 (2010).
- 1024 63 Kleywegt, G. T. & Jones, T. A. A super position. *CCP4/ESF-EACBM Newsletter on Protein*
1025 *Crystallogr.* 31, 9-14 (1994).
- 1026 64 Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen
1027 movements. *J. Struct. Biol.* 152, 36-51 (2005).
- 1028 65 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-
1029 electron microscopy. *Nat. Methods* 14, 331-332 (2017).
- 1030 66 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron
1031 micrographs. *J. Struct. Biol.* 192, 216-221 (2015).
- 1032 67 Biyani, N. *et al.* Focus: The interface between data collection and data processing in cryo-EM. *J.*
1033 *Struct. Biol.* 198, 124-133 (2017).
- 1034 68 Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure
1035 determination. *J. Struct. Biol.* 180, 519-530 (2012).
- 1036 69 Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis.
1037 25, 1605-1612 (2004).
- 1038 70 Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary
1039 conservation in macromolecules. *Nucleic Acids Res.* 44, W344-350 (2016).
- 1040 71 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator.
1041 *Genome Res.* 14, 1188-1190 (2004).
- 1042 72 Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for
1043 protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845-858 (2015).

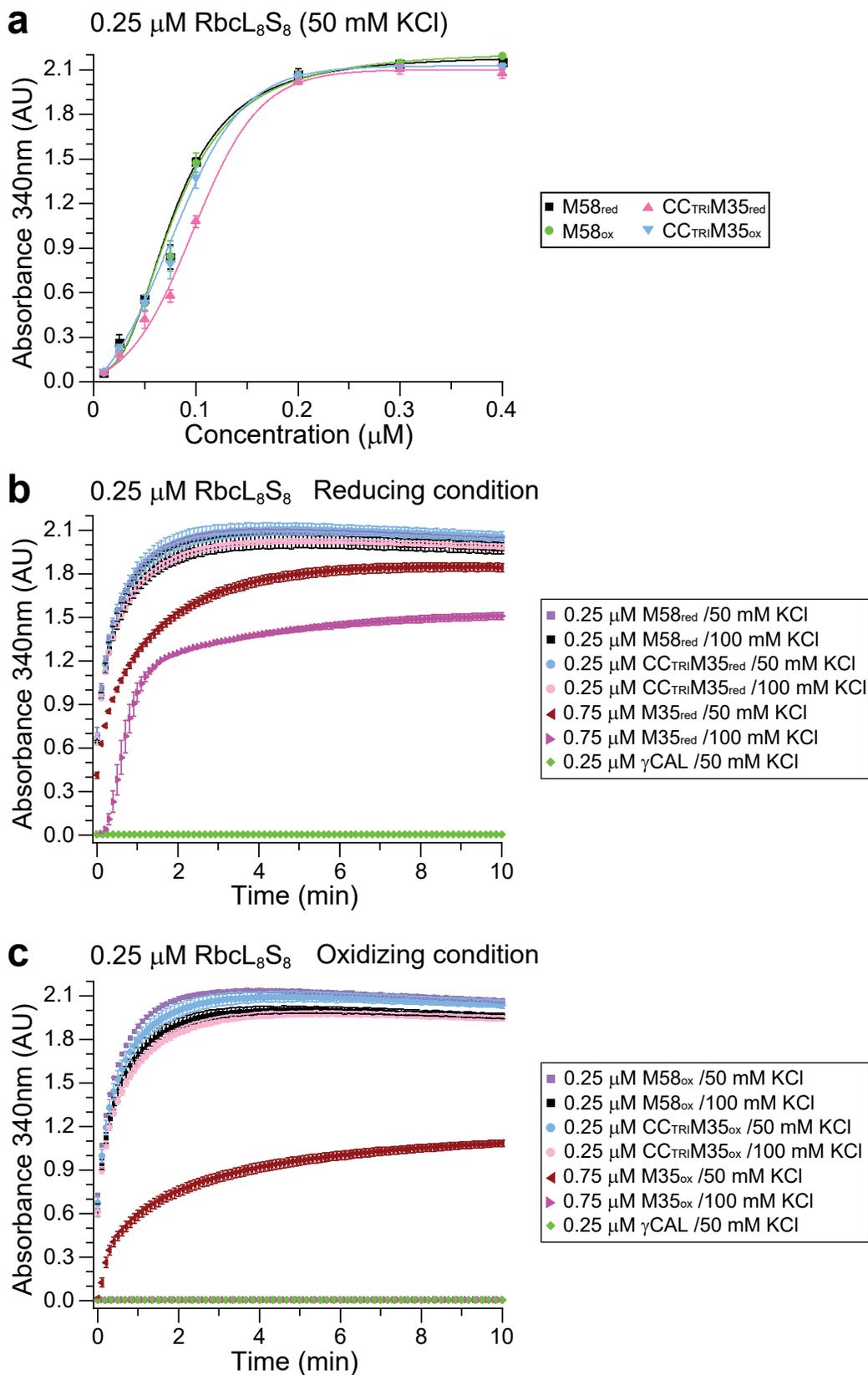


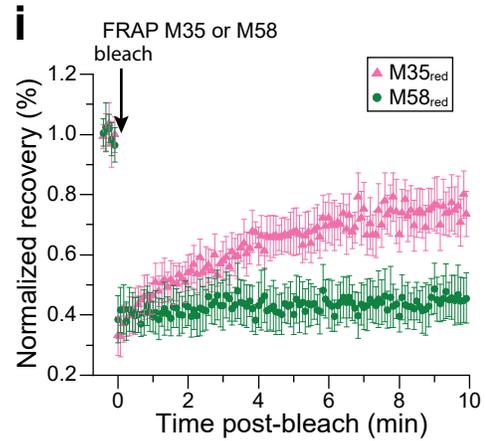
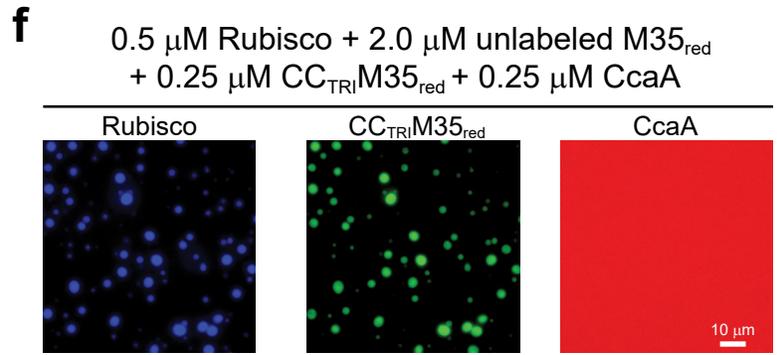
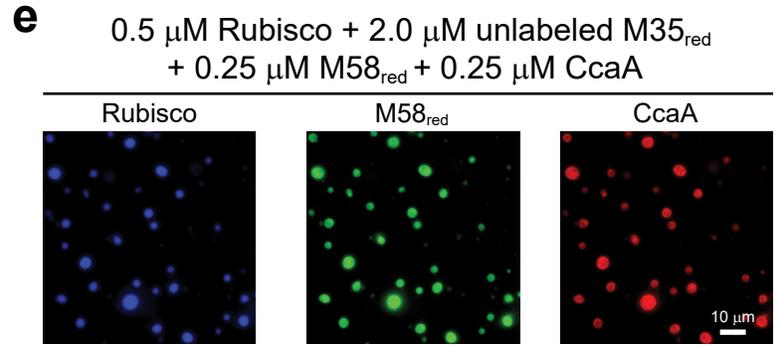
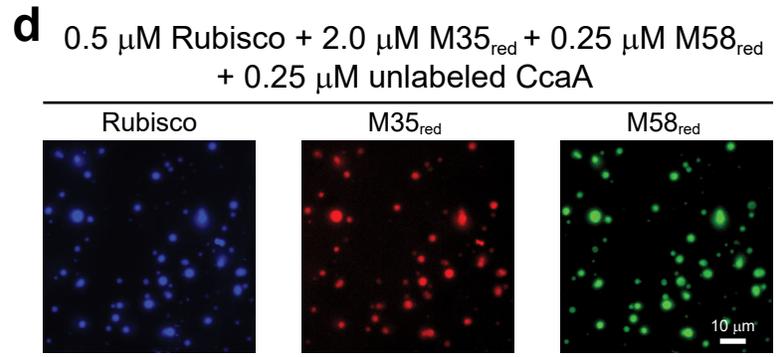
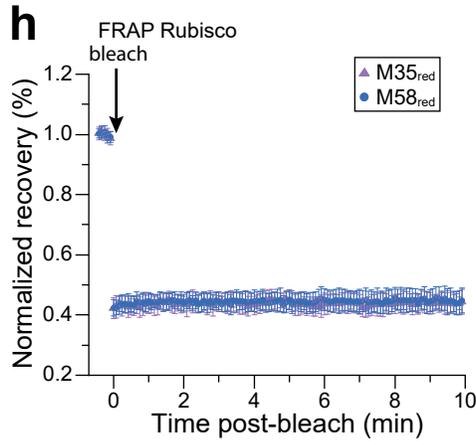
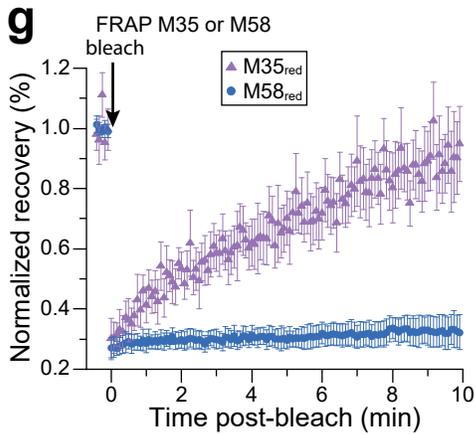
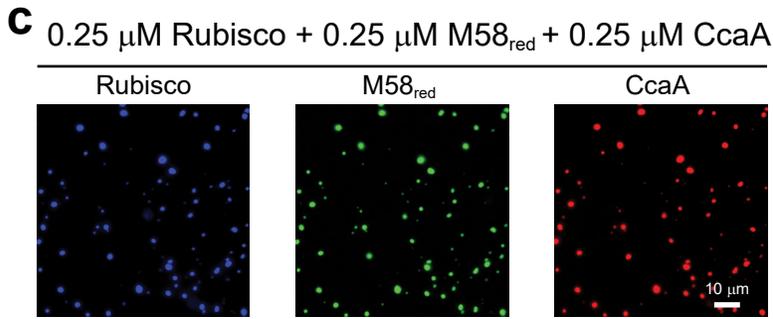
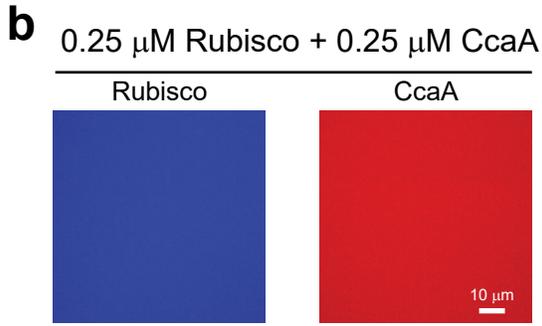
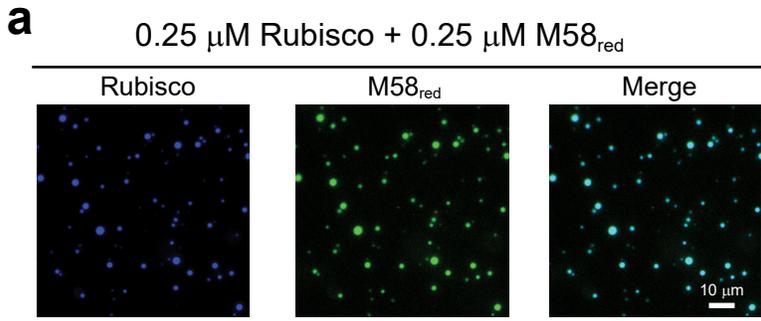




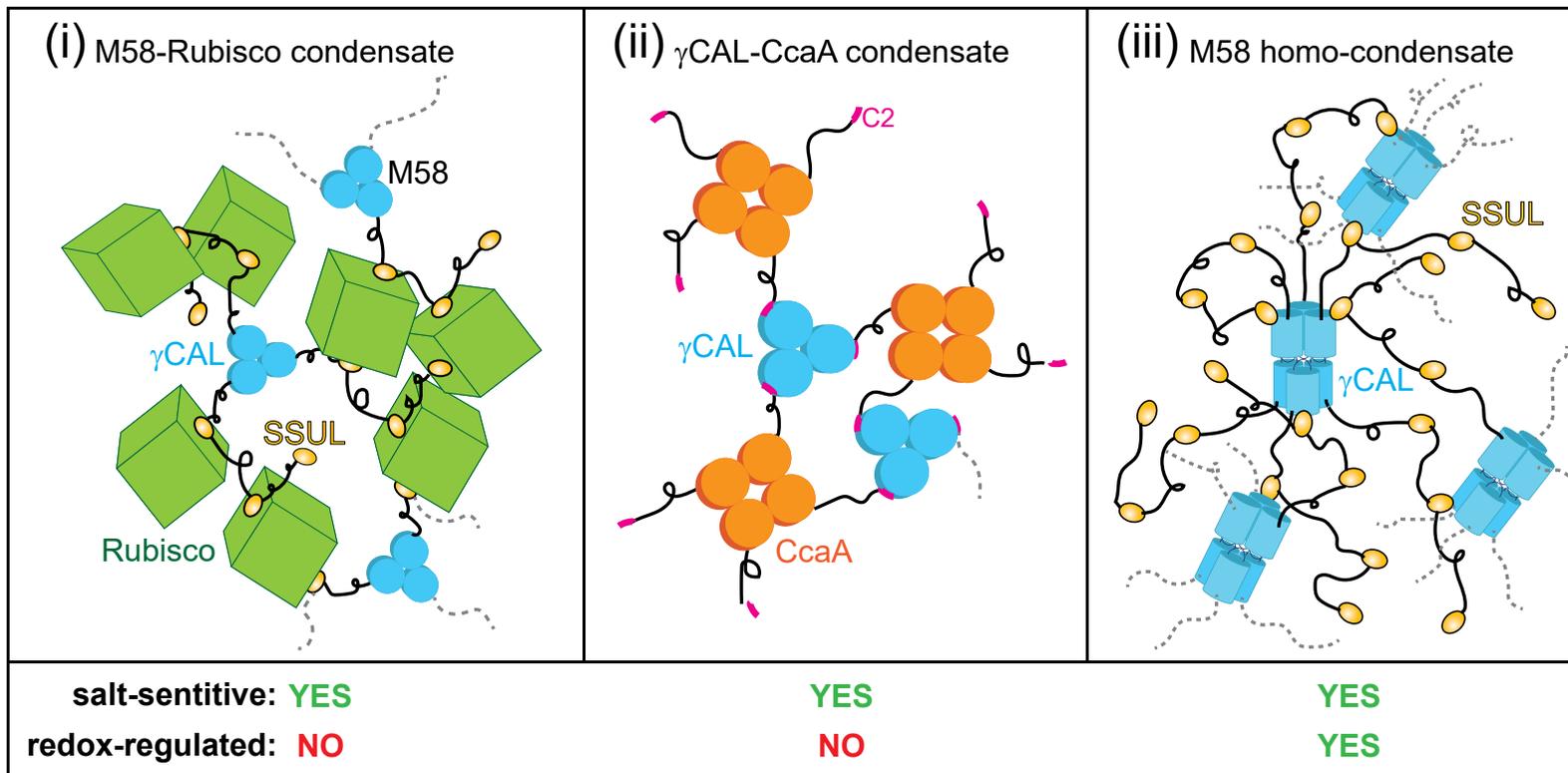




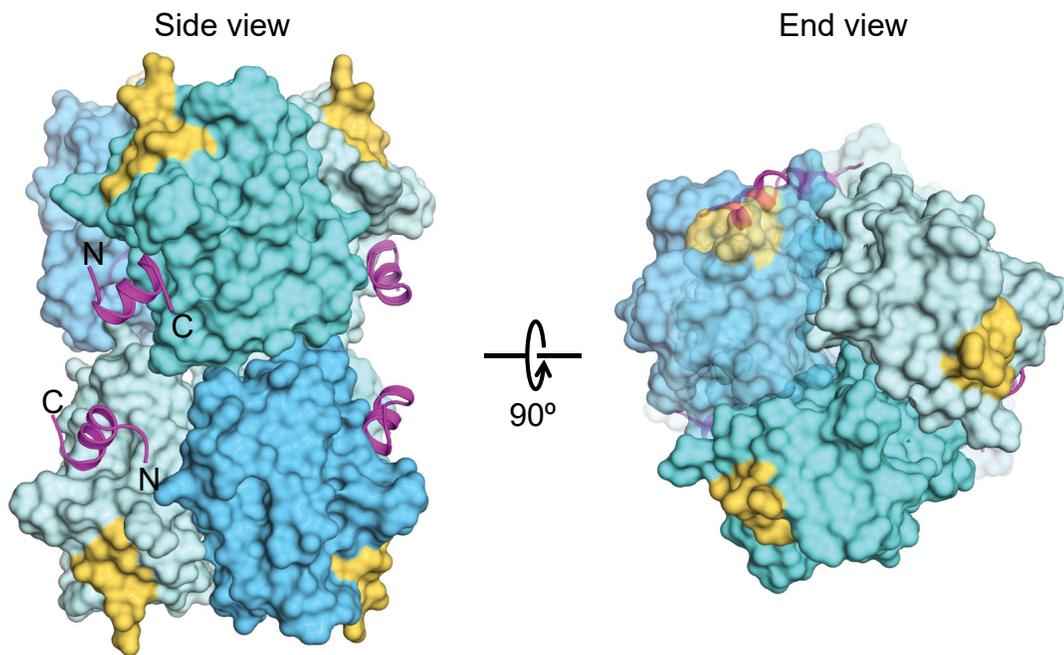




a



b



Supporting Information

Scaffolding Protein CcmM Directs Multiprotein Phase Separation in β - Carboxysome Biogenesis

Kun Zang^{1#}, Huping Wang^{1#}, F. Ulrich Hartl¹ and Manajit Hayer-Hartl^{1,*}

Affiliation:

¹Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, Am Klopferspitz
18, 82152 Martinsried, Germany

[#]These authors contributed equally

*Correspondence: mhartl@biochem.mpg.de (M.H-H.)

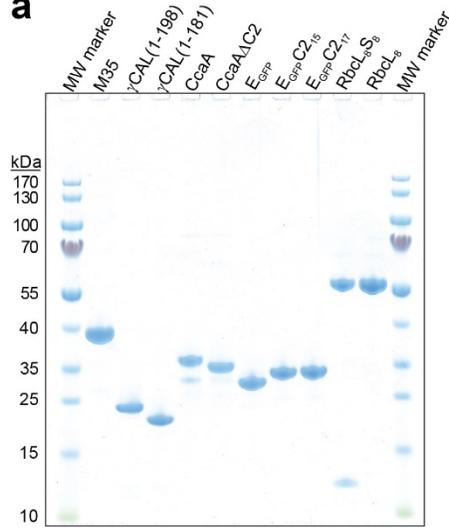
This PDF file includes:

1. Supplementary Figures 1-10

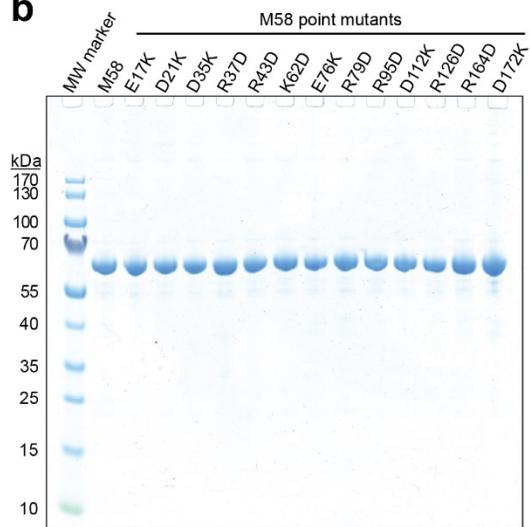
2. Supplementary Tables 1-6

1. Supplementary Figures

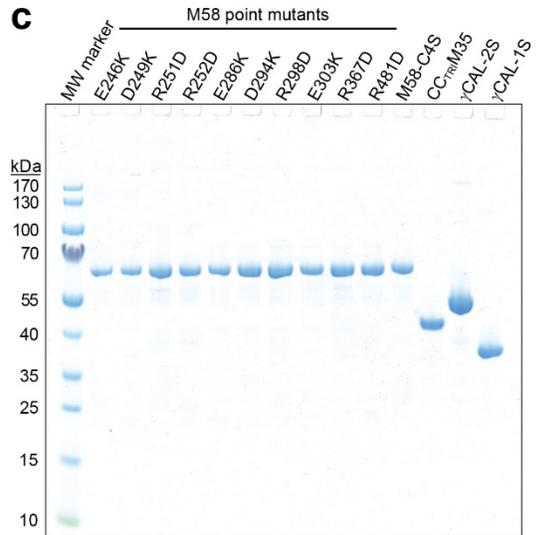
a



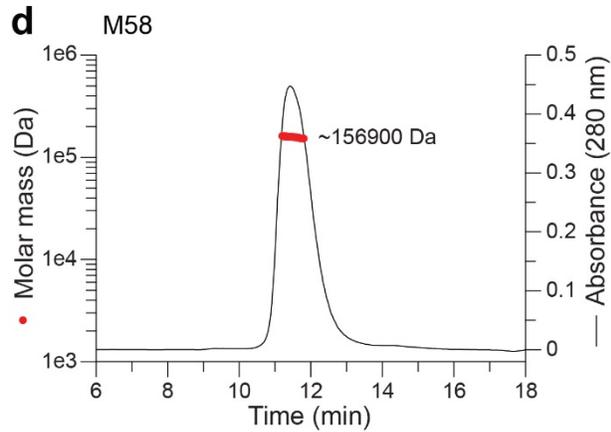
b



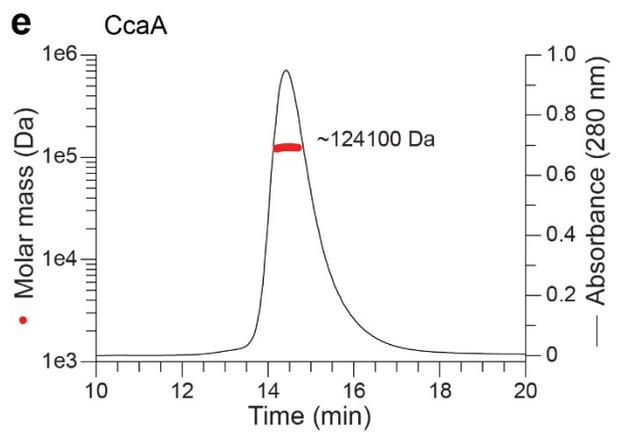
c



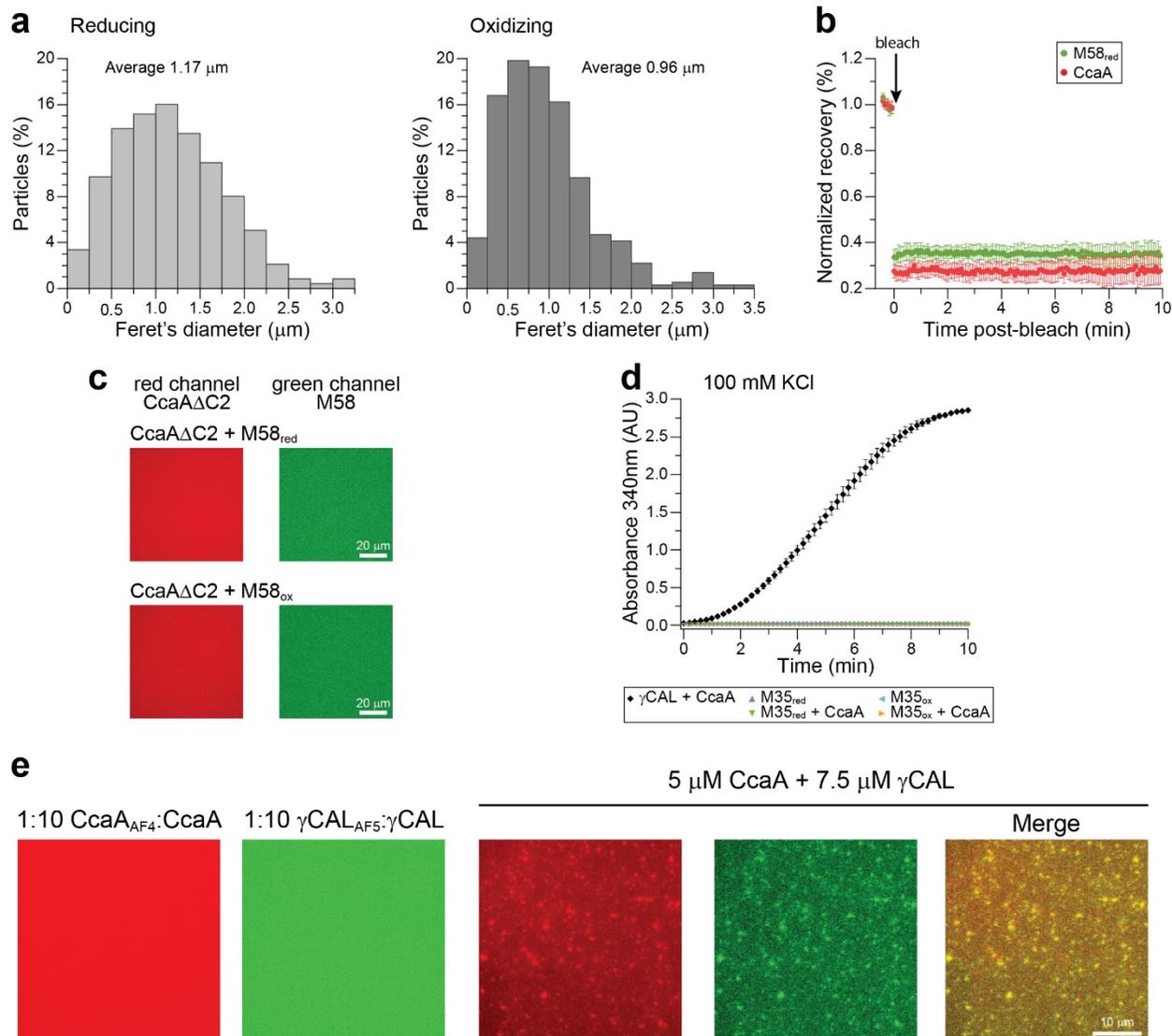
d



e

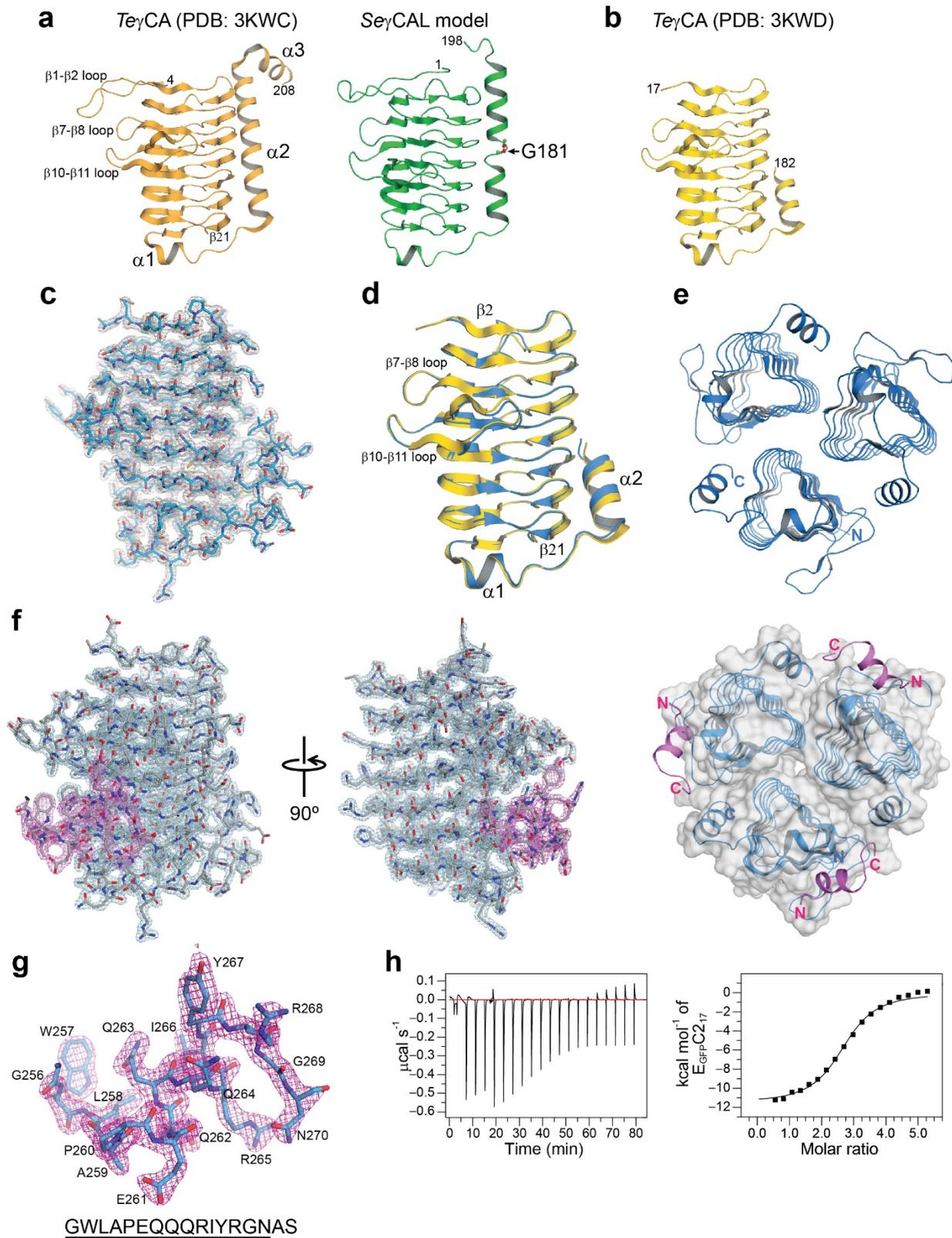


Supplementary Fig. 1 | Analysis of purified proteins. **a-c**, Purified proteins used in this study were analyzed by SDS-PAGE and Coomassie staining. **d-e**, SEC-MALS analysis of purified M58_{red} (**d**) and CcaA (**e**). The expected molecular mass of the M58 trimer is 173498.52 Da and of the CcaA tetramer 120741.44 Da. The observed mass values are indicated. M58 was analyzed once, while representative data is shown for CcaA (n=3).



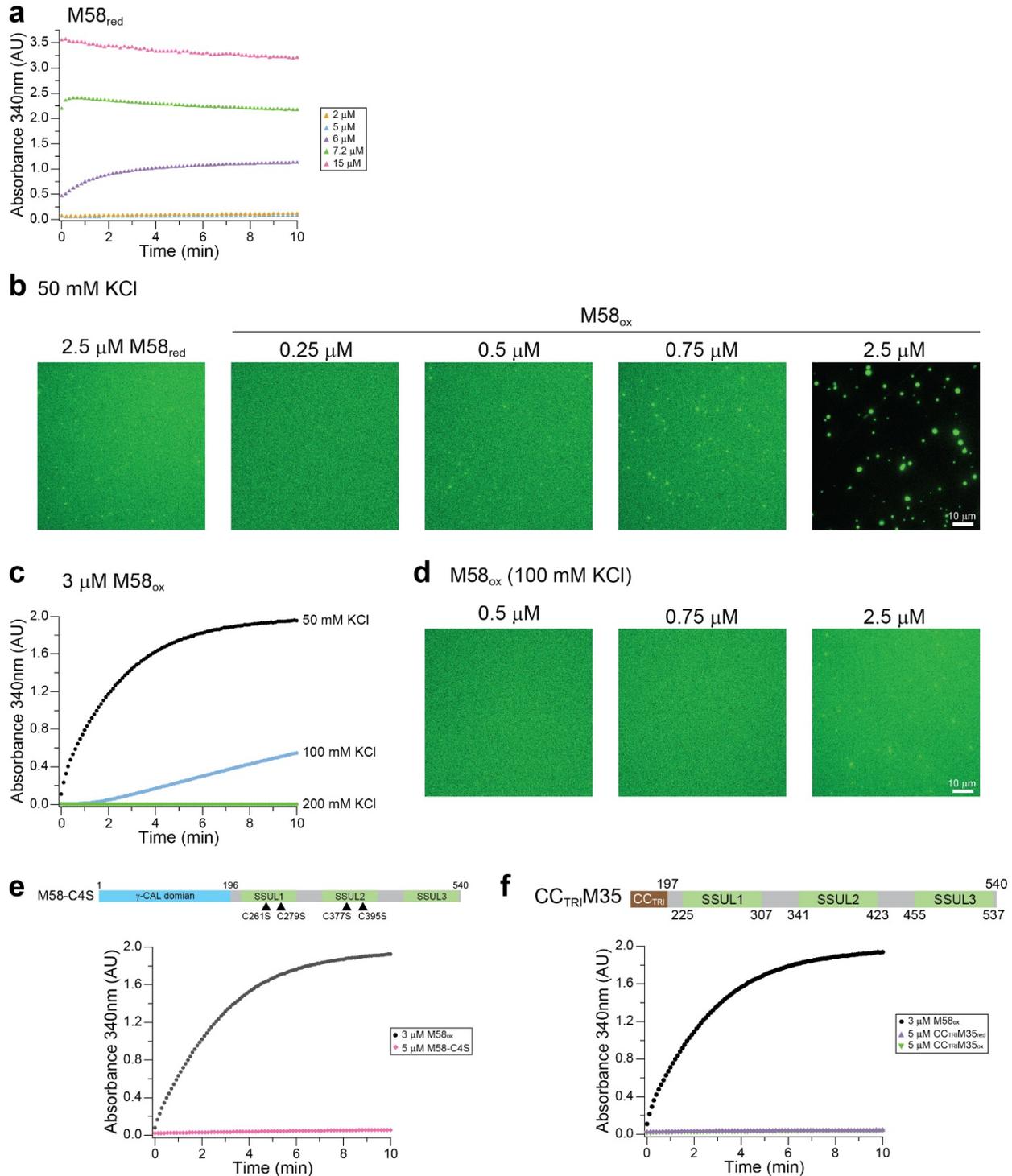
Supplementary Fig. 2 | M58-CcaA condensate formation. **a**, Size distribution (Ferret's diameter) of M58-CcaA condensates formed under reducing (left) and oxidizing conditions (right). 237 and 363 droplet-shaped condensates were analyzed under reducing or oxidizing conditions, respectively. Condensates were generated as in Fig. 1e,f. **b**, M58 and CcaA in condensates are immobile. Fluorescence recovery after photobleaching (FRAP) experiments were performed on condensates formed by fluorescence labeled M58_{red} (M58_{AF5}) with unlabeled CcaA or by unlabeled M58_{red} with labeled CcaA (CcaA_{AF4}). Condensates were generated as in Fig. 1e. Pre-bleach fluorescence is set to 1. Averages \pm SD from n=20 droplets. **c**, The C-terminal C2 sequence of CcaA is required for condensate formation with M58. M58 and CcaA Δ C2, fluorescence labeled as in Fig. 1e. (M58_{AF5}; CcaA Δ C2_{AF4}), were used as 1:10 mixtures with unlabeled protein (see Methods). Conditions were as in Fig. 1e,f. Images shown are from a single experiment (n=1). **d**, CcaA interacts with the γ CAL domain of M58, not with the SSUL modules. Condensate formation was analyzed by turbidity assay as in Fig. 1b,c with 5.0 μM CcaA, 7.5 μM γ CAL, 22.5 μM M35_{red}/M35_{ox} (100 mM KCl). Averages \pm SD are shown (n=3). **e**, Condensate formation of CcaA and γ CAL analyzed by fluorescence microscopy. N-

terminally labeled CcaA (CcaA_{AF4}) and γ CAL (γ CAL_{AF5}) were used as 1:10 mixtures with the respective unlabeled proteins and analyzed either alone or in combination at the concentrations as in (b). Images shown are from a single experiment (n=1).



Supplementary Fig. 3 | Structure of the γ CAL domain with and without bound C2 peptide.
a, Left: Crystal structure of the γ CA domain of CcmM from *Thermosynechococcus elongatus*

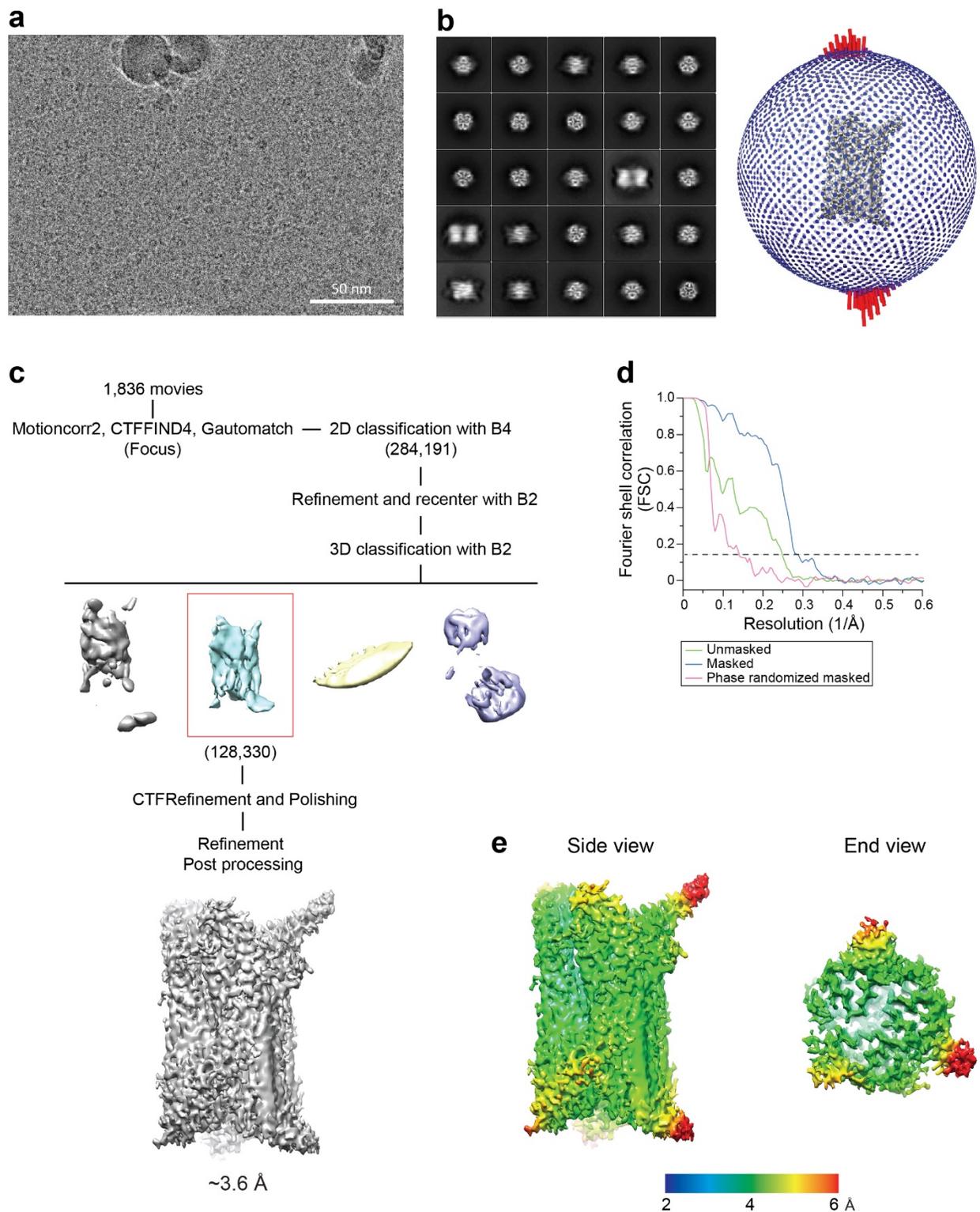
BP-1 (residues 1-209, PDB: 3KWC)³⁰ in ribbon representation. Right: Model of the structure of the *Se* γ CAL (1-198) generated with Phyre2⁷². The position of the putative break in the α 2 helix (residue G181) is indicated. **b**, Structure of the C-terminally truncated form of *Te* γ CA (residues 1-193, PDB: 3KWD) lacking the residues after the break in α 2. **c**, Experimental electron density for the *Se* γ CAL(1-181) protomer (for data collection and refinement statistics, see Supplementary Table 2). The weighted 2Fo–Fc density map contoured at 1.5 σ is shown as a gray meshwork with the final model in stick representation. **d**, Overlay of the structure of the γ CAL protomer (blue) and the structure of *Te* γ CA (PDB: 3KWD) (yellow). **e**, Model of the structure of the γ CAL trimer viewed along the long axis of the β -helices of the protomers. **f**, Left: Experimental electron density for the *Se* γ CAL-C2₁₇ complex (for data collection and refinement statistics, see Supplementary Table 2). The weighted 2Fo–Fc density map contoured at 1.5 σ is shown for γ CAL (gray) and the bound C2₁₇ peptide (magenta) as meshwork, with the final model in stick representation. Right: Model of the trimer of the γ CAL-C2₁₇ complex in transparent surface representation. **g**, The weighted 2Fo–Fc electron density map contoured at 1.5 σ for the 15 residues of the C2₁₇ peptide (underlined in the sequence below) that are resolved in the structure of the γ CAL-C2₁₇ complex (for data collection and refinement statistics, see Supplementary Table 2). **h**, Analysis by isothermal titration calorimetry (ITC) of the interaction of E_{GFP}C2₁₇ with γ CAL(1-198). Left: Change in enthalpy between the sample and the reference cell (buffer) upon titration of E_{GFP}C2₁₇ into buffer containing γ CAL(1-198). Right: The fitted integrated data from the left panel generated in Origin. Representative data are shown (n=2).



Supplementary Fig. 4 | Role of SSUL modules in M58 homo-condensate formation. a, Concentration dependence of M58_{red} homo-condensate formation. Turbidity assays were performed with M58_{red} at the concentrations indicated (50 mM KCl). Reactions were started by diluting M58_{red} from buffer containing 500 mM KCl as in Fig. 4c (n=2). **b,** M58 homo-condensate formation is enhanced under oxidizing conditions. Concentration dependence of

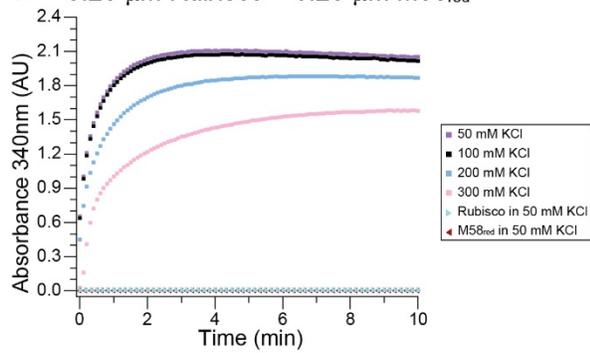
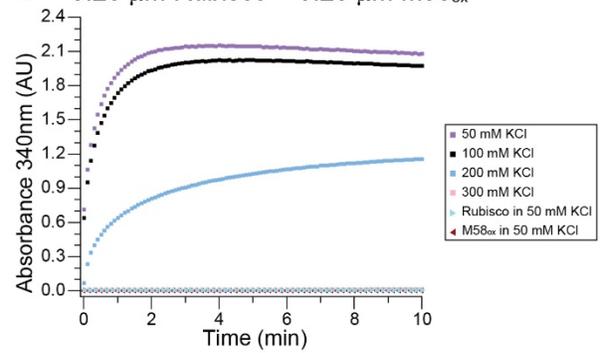
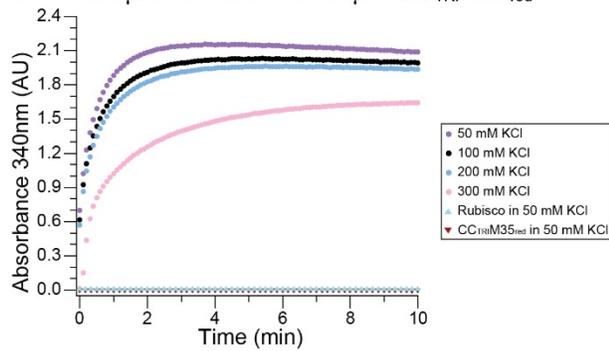
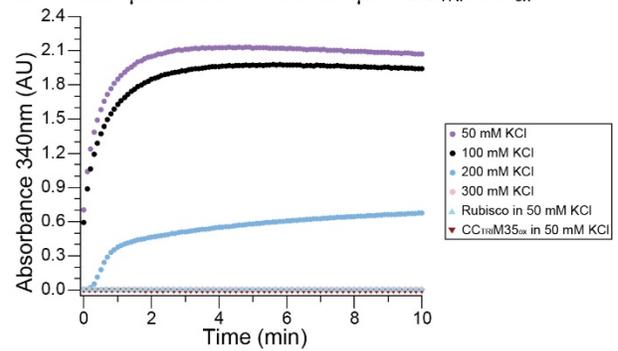
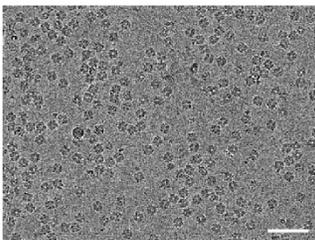
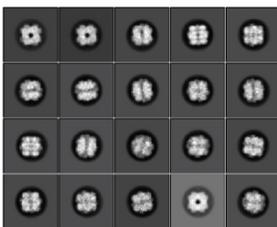
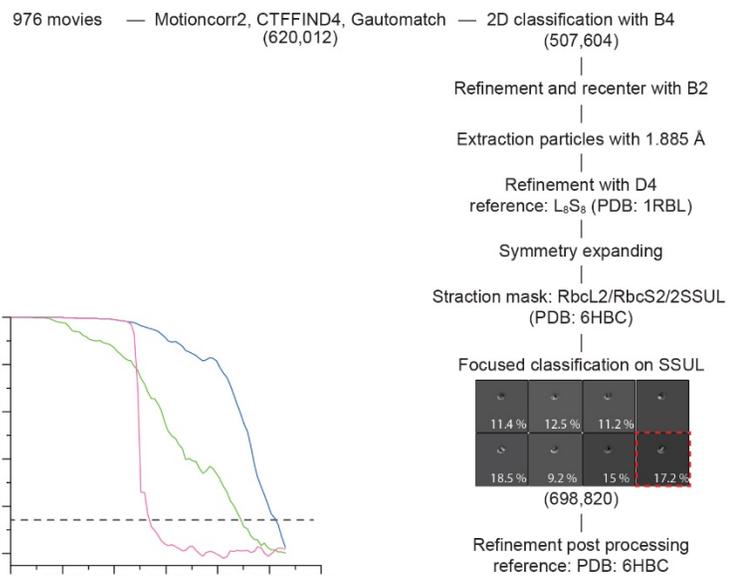
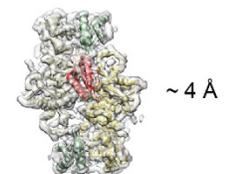
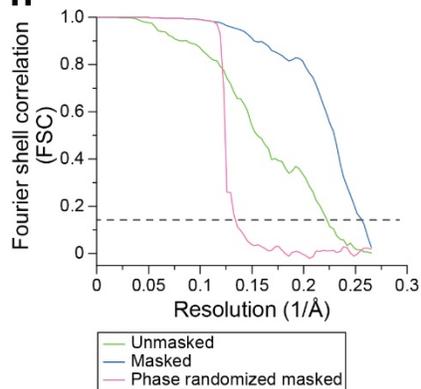
M58_{ox} condensate formation analyzed by fluorescence microscopy. M58 was N-terminally labeled with fluorophore Alexa532 (M58_{AF5}), and used as a 1:10 mixture with unlabeled protein at the final concentrations indicated. M58_{red} at 2.5 μ M is shown as control. Reactions were performed in 50 mM KCl. Representative data are shown (n=2). **c**, Salt dependence of M58_{ox} homo-condensate formation. Turbidity assays were performed with 3 μ M M58_{ox} at 50 to 200 mM KCl (n=2). **d**, M58_{ox} homo-condensate formation was analyzed by fluorescence microscopy as in (b) at 100 mM KCl at the concentrations of M58_{ox} indicated. Representative data are shown (n=2). **e**, Requirement of disulfide bond formation in SSUL modules for M58 homo-condensate formation. Top: Domain structure of M58 indicating the position of cysteine mutations disrupting disulfide bond formation in SSUL1 and SSUL2. Bottom: M58_{ox} (3 μ M) and M58-C4S (5 μ M) were analyzed by turbidity assay. Reactions were performed at 50 mM KCl. Representative data are shown (n=2). **f**, M58 homo-condensate formation requires the γ CAL domains of M58. Top: Domain structure of CC_{TRIM35}. Bottom: Turbidity assays were performed with 3 μ M M58_{ox}, 5 μ M CC_{TRIM35}_{red} and 5 μ M CC_{TRIM35}_{ox} at 50 mM KCl. Representative data are shown (n=2).

Conserved charged residues mapped on the γ CAL structure (PDB: 7O4Z) shown in surface representation. Mutated residues are circled.

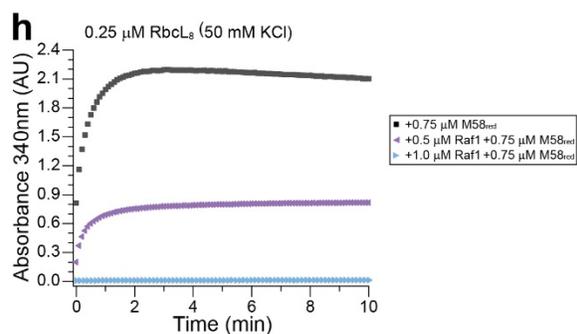
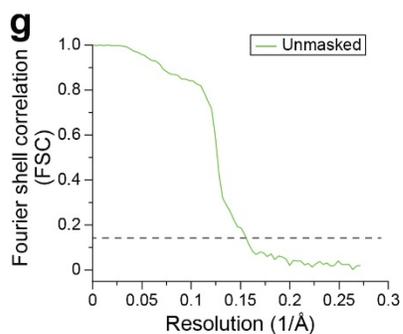
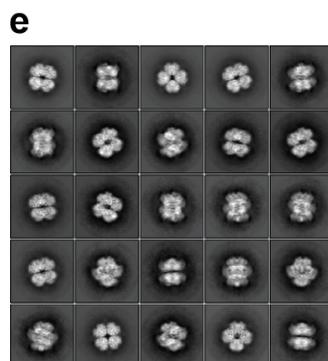
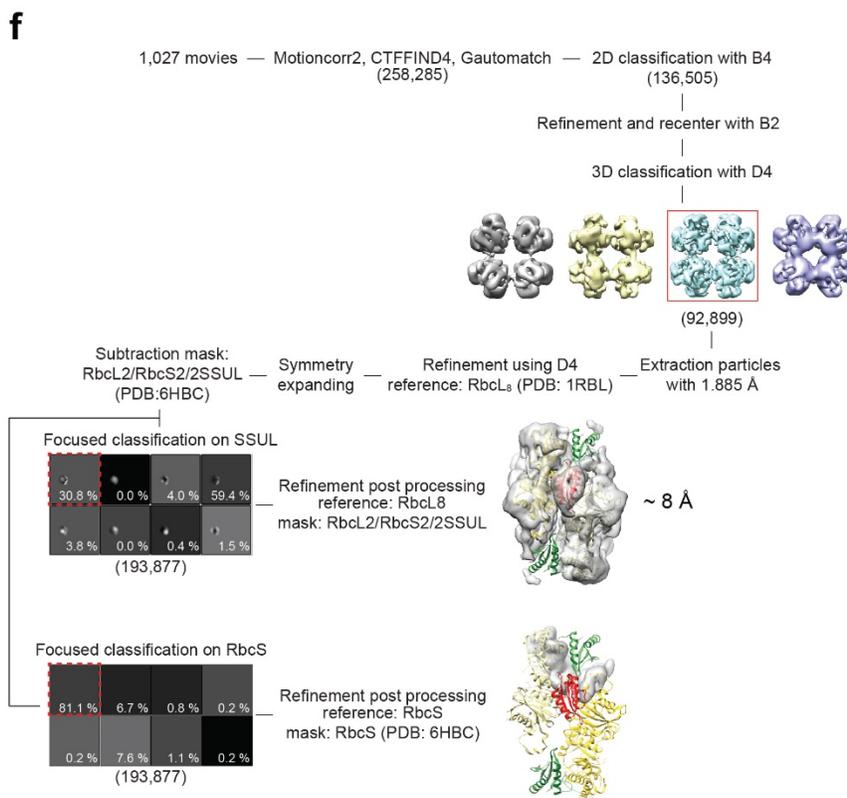
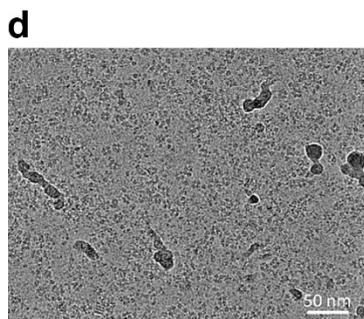
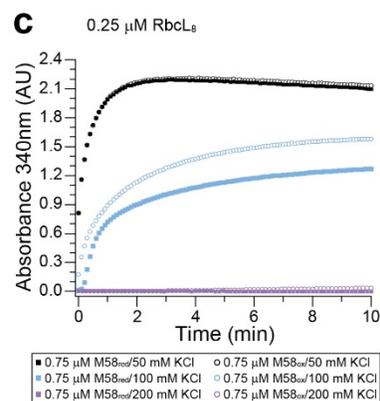
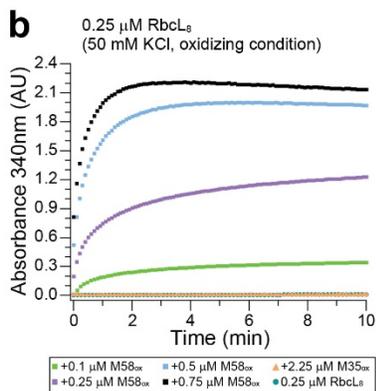
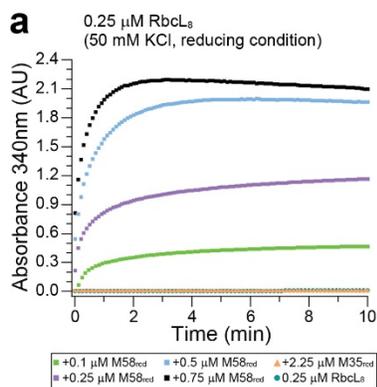


Supplementary Fig. 6 | Cryo-EM single-particle reconstruction of M58 in M58 homo-condensates. **a**, A representative micrograph of the M58_{ox} homo-condensate formed as in Fig. 4c. Scale bar, 50 nm. **b**, 2D class averages of particles in (a) (left) and the particle angular

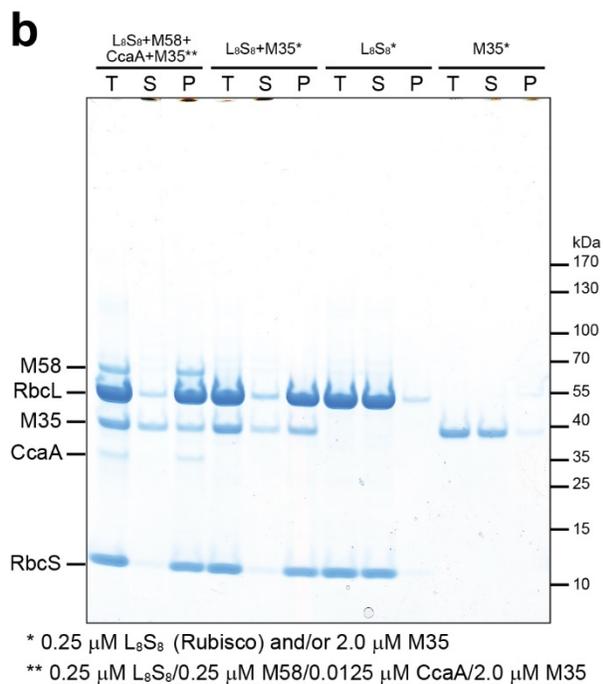
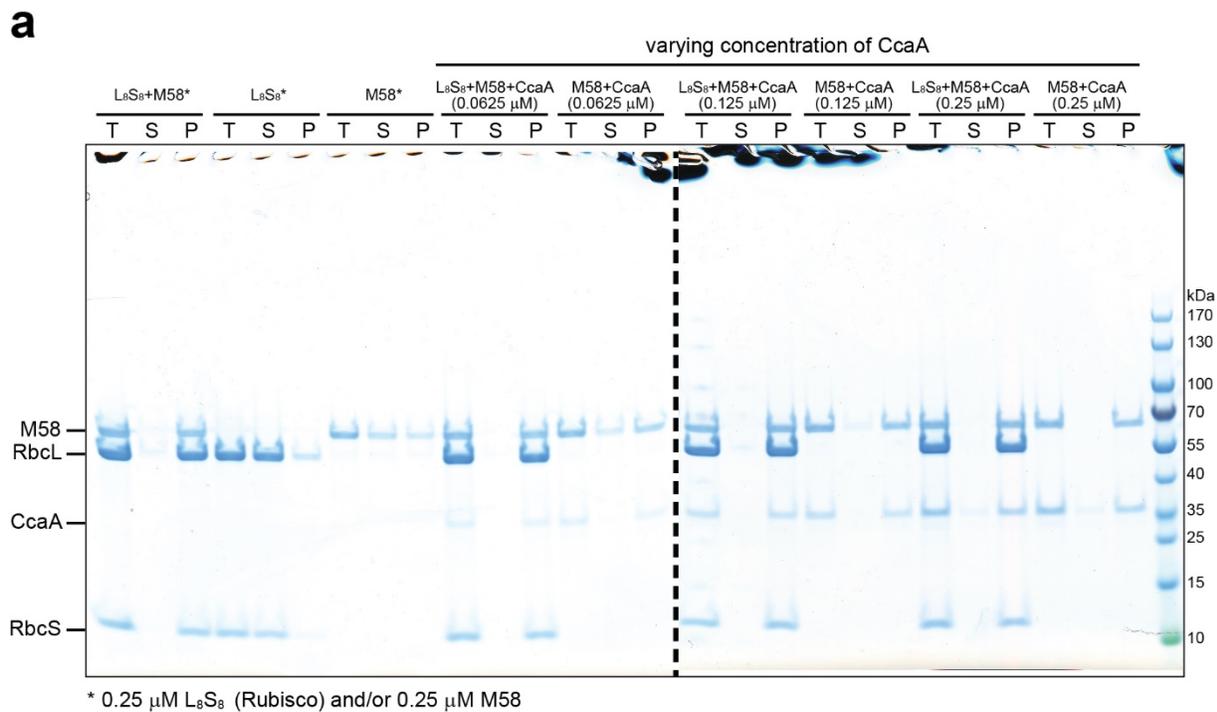
distribution plot (right) showing a preferred particle orientation in end-view. **c**, The single-particle data processing workflow. Particle numbers are indicated in parentheses. B4, 4x4 pixel-binned image. B2, 2x2 pixel-binned image. See Methods for details. **d**, Gold-standard FSC curves (EMDB: EMD-12730). **e**, Local resolution maps of side and end views. The color gradient from blue to red indicates local resolution from 2.0 to 6.0 Å.

a 0.25 μ M Rubisco + 0.25 μ M M58_{red}**b** 0.25 μ M Rubisco + 0.25 μ M M58_{ox}**c** 0.25 μ M Rubisco + 0.25 μ M CC_{TR1}M35_{red}**d** 0.25 μ M Rubisco + 0.25 μ M CC_{TR1}M35_{ox}**e****f****g****h**

Supplementary Fig. 7 | Co-assembly of M58 and Rubisco. **a**, Salt dependence of M58-Rubisco condensate formation under reducing conditions. M58_{red} and Rubisco (0.25 μ M each) were incubated in buffer containing 50-300 mM KCl and analyzed by turbidity assay. M58_{red} and Rubisco alone were analyzed as controls at 50 mM KCl. Representative data are shown (n=2). **b**, M58-Rubisco condensate formation is more salt sensitive under oxidizing conditions. M58_{ox} and Rubisco (0.25 μ M each) were incubated in buffer containing 50-300 mM KCl and analyzed by turbidity assay as in (a). Representative data are shown (n=2). **c**, Salt dependence of CC_{TRI}M35-Rubisco condensate formation under reducing conditions. CC_{TRI}M35_{red} and Rubisco (0.25 μ M each) were incubated in buffer containing 50-300 mM KCl and analyzed by turbidity assay. CC_{TRI}M35_{red} and Rubisco alone were analyzed as controls at 50 mM KCl. Representative data are shown (n=2). **d**, CC_{TRI}M35-Rubisco condensate formation is more salt sensitive under oxidizing conditions. CC_{TRI}M35_{ox} and Rubisco (0.25 μ M each) were incubated in buffer containing 50-300 mM KCl and analyzed by turbidity assay as in (c). Representative data are shown (n=2). **e-h**, Cryo-EM single particle analysis of the M58_{red}-Rubisco complex. A representative micrograph (**e**). Scale bar, 50 nm. 2D class averages of particles (**f**). The single-particle data processing workflow (**g**). Particle numbers are in parentheses. B4, 4x4 pixel-binned image. B2, 2x2 pixel-binned image. D4, *D4* symmetry applied. See Methods for details. Gold-standard FSC curves of the RbcL₂-RbcS₂-SSUL reconstruction from the complex of M58_{red}-Rubisco (**h**) (EMDB: EMD-12731).



Supplementary Fig. 8 | Interaction of SSUL with the RbcL₈ core complex. a-c, The high local concentration of SSUL modules in M58 enables interaction with RbcL₈. **(a)** M58-RbcL₈ condensate formation at different concentrations of M58 under reducing conditions. RbcL₈ (0.25 μM) was incubated with M58_{red} at the concentrations indicated. Reactions were analyzed by turbidity assay. M35_{red} in combination with RbcL₈ and RbcL₈ alone were analyzed as controls. Representative data are shown (n=2). **(b)** M58-RbcL₈ condensate formation at different concentrations of M58 under oxidizing conditions. RbcL₈ (0.25 μM) was incubated with M58_{ox} at the concentrations indicated. Reactions were analyzed by turbidity assay. M35_{ox} in combination with RbcL₈ and RbcL₈ alone were analyzed as controls. Representative data are shown (n=2). **(c)** Similar salt sensitivity of M58-RbcL₈ condensate formation under reducing and oxidizing conditions. RbcL₈ (0.25 μM) was incubated with M58_{red} or M58_{ox} (0.75 μM) in buffer containing 50-200 mM KCl. Reactions were analyzed by turbidity assay. Representative data are shown (n=2). **d-g,** Cryo-EM single particle analysis of the M58_{red}-RbcL₈ condensate. **(d)** A representative micrograph of M58_{red}-RbcL₈ complexes. Scale bar, 50 nm. **(e)** 2D class averages of particles. **(f)** The single-particle data processing workflow. Particle numbers are indicated in parentheses. B4, 4x4 pixel-binned image. B2, 2x2 pixel-binned image. D4, *D4* symmetry applied. See Methods for details. **(g)** Gold-standard FSC curve of the unsharpened RbcL₂-SSUL reconstruction solved from the complex of M58_{red}-RbcL₈ (EMDB: EMD-12732). **h,** The assembly chaperone Raf1 prevents the interaction of M58 with RbcL₈. Condensate formation of M58_{red} (0.75 μM) and RbcL₈ (0.25 μM) was analyzed by turbidity assay at 50 mM KCl in the presence of 0.5 μM and 1.0 μM of Raf1 (concentration of the functional Raf1 dimer). Representative data are shown (n=2).

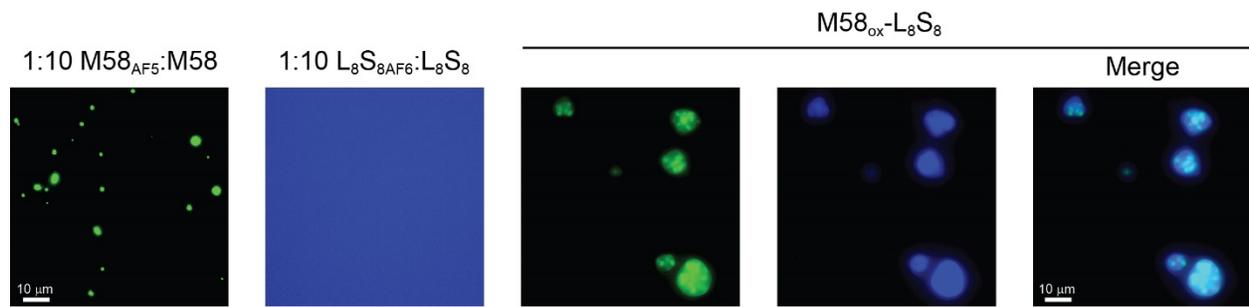


Supplementary Fig. 9 | Analysis of condensate formation by sedimentation. **a**, Efficient sequestration of Rubisco (L₈S₈) and CcaA by M58_{red}. M58_{red}-Rubisco and M58_{red}-CcaA condensates in the presence of 100 mM KCl were formed as described in Supplementary Fig. 7a and Fig. 1b, respectively, followed by fractionation into total (T), supernatant (S) and pellet (P) by centrifugation (20,817 x g for 20 min at 25 °C), and analysis by SDS-PAGE and Coomassie staining. The concentration of CcaA was varied from 0.0625 to 0.25 μM at a constant

concentration of M58_{red} of 0.25 μ M. Rubisco (L₈S₈) and M58_{red} were analyzed alone as controls. Representative data are shown (n=2). **b**, Condensates of Rubisco (L₈S₈), CcaA, M58_{red} and M35_{red} as well as of L₈S₈ and M35_{red} were formed at the concentrations indicated and analyzed by sedimentation as in (a). Representative data are shown (n=2).

Oxidizing (50 mM KCl)

M58_{ox} $\xrightarrow{10 \text{ min}}$ + Rubisco $\xrightarrow{2 \text{ min}}$ $\xrightarrow{\text{dilution}}$ Fluorescence microscopy



Supplementary Fig. 10 | Formation of biphasic M58-Rubisco condensates under oxidizing conditions. M58_{ox}-Rubisco condensates were analyzed by fluorescence microscopy using M58_{ox} and Rubisco N-terminally labeled with Alexa532 and Alexa647, respectively (M58_{AF5}; L₈S_{8AF6}), and used as 1:10 mixtures with unlabeled protein. M58_{ox} homo-condensates (3.5 μM) were first allowed to form for 10 min, followed by addition of Rubisco (L₈S₈) to reach final concentrations of M58_{ox} and Rubisco of 2.5 μM each. The reaction was diluted 10 times before fluorescence microscopy. Large M58_{ox}-Rubisco condensates formed containing dense M58_{ox} foci. The M58_{ox} homo-condensate and Rubisco were also analyzed alone. Images shown are from a single experiment (n=1). Note a similar result was observed at 100 mM KCl, however, the overall size of the condensates was smaller.

2. Supplementary Tables

Supplementary Table 1 | Molar mass and hydrodynamic radius of proteins determined by SEC-MALS.

| Protein | Measured Molar mass (Da) | Theoretical MW (Da) | | Hydrodynamic radius (nm) |
|-----------------------|--------------------------|---------------------|-----------|--------------------------|
| | | monomer | oligomer | |
| CcaA | ~124100* | 30185.36 | 120741.44 | ~5.4 |
| CcaAAC2 | ~114000* | 28472.47 | 113889.88 | ~5.2 |
| M58 | ~156900 | 57832.84 | 173498.52 | ~6.3 |
| M58-C4S | ~153600 | 57768.60 | 173305.80 | ~6.1 |
| γ CAL-2S | ~125600 | 46303.89 | 138911.67 | ~5.1 |
| γ CAL-1S | ~95190 | 33764.01 | 101292.03 | ~4.6 |
| γ CAL(1-198) | ~67050* | 20976.88 | 62930.62 | ~3.7 |
| γ CAL(1-181) | ~61160* | 19228.91 | 57686.73 | ~3.4 |
| EGFP | ~23990* | 27286.79 | 27286.79 | ~2.1 |
| EGFP-C2 ₁₅ | ~30960* | 28999.68 | 28999.68 | ~2.0 |
| EGFP-C2 ₁₇ | ~30810* | 29242.94 | 29242.94 | ~2.1 |
| M58-E17K | ~155800 | 57831.90 | 173495.70 | ~6.2 |
| M58-D21K | ~154800 | 57845.93 | 173537.79 | ~6.1 |
| M58-D35K | ~152100 | 57845.93 | 173537.79 | ~6.2 |
| M58-R37D | ~153700 | 57791.75 | 173375.25 | ~6.2 |
| M58-R43D | ~152400 | 57791.75 | 173375.25 | ~6.2 |
| M58-K62D | ~154300 | 57819.76 | 173459.28 | ~6.2 |
| M58-E76K | ~152900 | 57831.90 | 173495.70 | ~6.2 |
| M58-R79D | ~155800 | 57791.75 | 173375.25 | ~6.3 |
| M58-R95D | ~154100 | 57791.75 | 173375.25 | ~6.2 |
| M58-D112K | ~154900 | 57845.93 | 173537.79 | ~6.1 |
| M58-R126D | ~159500 | 57791.75 | 173375.25 | ~6.3 |
| M58-R164D | ~150500 | 57791.75 | 173375.25 | ~6.2 |
| M58-D172K | ~151600 | 57845.93 | 173537.79 | ~6.2 |
| M58-E246K | ~149200 | 57831.90 | 173495.70 | ~6.1 |
| M58-D249K | ~155800 | 57845.93 | 173537.79 | ~6.1 |
| M58-R251D | ~149000 | 57791.75 | 173375.25 | ~6.1 |
| M58-R252D | ~147700 | 57791.75 | 173375.25 | ~6.1 |
| M58-E286K | ~158000 | 57831.90 | 173495.7 | ~6.2 |
| M58-D294K | ~157400 | 57845.93 | 173537.79 | ~6.1 |
| M58-R298D | ~156300 | 57791.75 | 173375.25 | ~6.3 |
| M58-E303K | ~158300 | 57831.90 | 173495.7 | ~6.2 |
| M58-R367D | ~154400 | 57791.75 | 173375.25 | ~6.3 |
| M58-R481D | ~153600 | 57791.75 | 173375.25 | ~6.3 |
| M35 | ~33840 | 37158.29 | ----- | ~3.1 |
| CC _{TRI} M35 | ~101600 | 40375.11 | 121125.33 | ~5.1 |

*Measured in buffer 50 mM Tris-HCl pH 8.0/150 mM KCl.

All other proteins measured in buffer 50 mM Tris-HCl pH 8.0/500 mM KCl/5 mM DTT.

Supplementary Table 2 | Data collection, phasing and refinement statistics.

| | <i>Seγ</i> CAL(1-181)* | <i>Seγ</i> CAL(1-181)-C2 ₁₇ complex* |
|--|----------------------------|---|
| PDB: | 7O4Z | 7O54 |
| Data collection | | |
| Space group | <i>H</i> 32 | <i>H</i> 32 |
| Cell dimensions | | |
| <i>a</i> , <i>b</i> , <i>c</i> (Å) | 89.74, 89.74, 130.79 | 89.36, 89.36, 129.84 |
| α , β , γ (°) | 90.0, 90.0, 120.0 | 90.0, 90.0, 120.0 |
| Wavelength (Å) | 0.87313 | 0.99989 |
| Resolution (Å) ** | 66.81 – 1.67 (1.70 – 1.67) | 37.08 – 1.63 (1.66 – 1.63) |
| <i>R</i> _{sym} or <i>R</i> _{merge} | 0.081 (0.960) | 0.030 (0.684) |
| <i>I</i> / σ <i>I</i> | 13.8 (2.3) | 45.4 (2.2) |
| Completeness (%) | 99.9 (100.0) | 98.8 (91.3) |
| Redundancy | 9.7 (10.0) | 18.4 (9.2) |
| Refinement | | |
| Resolution (Å) | 66.81 – 1.67 | 37.08 – 1.63 |
| No. reflections | 22708 | 23554 |
| <i>R</i> _{work} / <i>R</i> _{free} | 0.164 / 0.200 | 0.177 / 0.200 |
| No. atoms | | |
| Protein | 1248 | 1376 |
| Ligand/ion | 2 | 2 |
| Water | 127 | 103 |
| <i>B</i> -factors (Å ²) | | |
| Protein | 30 | 37 |
| Ligand/ion | 32 | 36 |
| Water | 42 | 44 |
| R.m.s deviations | | |
| Bond lengths (Å) | 0.015 | 0.017 |
| Bond angles (°) | 1.95 | 2.10 |
| Ramachandran Plot | | |
| Favored (%) | 98.2 | 98.3 |
| Allowed (%) | 1.8 | 1.7 |
| Disallowed (%) | - | - |

*One crystal per structure.

**Values in parentheses are for highest-resolution shell.

Supplementary Table 3 | Cryo-EM statistics and model validation.

| | <i>SeM58_{ox}</i> (EMDB: EMD-12730) | <i>SeM58_{red}-SeRubisco</i> (EMDB: EMD-12731) | <i>SeM58_{red}-SeRbcL₈</i> (EMDB: EMD-12732) |
|--|--|---|--|
| Data collection and processing | | | |
| Magnification | 215000 | 22000 | 22000 |
| Voltage (kV) | 300 | 200 | 200 |
| Electron exposure (e-/Å ²) | 60 | 45 | 45 |
| Defocus range (μm) | -0.65 to -2.15 | -0.7 to -4.5 | -0.7 to -4.5 |
| Pixel size (Å) | 0.4114 | 1.885 | 1.885 |
| Symmetry imposed | C1 | C1 | C1 |
| Initial particle images (no.) | 349,391 | 620,012 | 258,285 |
| Final particle images (no.) | 128,330 | 698,820 | 193,877 |
| Map resolution (Å) | 3.57 | 4.0 | 8.0 |
| FSC threshold | 0.143 | 0.143 | 0.143 |
| Map resolution range (Å) | 2.4-23.57 | 3.5-50.0 | 8.0-50.0 |
| Refinement | | | |
| Initial model used (PDB code) | PDB: 7O4Z | 1RBL, 6HAS | 1RBL, 6HAS |
| Model resolution (Å) | 1.67 | 2.2, 8.8 | 2.2 8.8 |
| FSC threshold | 0.5 | 0.5 | 0.5 |
| Model resolution range (Å) | 1.67-66.81 | 2.2-130 | 2.2-130 |

Supplementary Table 4 | Size distribution of condensates.

| Sample | Average Feret's diameter (μm) |
|--|--|
| 0.25 μM M58 _{red} /0.25 μM CcaA | ~1.2 |
| 0.25 μM M58 _{ox} /0.25 μM CcaA | ~1.0 |
| 0.25 μM Rubisco/0.25 μM M58 _{red} | ~1.6 |
| 0.25 μM Rubisco/0.25 μM M58 _{red} /0.25 μM CcaA | ~1.5 |
| 0.5 μM Rubisco/2.0 μM M35 _{red} /0.25 μM M58 _{red} /0.25 μM CcaA | ~2.25 |
| 0.5 μM Rubisco/2.0 μM M35 _{red} /0.25 μM CC _{TRI} M35 _{red} /0.25 μM CcaA | ~2.6 |

Supplementary Table 5 | Oligo sequences used in this study.

See separate Excel spreadsheet.

Supplementary Table 6 | Plasmids generated or used in this study.

| Plasmid/Strain | Protein/Strain | Source |
|---|-----------------------------------|-------------------|
| pHUE- <i>SeM58</i> | <i>SeM58</i> | This study |
| pHUE- <i>SeM58-E17K</i> | M58-E17K | This study |
| pHUE- <i>SeM58-D21K</i> | M58-D21K | This study |
| pHUE- <i>SeM58-D35K</i> | M58-D35K | This study |
| pHUE- <i>SeM58-R37D</i> | M58-R37D | This study |
| pHUE- <i>SeM58-R43D</i> | M58-R43D | This study |
| pHUE- <i>SeM58-K62D</i> | M58-K62D | This study |
| pHUE- <i>SeM58-E76K</i> | M58-E76K | This study |
| pHUE- <i>SeM58-R79D</i> | M58-R79D | This study |
| pHUE- <i>SeM58-R95D</i> | M58-R95D | This study |
| pHUE- <i>SeM58-D112K</i> | M58-D112K | This study |
| pHUE- <i>SeM58-R126D</i> | M58-R126D | This study |
| pHUE- <i>SeM58-R164D</i> | M58-R164D | This study |
| pHUE- <i>SeM58-D172K</i> | M58-D172K | This study |
| pHUE- <i>SeM58-E246K</i> | M58-E246K | This study |
| pHUE- <i>SeM58-D249K</i> | M58-D249K | This study |
| pHUE- <i>SeM58-R251D</i> | M58-R251D | This study |
| pHUE- <i>SeM58-R252D</i> | M58-R252D | This study |
| pHUE- <i>SeM58-E286K</i> | M58-E286K | This study |
| pHUE- <i>SeM58-D294K</i> | M58-D294K | This study |
| pHUE- <i>SeM58-R298D</i> | M58-R298D | This study |
| pHUE- <i>SeM58-E303K</i> | M58-E303K | This study |
| pHUE- <i>SeM58-R367D</i> | M58-R367D | This study |
| pHUE- <i>SeM58-R481D</i> | M58-R481D | This study |
| pHUE- <i>SeM58-C4S</i> (C261S/C279S/C377S/C395S) | M58-C4S | This study |
| pHUE-CC _{TRI} M35 | CC _{TRI} M35 | This study |
| pHUE- <i>Se</i> γ CAL-2S(1-429) | γ CAL-2S | This study |
| pHUE- <i>Se</i> γ CAL-1S(1-313) | γ CAL-1S | This study |
| pHUE- <i>Se</i> γ CAL(1-198) | <i>Se</i> γ CAL(1-198) | This study |
| pHUE- <i>Se</i> γ CAL(1-181) | <i>Se</i> γ CAL(1-181) | This study |
| pHUE- <i>SeCcaA</i> | <i>SeCcaA</i> | This study |
| pHUE- <i>SeCcaA</i> Δ C2 | <i>SeCcaA</i> Δ C2 | This study |
| pHUE- <i>SeCcaA-C2</i> ₁₇ | <i>SeCcaA-C2</i> ₁₇ | This study |
| pHUE-E _{GFP} | E _{GFP} | This study |
| pHUE-E _{GFP} C2 ₁₅ | E _{GFP} C2 ₁₅ | This study |
| pHUE-E _{GFP} C2 ₁₇ | E _{GFP} C2 ₁₇ | This study |
| pHUE_ <i>Syn6301 ccmM</i> M35 | M35 | Ref. 15 |
| pET11a_ <i>Syn6301 rbcL</i> | RbcL ₈ | Ref. 45 |
| pET11a_ <i>Syn6301 rbcS</i> | RbcS | Ref. 45 |
| pHUE- <i>Se7942_rafI</i> | SeRafI | Ref. 47 |
| pEF- <i>gfp</i> | E _{GFP} | Addgene Cat#11154 |

4.4 Phase separation of Rubisco by folded SSUL modules of CcmM in β -carboxysome biogenesis (accepted)

Phase separation of Rubisco by the folded SSUL domains of CcmM in beta-carboxysome biogenesis

Huping Wang¹ and Manajit Hayer-Hartl^{1,*}

¹Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany

*Correspondence: mhartl@biochem.mpg.de (M.H-H.)

Running Head: Phase separation by structured domain repeats

Summary/Abstract

Carboxysomes are large, cytosolic bodies present in all cyanobacteria and many proteobacteria that function as the sites of photosynthetic CO₂ fixation by the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco). The carboxysome lumen is enriched with Rubisco and carbonic anhydrase (CA). The polyhedral proteinaceous shell allows the passage of HCO₃⁻ ions into the carboxysome, where they are converted to CO₂ by CA. Thus, the carboxysome functions as a CO₂ concentrating mechanism (CCM), enhancing the efficiency of Rubisco in CO₂ fixation. In β-cyanobacteria, carboxysome biogenesis first involves the aggregation of Rubisco by CcmM, a scaffolding protein that exists in two isoforms. Both isoforms contain a minimum of three Rubisco small subunit-like (SSUL) domains, connected by flexible linkers. Multivalent interaction between these linked SSUL domains with Rubisco results in phase separation and condensate formation. Here, we use Rubisco and the short isoform of CcmM (M35) of the β-cyanobacterium *Synechococcus elongatus* PCC7942 to describe the methods used for *in vitro* analysis of the mechanism of condensate formation driven by the SSUL domains. The methods include turbidity assays, bright-field and fluorescence microscopy, as well as transmission electron microscopy (TEM) in both negative-staining and cryo conditions.

Key Words: Cyanobacteria; Proteinaceous microcompartment; carboxysome; CO₂ concentrating mechanism (CCM); Rubisco; Rubisco activase; CcmM; Rubisco small subunit-like (SSUL); liquid-liquid phase separation (LLPS); cryo-electron microscopy.

1. Introduction

Three types of bacterial membrane-less microcompartments in bacteria have been extensively characterized and shown to function in regulating and optimizing specific metabolic pathways [1,2]. Each consists of an enzymatic core that is encapsulated by a selectively permeable protein shell [3]. One such microcompartment is the carboxysome found in cyanobacteria and in some chemotrophic bacteria. Carboxysomes are polyhedral bodies filled with the enzymes ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), the protein that catalyzes carbon fixation in photosynthesis, and carbonic anhydrase (CA) [4-6]. They are thought to have evolved as a consequence of the increase in oxygen concentration in the ancient atmosphere, to overcome the competition between oxygen and carbon dioxide as substrates for Rubisco [7]. The proteinaceous shell of the carboxysome is made of pentameric and hexameric oligomers. Some oligomers harbor a highly positively charged narrow pore in the center. The generated attractive force is purported to be the reason of selective permeability that allows the passage of dissolved CO₂ in the form of HCO₃⁻, but prevents the escape of non-charged CO₂ that is generated by CA inside the carboxysome [8]. As a result CO₂ is concentrated in the vicinity of Rubisco [6,9], avoiding the costly photorespiratory reactions with oxygen[10].

The form of Rubisco (form I) found in cyanobacteria, algae and plants is a hexadecameric complex consisting of eight large (RbcL; ~52 kDa) and eight small (RbcS, ~15 kDa) subunits. The large subunits are arranged as tetramers of anti-parallel dimers, forming the cube-shaped core of the enzyme, with four RbcS subunits bound at the top and four at the bottom [11] (Figure 1a). The form I Rubisco proteins are sub-categorized into form 1A to 1D based on the sequence of their RbcL subunits [12]. Interestingly, two distinct types of carboxysomes have evolved, differing in the type of Rubisco they encapsulate and their assembly mechanism: α- and β-

carboxysomes, containing 1A and 1B Rubisco, respectively [4-6]. More recently, it has been shown that in α -carboxysomes the ~900 residue intrinsically disordered protein CsoS2 is involved in aggregating the Rubisco during α -carboxysome biogenesis [13], while in β -carboxysomes the structured Rubisco small subunit-like (SSUL) domains of the protein CcmM (~540 residues) bind and aggregate Rubisco into biomolecular condensates [14,15]. In some β -cyanobacteria, a protein that is required for Rubisco's functional maintenance, Rubisco activase, also contains an SSUL domain, homologous to that of CcmM, which is used to 'piggyback' on Rubisco for recruitment into carboxysomes [16]. Although the Rubisco linker proteins of α - and β -carboxysomes, CsoS2 and CcmM respectively, have no sequence or structural homology to each other, their multivalent electrostatic interactions and the overlap in binding regions on Rubisco, suggest that they represent a remarkable example of convergent evolution [14,15,13].

Here we use the mechanism of the CcmM protein of the β -cyanobacterium *Synechococcus elongatus* PCC7942 (*Se7942*) in Rubisco condensate formation as a paradigm to illustrate the methods useful for investigating liquid-liquid phase separation (LLPS) [17-20] mediated by structured protein domains. This process represents the first step in pre-carboxysome biogenesis. CcmM exists as two isoforms: The full-length isoform (*SeM58*) consists of a N-terminal CA-like domain, followed by three SSUL domains connected by ~30 residue flexible linkers [21] (Figure 1b). The shorter isoform (*SeM35*), resulting from an internal ribosome binding site, consists only of the three SSUL domains (Figure 1b). For simplicity, we focus on the interaction of the short isoform of CcmM (*SeM35*) with its cognate Rubisco. The crystal structure of the SSUL module showed that although its backbone resembles that of RbcS, it has two distinctive features: a short helical insertion ($\alpha 2$) which replaces the A-B loop in RbcS, and the presence of a disulfide bond between Cys261 and Cys279 (Figure 1c). Both of these elements are critical for the interaction of

SSUL with Rubisco and condensate formation [14]. Cryo-electron microscopy (cryo-EM) and single particle analysis (SPA) revealed that the SSUL module binds asymmetrically in a cleft between RbcL dimers of Rubisco making critical salt-bridges between the short helix $\alpha 2$ of SSUL and the RbcL subunits and adjacent RbcS (Figure 1d), consistent with the salt sensitivity of the interaction. Furthermore, the apparent affinity of oxidized SeM35 for Rubisco was ~ 5 -fold lower than that of the reduced protein as monitored by turbidity assay at 340 nm, consistent with the observation of fewer and smaller liquid droplets by fluorescence microscopy [14]. Interestingly, the interaction is redox-regulated with tighter binding in the reducing environment of the cytosol and a more dynamic association in the oxidizing environment of the carboxysome [22,21,23]. The four-fold symmetry of Rubisco together with the three flexibly-linked SSUL domains provides the multivalency necessary to drive LLPS.

We first describe turbidity assays, bright-field microscopy and fluorescence microscopy as the methods used to monitor condensate formation during carboxysome biogenesis, and then discuss negative stain and cryo-EM as techniques that can provide structural insight into the underlying mechanism. Our focus is on understanding how multi-valent interactions with structured protein domains mediate phase separation, as opposed to the role that unstructured and low complexity sequences play in many other LLPS processes [17-20].

2. Materials

Prepare all solutions using dH₂O (ASTM Type 1 with 0.2 μm membrane filter purification system with a resistivity of 18.2 M Ω cm⁻¹ at 25 °C) and analytical grade reagents. Prepare and store all buffers at room temperature (~ 20 -25 °C), unless otherwise indicated.

2.1. Proteins

1. *Se*Rubisco stored as aliquots at $-80\text{ }^{\circ}\text{C}$ in storage buffer (50 mM Tris-HCl pH 8.0, 50 mM KCl, 10 mM Mg(OAc)₂, 10 % glycerol) at a concentration of $\sim 80\text{ }\mu\text{M}$. Holoenzyme concentration is determined spectroscopically using the extinction coefficient at 280 nm (ϵ_{280}) of $727400\text{ cm}^{-1}\text{ M}^{-1}$.
2. *Se*M35_{red} ($\sim 200\text{ }\mu\text{M}$; $\epsilon_{280}\text{ }29910\text{ cm}^{-1}\text{ M}^{-1}$) aliquots in storage buffer containing 5 mM DTT stored at $-80\text{ }^{\circ}\text{C}$.
3. *Se*M35_{ox} ($\sim 200\text{ }\mu\text{M}$; $\epsilon_{280}\text{ }30160\text{ cm}^{-1}\text{ M}^{-1}$) aliquots in storage buffer stored at $-80\text{ }^{\circ}\text{C}$.

2.2. Turbidity Assay

1. Assay buffer: 50 mM Tris-HCl pH 8.0, 10 mM Mg(OAc)₂ containing different concentrations of KCl with or without 5 mM DTT (see **Note 1**)
2. Machine: Jasco V-560 spectrophotometer set to $25\text{ }^{\circ}\text{C}$ and programmed for a time course measurement at 340 nm for 10 min (see **Note 2**), with data collection every 2 s.
3. Cuvette: Hellma 105.200-QS, volume 160 μL ; light path 10 mm (see **Note 3**).
4. Cuvette washer apparatus (Sigma-Aldrich) connected to a vacuum pump.
5. Eppendorf Thermomixer comfort set to $25\text{ }^{\circ}\text{C}$.
6. Low protein-binding microcentrifuge (MC) tubes (0.65 mL; Sorenson, Cat#11300).
7. Scientific Industries SITM Vortex-GenieTM 2.

2.3. Fluorophore Labeling of Proteins

1. Labeling buffer: 20 mM HEPES-KOH pH 8.0, 50 mM KCl and 10 mM Mg(OAc)₂ (see **Note 4**).

2. Reaction buffer: 50 mM Tris-HCl pH 8.0, 50 mM KCl and 10 mM Mg(OAc)₂, with or without 5 mM DTT.
3. Amine reactive fluorophore: Alexa Fluor® 532 NHS ester (succinimidyl ester) (henceforth named Alexa-532) (ThermoFisher Scientific; Cat# A20101MP; MW 723.8 Da; extinction coefficient 81000 cm⁻¹ M⁻¹ at 530 nm and a 0.09 reduction factor for absorbance at 280 nm). Excitation and emission maxima of 530 and 555 nm, respectively. Prepare a fresh solution of the fluorophore (~600 μM) in labeling buffer for immediate use (see **Notes 5 and 6**).
4. NT-650-NHS amine reactive fluorophore kit (NanoTemper; Cat#MO-L011): NT-650-NHS (MW1066 Da; extinction coefficient 195000 cm⁻¹ M⁻¹ at approximately 650 nm and a 0.04 reduction factor for absorbance at 280 nm; excitation/emission maxima ~650/670 nm). Prepare a fresh solution of NT-650-NHS by dissolving 10 μg in 25 μL 100% DMSO to a final concentration of ~600 μM (see **Note 7**).
5. Low protein-binding microcentrifuge (MC) tubes (0.65 mL; Sorenson, Cat#11300) (see **Note 8**).
6. Bio-Spin® P-6 Gel Columns, Tris Buffer (Bio-Rad Cat#7326227).
7. Benchtop centrifuge: Eppendorf 5427 R, temperature set to 4 °C.
8. Spectrophotometer: NanoDrop™ 1000 (ThermoFisher Scientific).

2.4. Fluorescence Microscopy

1. Assay buffer: 50 mM Tris-HCl pH 8.0, 10 mM Mg(OAc)₂ containing different concentrations of KCl with or without 5 mM DTT.

2. Leica Thunder inverted widefield microscope equipped with a LEICA HCX PL APO 63x/NA1.4 oil immersion objective with a coolLED pe-4000 LED source at 635 nm and 460 nm for fluorescence imaging and a 12 V/100 W halogen lamp for bright-field imaging (Wetzlar, Germany). The images were recorded by a Leica DFC9000 GTC camera.
3. Workstation for imaging (hardware configuration: Model: HP Z800; CPU: 2 x Intel Xeon X5650 2.67 GHz; Memory: 48 GB RAM; Graphic card: Nvidia GeForce GTX 1080 8 GB GDDR5X video memory; OS: Windows 10 Pro) is installed with operating software Leica LAS AF (3.2.1.9702) and Leica LAS X (2.0.0.14332.2).
4. μ -Slide Angiogenesis, uncoated polymer coverslip (Ibidi Cat# 81506).
5. Eppendorf Thermomixer comfort set to 25 °C.
6. Bovine Albumin Fraction V (BSA) (Serva Cat#11924.03). Prepare a 6 mg mL⁻¹ solution of BSA in assay buffer (see **Note 9**).

2.5. Negative stain Electron Microscopy Grid Preparation

1. EM buffer: 50 mM Tris-HCl pH 8.0, 10 mM Mg(OAc)₂, 5 mM DTT (see **Note 10**).
2. Negative-staining EM grids: Quantifoil Carbon support films 200 mesh (Quantifoil, Cat# N1-C73nCu20-01)
3. Prepare a 2% EM-grade uranyl acetate (UA) (Merck, Darmstadt, Cat#8473.0025) stock in deionized water (dH₂O) (see **Note 11**).
4. DUMONT ANTIAPAPILLA Reversing Tweezers (PLANO, Cat# T5301).
5. Plasma cleaner (Harrick Plasma, Cat# PDC-32G).
6. Blotting filter paper (Waterman, Cat# WH1005-125).
7. A desiccator for storage of the prepared grids.

8. Cell culture dish (ThermoFisher, Cat#150318).

2.6. Cryo-Electron Microscopy Grid Preparation

1. EM buffer: 50 mM Tris-HCl pH 8.0, 10 mM Mg(OAc)₂, 5 mM DTT (see **Notes 10 and 12**).
2. Cryo-EM grids: Quantifoil R2/1 300 mesh. (PLANO, Cat#S174-2).
3. Vitrobot Mark 4 (FEI) equipped with foam cryo dewar and metal ethane container. Set the Vitrobot to operate at 25 °C and 95% humidity.
4. Plasma cleaner (Harrick Plasma, Cat#PDC-32G).
5. Ethane gas.
6. Liquid nitrogen (N₂).
7. Blotting filter paper (Waterman, Cat# WH1001-055).
8. Vitrobot cryo-plunger tweezers (FEI, Ted Pella 47000-500)
9. Cryo-boxes (PLANO; Cat#160-40) for storage of cryo-EM grids.

2.7. Transmission Electron Microscopes

1. Titan Halo equipped with Falcon II direct electron detector, operating at 300 Kv.
2. Talos-Actica equipped with Falcon III direct electron detector, operating at 200 Kv.
3. Glacios equipped with Gatan K2 Summit direct electron detector, operating at 200 Kv.
4. Titan Krios equipped with Gatan K3 direct electron detector, operating at 300 Kv.

2.8. Cryo-Electron Microscopy Image Processing

Use of a local GPU workstation or a GPU/CPU cluster with following programs installed:

MotionCor2 (<https://emcore.ucsf.edu/ucsf-software>) [24]; CTFFIND4

(<https://grigoriefflab.umassmed.edu/ctffind4>) [25]; Gautomatch (<http://www.mrc-lmb.cam.ac.uk/kzhang/>); RELION (https://www3.mrc-lmb.cam.ac.uk/relion/index.php?title=Main_Page) [26]; Chimera (<https://www.cgl.ucsf.edu/chimera/>) [27].

3. Methods

All procedures carried out at room temperature unless otherwise specified.

3.1. Recombinant Protein Expression and Purification

The proteins *Se*Rubisco, *Se*M35_{ox} and *Se*M35_{red} were recombinantly expressed in *E. coli* and purified using previously published protocols [28,29,14].

3.2. Turbidity assay

The following protocol describes the detection of protein condensate formation by absorbance measurement at 340 nm at 25 °C using a UV/Vis spectrometer (see **Note 13**).

1. Remove aliquots of the purified proteins *Se*Rubisco, *Se*M35_{red} and *Se*M35_{ox} from storage at –80 °C, place on ice and allow to thaw.
2. Prepare dilutions in turbidity assay buffer of *Se*Rubisco, *Se*M35_{red} and *Se*M35_{ox} to final concentrations of 5 μM, 20 μM and 20 μM, respectively, and allow to equilibrate to 25 °C on a Thermomixer (see **Note 14**). Use low protein-binding MC tubes (see **Note 8**).
3. Pipette 100 μl of the assay buffer into a clean, dry cuvette (see **Note 3**). Avoid formation of air bubbles (see **Note 15**) and allow to equilibrate to 25 °C in the spectrophotometer. Then zero the absorbance signal of buffer at 340 nm (see **Notes 16 and 17**).

4. The procedure for preparing a turbidity measurement on a mixture of two proteins, in this case, *SeRubisco* and *SeM35_{red}* or *SeM35_{ox}*, is as follows:

First prepare 100 μL of 0.5 μM *SeRubisco* in assay buffer (with or without DTT) using a 0.5 mL low-binding MC tube. Next transfer 50 μL into a clean/dry cuvette containing 40 μL assay buffer (with or without DTT). Mix well by gentle pipetting up and down, and avoid formation of bubbles. Equilibrate to 25 $^{\circ}\text{C}$ for \sim 2-3 min in the spectrophotometer. Next, add 10 μL 20 μM *SeM35_{red}* or *SeM35_{ox}* to the cuvette, and immediately vortex (speed 8) for \sim 1-2 s to mix thoroughly (see **Note 18**). The final concentrations of *SeRubisco* and *SeM35_{red}* or *SeM35_{ox}* in the mixture are 0.25 μM and 2 μM , respectively.

5. Start the continuous measurement at 340 nm, collecting data at 1-2 s interval, for 10 min (see **Notes 2 and 19**).

The absorbance of the individual proteins alone, 0.25 μM *SeRubisco* and 2 μM *SeM35_{red}* or *SeM35_{ox}*, should always be included in a series of measurements (Fig. 2a).

The turbidity assay can also be used to obtain apparent affinities of the linker protein (*SeM35_{red}* or *SeM35_{ox}*) for the target (*SeRubisco*). This is done by keeping the concentration of the target protein constant and varying the concentration of the linker protein. For each concentration the absorbance signal is recorded after, for example, 10 mins of incubation (Fig. 2b). Furthermore, relative affinities of homologous repeat domains or peptide sequences of the linker protein can be analyzed either by generating truncation mutants of the linker protein or by measuring the ability of individual repeats to displace the full-length linker protein [14]. Such an analysis also provides insight into the dynamics of the condensate.

3.3. Fluorophore Labeling of Proteins for Fluorescence Microscopy

To monitor LLPS by fluorescence microscopy, the proteins have to be labeled by fluorophores that differ in their excitation/emission maxima within the spectral range of visible light (wavelengths 380 to 750 nm). Generally, fluorescent dyes react with amine or thiol groups of amino acids. Since *SeM35* contains functionally important cysteines, we selected the fluorescent dye NT-650-NHS for preferentially labeling the amino group at the N-terminus. Similarly, the fluorescent dye Alexa-532 was selected for labeling an amino group of the *SeRubisco* holoenzyme (see **Note 20**).

1. Remove aliquots of the purified proteins *SeRubisco*, *SeM35_{red}* and *SeM35_{ox}* from storage at $-80\text{ }^{\circ}\text{C}$, place on ice and allow to thaw.
2. Since the protein stocks are stored in Tris buffer, for efficient amino group labeling the proteins were first buffer exchanged (see **Note 21**) to labeling buffer using Bio-Spin P-6 columns following the manufacturer's instructions (Bio-Rad Document# 4006051).
3. *SeRubisco*, *SeM35_{red}* and *SeM35_{ox}* protein stocks were prepared by diluting the proteins in labeling buffer to final concentrations of $10\text{ }\mu\text{M}$, $40\text{ }\mu\text{M}$ and $40\text{ }\mu\text{M}$, respectively, in a volume of $50\text{ }\mu\text{L}$. Use low protein-binding MC tubes (see **Note 8**).
4. The concentrations of the buffer-exchanged proteins are determined using NanoDrop™ 1000 spectrophotometer and the program "Protein A280". The molar concentration is calculated based on the extinction coefficients of the respective proteins (see **section 2.1**).
5. Determine also the concentration of the freshly prepared fluorescent dye solutions (see **Notes 5-7**) using NanoDrop™ 1000 spectrophotometer. The molar concentration is calculated based on the extinction coefficients of the respective dyes (see **section 2.3**).
6. To label a protein, for example *SeM35_{red}* or *SeM35_{ox}*, with the fluorophore NT-650-NHS follow the manufacturer's instructions in the kit. In brief:

- a) Prepare 100 μL of a 3-fold higher concentrated fluorescent dye solution relative to the concentration of protein to be labeled, using the labeling buffer provided in the NanoTemper fluorophore kit.
 - b) Mix 1:1 (v/v) protein and dye in low-binding MC tube.
 - c) Incubate the labeling reaction in the dark at 25 $^{\circ}\text{C}$ for 1 h.
7. To label the second protein, for example *SeRubisco*, with the fluorophore Alexa-532, first calculate the volume of Alexa-532 required using the following equation to obtain a molar ratio of dye to N-terminal amino groups of 2:1,

$$\text{Vol}_{\text{Rubisco}} * \text{Conc}_{\text{Rubisco}} * 8 * 2 = \text{Vol}_{\text{Dye}} * \text{Conc}_{\text{Dye}},$$

where $\text{Vol}_{\text{Rubisco}}$ and Vol_{Dye} refer to the volume of Rubisco and dye used, respectively, and $\text{Conc}_{\text{Rubisco}}$ and Conc_{Dye} the respective concentrations of Rubisco and dye.

8. Pipette the protein solution and dye into the labeling buffer (final volume 100 μL), mix thoroughly, and incubate the labeling reaction in the dark at 25 $^{\circ}\text{C}$ for 1 h.
9. The labeling reactions are stopped by rapid buffer exchange using a Bio-Spin 6-P Tris column pre-equilibrated in reaction buffer (Bio-Rad Document# 4006051). Elute the respective labeled protein at 800 x g, 4 min (see **Note 22**).
10. Determine the protein concentration and dye concentration using NanoDropTM 1000 spectrophotometer. In the “Dye/Chromophore list” insert the extinction coefficients and the reduction factor for the respective dyes. Then use the program “Proteins & Labels” to measure and readout the concentrations (see **Notes 23-24**). Labeling efficiency is calculated by dividing the dye to protein concentration. Generally we obtain a labeling efficiency of ~260% for ~2.5 μM *SeRubisco* and ~60% for ~10 μM *SeM35* (see **Note 25**).

11. The freshly labeled proteins can be used immediately or aliquoted and stored at $-80\text{ }^{\circ}\text{C}$, and used within a few months.

3.4. Bright-Field and Fluorescence Microscopy

As mentioned above, the turbidity assay does not directly report condensate formation. LLPS (or demixing) must be confirmed by another technique, for example by the observation of liquid droplets monitored by light microscopy. Bright-field microscopy relies on the differences in absorption of light due to differences in densities between various parts of a sample, and thus may be used to initially distinguish between protein aggregation (no defined structure) and condensate (round droplets) formation. Fluorescence microscopy, on the other hand, has the advantage of higher resolution, as well as allowing one to collect images in more than one color. This permits the visualization and confirmation of LLPS of proteins labeled with different fluorophores within the same liquid droplet.

1. Remove aliquots of the purified unlabeled and fluorophore-labeled proteins *SeRubisco*, *SeM35_{red}* and *SeM35_{ox}* from storage at $-80\text{ }^{\circ}\text{C}$, place on ice and allow to thaw, or use freshly labeled proteins.
2. Use low-protein binding MC tubes in all steps (see **Note 8**).
3. Prepare dilutions of unlabeled *SeRubisco*, *SeM35_{red}* and *SeM35_{ox}* in assay buffer to final concentrations of $10\text{ }\mu\text{M}$, $40\text{ }\mu\text{M}$ and $40\text{ }\mu\text{M}$, respectively, in a volume of $50\text{ }\mu\text{L}$. Place on ice until ready to use.
4. Prepare dilutions of the fluorophore-labeled *SeRubisco*, *SeM35_{red}* and *SeM35_{ox}* in assay buffer to final concentrations of $2.5\text{ }\mu\text{M}$, $10\text{ }\mu\text{M}$ and $10\text{ }\mu\text{M}$, respectively, in a volume of $25\text{ }\mu\text{L}$, and place on ice until ready to use.

5. Switch on power of Leica Thunder inverted widefield microscope, the control computer and set the temperature of the microscope stage to 25 °C.
6. It is highly recommended to first screen the formation of the protein droplets using bright field. For this, prepare mixtures containing 0.25 μM unlabeled Rubisco and 1 μM M35_{red} or M35_{ox}. One at a time, the mixtures (a-e) are prepared and analyzed following steps 7-9 (see **Note 26**).
 - a) 2.5 μM unlabeled Rubisco = 15 μL assay buffer + 5 μL of 10 μM unlabeled Rubisco.
 - b) 10 μM unlabeled M35_{red} = 15 μL assay buffer + 5 μL of 40 μM unlabeled M35_{red}.
 - c) 10 μM unlabeled M35_{ox} = 15 μL assay buffer + 5 μL of 40 μM unlabeled M35_{ox}.
 - d) Rubisco/M35_{red} reaction = 16 μL assay buffer + 2 μL of 2.5 μM unlabeled Rubisco + 2 μL of 10 μM unlabeled M35_{red}.
 - e) Rubisco/M35_{ox} reaction = 16 μL assay buffer + 2 μL of 2.5 μM unlabeled Rubisco + 2 μL of 10 μM unlabeled M35_{ox}.
7. Incubate the mixture at 25 °C for 5 min using a Thermomixer.
8. Pipette the mixture into a well of the uncoated coverslip (μ -Slide Angiogenesis; Ibidi) pretreated with BSA, cover with lid, transfer to microscope stage and incubate for a further 5 min at 25 °C (see **Notes 27 and 28**)
9. After a total incubation time of 10 min, take a snapshot with the camera focused at the bottom of the slide. Figure 3a displays a typical image of liquid droplets as visualized by bright field.

10. To obtain a higher resolution image and to investigate whether both proteins are colocalized within the liquid droplets, fluorescence microscopy of the differently labeled proteins is performed.
11. Prepare mixtures of 10:1 ratio of unlabeled:labeled protein (see **Note 29**) as follows:
- a) 2.5 μM 10 unlabeled:1 labeled Rubisco = 13.5 μL assay buffer + 4.5 μL of 10 μM unlabeled Rubisco + 2.0 μL of 2.5 μM labeled Rubisco.
 - b) 10 μM 10 unlabeled:1 labeled M35_{red} = 13.5 μL assay buffer + 4.5 μL of 40 μM unlabeled M35_{red} + 2.0 μL of 10 μM labeled M35_{red}.
 - c) 10 μM 10 unlabeled:1 labeled M35_{ox} = 13.5 μL assay buffer + 4.5 μL of 40 μM unlabeled M35_{ox} + 2.0 μL of 10 μM labeled M35_{ox}.
12. One at a time, prepare mixtures (a and b) containing 0.25 μM of 10 unlabeled:1 labeled Rubisco and 1 μM of 10 unlabeled:1 labeled M35_{red} or M35_{ox} following the procedure as in steps 7 to 9 (see **Notes 26 and 30**).
- a) Rubisco/M35_{red} mixture = 16 μL assay buffer + 2 μL of 2.5 μM of 10 unlabeled:1 labeled Rubisco + 2 μL of 10 μM 10 unlabeled:1 labeled M35_{red}.
 - b) Rubisco/M35_{ox} mixture = 16 μL assay buffer + 2 μL of 2.5 μM of 10 unlabeled:1 labeled Rubisco + 2 μL of 10 μM 10 unlabeled:1 labeled M35_{ox}.

Figure 3b displays the demixing of two differentially labeled proteins into round fluorescent condensates, as visualized in the merged image. The absorbance of the individual proteins alone, 0.25 μM of unlabeled:labeled (10:1) Rubisco and 1 μM of unlabeled:labeled (10:1) M35_{red} or M35_{ox}, should always be included as controls showing a diffuse distribution of the proteins on their own (Fig. 3c).

Phase separation assays can also be used to obtain additional information about the process, for example, the size distribution of the droplets can be analyzed and the average Feret's diameter of the distribution determined using the software Fiji [30]. Moreover, analysis of fusion events can be monitored by taking time lapse images of the droplets, and the technique of fluorescence recovery after photobleaching (FRAP) to obtain information on the dynamics of the respective proteins within the condensate [31].

3.5. Negative Stain Electron Microscopy

Negative-staining is an established method in which the protein sample is adsorbed onto an EM grid substrate, then a heavy metal stain such as uranyl acetate is applied, blotted, and dried to produce a thin layer of electron dense stain in which protein particles are embedded [32,33]. Although, negative stain EM results in better contrast of the protein particles to the background, the relatively large particle size of the heavy metal limits the resolution that can be attained (a maximum of 18-20 Å). Nevertheless, negative-staining provides a fast screening method for optimizing the protein concentration for homogeneous distribution and structural integrity of particles, before cryo-EM grid preparation.

Here we describe the steps to prepare grids for the reaction between Rubisco and *SeM35_{red}*.

1. Remove aliquots of the purified proteins Rubisco and *M35_{red}* storage at $-80\text{ }^{\circ}\text{C}$, place on ice and allow to thaw.
2. Using the fine-tweezers, place several of the carbon-coated grids on a clean glass slide with the carbon side up (dark side of the grip up) (see **Notes 31 and 32**).

3. Transfer the glass slide with the grids to a plasma cleaner chamber to remove contaminants and to make the grids hydrophilic. Operate the machine for 15 s at medium power (see **Notes 33 and 34**).
4. Store the slide containing the plasma cleaned grids until ready to use (see **Note 35**).
5. Prepare 80 μM and 200 μM solutions of Rubisco and M35_{red}, respectively, in EM-buffer.
6. Prepare a mixture containing 6.25 μM Rubisco and 50 μM M35_{red} as follows:

26.9 μL EM buffer + 3.1 μL of 80 μM Rubisco + 10 μL of 200 μM M35_{red}

7. Incubate the mixture at 25 °C for 10 min using a Thermomixer.
8. Dilute the mixture 10-fold with EM buffer (equilibrated at 25 °C) and place in the 25 °C thermomixer.
9. Using the reverse EM tweezer pick one of the freshly plasma cleaned grid, and with the grid held by the tweezer perform the next steps.
10. Apply 5 μL of the 10-fold diluted protein mixture on the carbon coated side of grid.
11. Incubate for 1 min to allow the protein sample to settle on the EM grid (see **Note 36**).
12. Blot the excess protein sample with the filter paper held at the edge of the grid (see **Note 37**).
13. Wash the grid with 5 μL staining buffer (2 % UA). Blot the excess UA by filter paper.
14. Next apply 5 μL the 2% UA to the grid, incubate for 1 min to allow staining and blot the excess UA from the grid as above.
15. Finally, place the grid, with carbon side up, on a filter paper in a cell culture dish. Cover the dish and store in a desiccator overnight or for at least 12 h (see **Note 38**).

16. The following day, the grid is loaded onto the single tilt holder and inserted into Titan halo Specs microscope for screening. The microscope is controlled by SerialEM [34] for image collection.

Figure 4a shows the negative stain image of the mixture of $\sim 0.63 \mu\text{M}$ Rubisco and $5 \mu\text{M}$ M35_{red}, with the protein particles in white and the background grey (see **Note 39**). The Rubisco particles (~ 550 kDa) are clearly discernable, while the SSUL domains (~ 11 kDa) of M35 are too small to be seen in the micrograph. Importantly, the protein Rubisco complexes are relatively well-distributed, indicating good sample conditions for preparation of cryo-EM grids (see **Note 40**).

3.6. Cryo-Electron Microscopy

Electron microscopy of frozen hydrated samples (cryo-EM) combined with single-particle analysis (SPA) is a rapidly developing method to visualize macromolecular complexes in an almost physiological state. Recent advances in direct electron detector technologies, improved sample preparation protocols and the development of new data processing algorithms has resulted in making cryo-EM a powerful structure determination technique [35]. In cryo-EM the proteins or macromolecular complexes are preserved in a thin layer of vitreous ice and, in contrast to negative stain-EM, the signal of the image results directly from the protein. Thus cryo-EM allow images of exceptional quality to be recorded and structures to be determined at near-atomic resolution. For further reading on performing high-resolution cryo-EM the following references are highly recommended [36-41], and the online course for beginners to the field provided at <https://cryoem101.org/> (NIH Common Fund CryoEM initiative). Despite many advances in the automation of the cryo-EM microscopes, performing cryo-EM and SPA requires

extensive training, beginning with sample preparation, handling of microscopes and finally data processing.

In this article, we describe the protocol we used for investigating the weak and transient multivalent interactions between Rubisco and M35, resulting in LLPS [14].

1. Remove aliquots of the purified proteins Rubisco and M35_{red} from storage at $-80\text{ }^{\circ}\text{C}$, place on ice and allow to thaw.
2. For plunge freezing of the cryo-EM grids, we use the semi-automated Vitrobot Mark 4 (FEI) (see **Notes 41 and 42**).
3. Switch on the Vitrobot and set the chamber temperature to $25\text{ }^{\circ}\text{C}$ and 90% humidity.
4. Mount fresh filter papers on both sides of the blotting pads.
5. Using the fine-tweezers, place several of the cryo-EM grids on a clean glass slide with the carbon side up (see **Notes 31 or 32**).
6. Transfer the glass slide with the cryo-EM grids to a plasma cleaner chamber to remove contaminants and to make the grids hydrophilic. Operate the machine for 30 s at medium power (see **Notes 33-35**).
7. Cool down the foam cryo dewar, containing the metal container for liquid ethane and cryo boxes for grid storage, with liquid N_2 ($-196\text{ }^{\circ}\text{C}$) (see **Note 43**).
8. When the metal container for liquid ethane has cooled down for at least 15 min by the liquid N_2 , start filling it with ethane gas which liquifies at around $-88.6\text{ }^{\circ}\text{C}$. Stop filling when the liquid ethane reaches the top of the container and a thin layer of ethane starts to solidify on the surface of the metal container. This happens when the temperature has reached approximately $-182.8\text{ }^{\circ}\text{C}$ (see **Note 44**).

9. Prepare the mixture containing 6.25 μM Rubisco and 50 μM M35_{red} as in steps 5 to 7 for negative stain EM above.
10. Mount a freshly discharged grid (plasma cleaned) on the Vitrobot cryo-plunger tweezer and attach the tweezer-grid assembly to the Vitrobot robotic arm. Instruct machine to move the arm into the Vitrobot chamber.
11. Place the foam cryo dewar containing the liquid ethane with the thin crust, which is ready for freezing the sample, onto the Vitrobot stage and instruct the program to move the stage into freezing position.
12. Next instruct the machine to move the tweezer assembly to sample loading position, and apply 3 μl of the reaction (incubated for 10 min at 25 $^{\circ}\text{C}$) onto the carbon coated side of the grid (see **Notes 45 and 46**).
13. Start the semi-automatic blot and freezing procedure with customized blot force and blot time (see **Note 47**). After blotting according to the pre-set parameters, the robotic arm plunges the tweezer-grid assembly into the liquid ethane.
14. While keeping the grid in the liquid ethane, unmount the tweezer-grid assembly from the robotic arm.
15. Transfer the grid into the cryo-box which is under liquid N_2 in the foam cryo dewar.
16. Repeat steps 10-15 for all the other grids, each time using slightly different conditions, for example, changing the blot force (or blot time) for each grid.
17. Store the cryo-boxes containing the frozen grids in a liquid N_2 storage dewar (see **Note 48**).

18. To check the suitability of the frozen sample grids for high resolution data collection, the grids are first screened on either the transmission electron microscopes (TEM) Talos Actica or Glacios, operating at 200 kV and equipped with autoloader.
19. To load the frozen grids into the autoloader, they are first assembled in “auto-grid” cartridges. Autogrids are assembled in dedicated workstations by seating cryo-EM grids over a clip ring and then locking the grid and clip ring together with a C-shaped spring, or C-clip (see **Note 49**).
20. The clipped grids are then carefully inserted into slots within the liquid N₂ cooled autoloader cassette (cassettes, containing up to 12 Autogrids).
21. Cool the cryo-stage and cryo-nanocup with liquid N₂ (−196 °C).
22. Transfer the autoloader cassette into a cryo-nanocup and mount the cryo-nanocup on the TEM sample loading stage.
23. Next load the cassette with grids into the cryo-autoloader.
24. When the temperature of the cryo-autoloader reaches −165 °C, start autoloader inventory (see **Note 50**).
25. Run the SerialEM [34] script to automatically collect grid maps of all the grids (see **Note 51**).
26. Switch on low-dose mode in SerialEM.
27. Select an area of grid displaying thin vitreous ice and move the stage to that position (see **Note 52**).
28. Then using “search” mode, find the proper Z high position of the stage using the automated script in SerialEM or adjust the position manually.

29. Center the stage into one hole under “search” mode. Set the autofocus position on support carbon to obtain enough signal. Set target defocus to -2 μm from the back focal plane and perform autofocus.
30. Check “record” setting and take one image in “record” mode.
31. Repeat steps 27-30 for all the grids or until a suitable area for test data collection has been found for a grid (see **Notes 53 and 54**).
32. Upon finding a grid with perfect particle distribution and ice condition, it is transferred to the high-voltage (300kV) TEM Titan Krios for high resolution data collection.
33. To load the grid into the cryo-autoloader Titan Krios follow steps 21-23 and follow steps 24-25 for autoloader inventory and grid map collection, respectively.
34. Align the microscope carefully with the help from your EM manager. Check the beam and camera settings in the different illumination modes (especially: “search”, “focus” and “record”) in low-dose.
35. Set up for high quality data collection in SerialEM. The magnification used in the “record” mode yields a pixel size with 0.822 Å. Set up to collect movies at 0.15 s per frame and in total 7 s exposure. Measure the dose rate under current illumination condition as instructed for the microscope and camera. The automated data collection is set to perform 3*3 pattern and two different exposures per-hole.
36. Select squares (see **Note 54**) with the appropriate thickness of ice and start the data collection. During the procedure the cryo-stage remains stagnant over one 3*3 pattern of holes and collects 18 movies from 9 holes (2 exposures for each hole). The stage then moves to another 3*3 pattern of holes and so on, until a significant number of movies have been collected (see **Notes 55 and 56**).

Figure 4b shows the cryo-EM image of the mixture of $\sim 6.3 \mu\text{M}$ Rubisco and $50 \mu\text{M}$ M35_{red}, with the protein particles in dark grey and the background light grey (see **Note 57**). The top view of Rubisco has a well defined four-fold symmetric shape with a bright hole in the middle. Some of the Rubisco particles can be observed as clustered with a constant distance between particles, indicating they are in a protein network. The SSUL domains (~ 11 kDa each) of M35 are too small to be directly detected on a micrograph. (Fig. 4b).

3.7. Cryo-Electron Microscopy Image Processing

Single particle analysis (SPA) is a method that studies *in vitro* the structure of isolated macromolecules and complexes. It is based on alignment and averaging of many thousands of images of individual complexes followed by 3D reconstruction [36]. Thus, processing of the recorded movies consumes another major part of the workload of a SPA project, besides optimizing sample grid preparation. The main steps include motion correction to generate one micrograph from a raw movie stack, estimation of the CTF parameters for individual micrographs, particles coordinate determination, generation of an initial reference, structural heterogeneity classification, particles alignment in refinement, interpretation of the final 3D density map and assessment of resolution. A local GPU work station or a CPU/GPU cluster and several software packages are required for data processing. Described here briefly is the workflow applied for the *Se*Rubisco-*Se*M35_{red} complex analysis.

1. The MotionCor2 [24] software was used in 5x5 patches (see **Note 58**) with dose-weighting function to generate one micrograph from each movie.
2. Software CTFFIND4 [25] was used to estimate the Contrast Transfer Function (CTF) parameters in each micrograph.

3. Images with melted ice and bad quality were removed based on power spectrum (see **Notes 59 and 60**).
4. Particles picked using Gautomatch (<http://www.mrc-lmb.cam.ac.uk/kzhang/>) and templates from the screening data-set (see **Note 53**).
5. Performed 2D classification to remove bad particles with the software RELION 2.1 [26].
6. An initial model was generated by RELION de novo 3D model generation module with D4 symmetry (see **Note 61**).
7. Performed one round of 3D classification with D4 symmetry using the initial model as reference and applied a low pass filter to 60 Å. The classes with good structural features were selected for further refinement.
8. To sort out multiple SSUL domains of M35 bound per Rubisco holoenzyme, we applied the symmetry expanding protocol. In brief, particles were first aligned with D4 symmetry in an auto-refine job. Using the `relion_particle_symmetry_expand` command, the particle star file (`run_data.star`), the output from refinement, was used to generate eight-fold aligned particles. The command is as follows:

```
relion_particle_symmetry_expand --i run_data.star --o expanded_data.star --sym D4
```

A particle subtraction was performed, using the symmetry expanded particle star file as the input, with a mask covering 2 RbcL, 2 RbcS and 2 SSUL, to take each asymmetric unit as individual particles.
9. Select best homogeneous particles from a focused classification job with subtracted particle.

10. Run one round of auto-refine with only local sampling. This refinement output was used to improve particle signal to noisy ratio in CTF refinement and particle polishing to generate “shiny” particles.
11. Run the final round of auto-refine with “shiny” particles and further improve the reconstruction quality by post-processing with a soft mask applied.

The extra density corresponding to the SSUL domain could already be observed clearly in some 2D classes (step 5) (Fig. 5a). Generally the 3D classification results in several classes, with at least one class displaying well-defined structure and one class displaying undefined structure (step 7), indicating that the classification was successful (Fig. 5b). In the final EM density map at 2.8 Å resolution of the Rubisco-M35 complex, the sidechain electron densities of residues Arg251 and Arg252 of the SSUL module of M35 are clearly discernible. They are stabilized by formation of salt bridges with two Asp residues in Rubisco (Asp76 of RbcL and Asp93 of RbcS) (Fig. 5c).

4. Notes

1. Assay buffer is freshly prepared using the following stock solutions 1 M Tris-HCl pH 8.0, 1 M Mg(OAc)₂, 3 M KCl and 1 M DTT. All stock solutions are stored at R.T., except for DTT which is stored as aliquots of 0.5 mL at –80 °C.
2. The measurement time has to be adjusted such that the absorbance signal reaches a plateau (saturation). This depends on several parameters, for example, protein concentration temperature, pH and salt concentration.
3. The cuvette has to be clean and dry. To clean a cuvette, incubate overnight in 2 ml 6 M urea solution to remove any protein adherent to the cuvette. Using a cuvette washer

connected to a vacuum pump, wash several times with dH₂O water, followed by 95% ethanol and let the cuvette dry completely. The cuvette is also washed after each measurement.

4. HEPES, MES, MOPS and phosphate-based buffers or any other buffer free of ammonium ions or primary amines may be used for labeling. Note that the labeling is performed in the absence of DTT to avoid interference in the labeling reaction.
5. Using a 200 μ L tip, scratch a small amount of Alexa Fluor® 532 fluorescent dye from the vial (~10 μ g) into a 0.5 mL low protein-binding MC tubes. Dissolve the Alexa Fluor® 532 in ~20 μ L labeling buffer to a final concentration of ~600 μ M, and use freshly for labeling.
6. Alexa Fluor® 532 is highly water soluble. However, it is somewhat unstable upon exposure to moisture, hydrolyzing into the non-reactive free acid in aqueous solutions.
7. NT-650-NHS fluorescent dye is purchased as 10 μ g aliquots, and stored at -80 °C.
8. Low protein-binding MC tubes are used to prevent unspecific binding to tube walls and hence, loss of protein.
9. 20 μ l aliquots of 6 mg mL⁻¹ BSA in dH₂O, stored at -80 °C.
10. As a general rule in EM analysis, especially for cryo-EM, it is recommended to keep the concentration of salt as low as possible, ideally below ~300 mM, in order to minimize contrast problem.
11. UA is light sensitive, and can be stored as aliquots in a container wrapped with aluminum foil for up to 1 year at 4°C.
12. In high resolution cryo-EM the presence of >5% glycerol greatly affects the contrast in the micrographs. If possible glycerol should be avoided in the sample buffer.

13. Turbidity assay does not directly report condensate formation. Condensate formation must be confirmed by another technique, e.g. the formation of liquid droplets as monitored by light microscopy.
14. The reaction rate is highly temperature sensitive, and for consistency in measurements the assay buffer and protein samples should be pre-equilibrated at 25 °C.
15. Avoid air bubbles, as they create a very unstable background scatter in the absorbance signal.
16. The detection wavelength may be set either at 320 or 340 nm. The shorter wavelength is more sensitive for smaller aggregates (scattering $\propto \lambda^{-4}$).
17. If necessary, check signal stability of the spectrophotometer by performing a continuous measurement of the assay buffer for the time period used in the turbidity assay.
18. Since the initial slope of the reaction can be fast, a quick and thorough mixing avoids delays in the start of measurements. It is important to avoid long vortexing times as this can result in the formation of tiny bubbles that would affect the intensity of the measured signal.
19. The interval time for data collection will depend on the rate of the reaction. For a fast reaction, shorter interval time will result in a more accurate rate determination.
20. The selected fluorescent dyes should ideally have non-overlapping excitation and emission wavelengths to avoid fluorescent quenching and bleed-through.
21. If the purified protein sample contains amino or thiol group containing chemicals such as, free ammonium ions, Tris buffer, imidazole, DTT or β -mercaptoethanol (use TCEP if reducing agent is required), a buffer exchange of the protein is required depending on the amino acid group being labeled. This is to ensure efficient labeling of the protein of

interest. Also, partially purified protein samples or protein samples containing carriers like BSA should not be used for labeling.

22. The free unreacted dye remains trapped in the column.
23. In the absence of a spectrophotometer that has a program that calculates the concentration of the labeled protein, the following equation is used to calculate the protein concentration of the sample:

$$\text{protein concentration (M)} = \frac{[A_{280} - (A_{530 \text{ or } 647} \times 0.09 \text{ or } 0.04)] \times \text{dilution factor}}{\text{molar extinction coefficient of protein}}$$

where 0.09 is the correction factor for Alexa-532 and 0.04 for NT-650-NHS to account for absorption of the respective dyes at 280 nm.

24. In the absence of a spectrophotometer that has a program that calculates the concentration of the dye, the following equation is used to calculate the degree of labeling:

$$\text{Moles dye per mole protein} = \frac{(A_{530 \text{ or } 647} \times \text{dilution factor})}{(81000 \text{ or } 195000) \times M}$$

where 81000 cm⁻¹ M⁻¹ is the molar extinction coefficient of the Alexa-532 dye at 530 nm, 195000 cm⁻¹ M⁻¹ is the molar extinction coefficient of NT-650-NHS at 650 nm, and M is the protein concentration.

25. Ideally, labeling should result in only 1-2 dyes per protein, and the protein structure or function should not be affected by the attached dye. This ought to be confirmed by circular dichroism spectroscopy and/or a functional assay.
26. Prepare each reaction mix sequentially, as an experiment is completed (steps 6 to 9).
27. Avoid air bubbles when pipetting the reaction into a well of the coverslip. Air bubbles will affect image contrast.
28. To reduce unspecific interaction of the proteins with the well of an uncoated coverslip (μ-Slide Angiogenesis; Ibidi; 15 wells of 50 μL), pretreat the well with 15 μL of BSA (6 mg

mL⁻¹) for 1 min and wash two times with 20 μ L assay buffer. Remove most of the wash buffer from the well before adding the reaction.

29. It is advisable to use predominantly unlabeled protein with only a small percentage of the fluorescently tagged version to avoid effects of the fluorescent tag on phase separation properties, which are generally charge sensitive.
30. The exposure time should be adjusted based on labeling efficiency. We generally use 200 ms.
31. Grip the grid at the edge with the tweezers. Avoid contact with the grid center and tearing the carbon coating.
32. There are several ways to determine which side of the grid is carbon coated. If the vendor is Quantifoil or PLANO, the grids containing box has a row of labels that point to the carbon side. One can also look at the grid under light. The carbon side will have a colorful diffraction, while the metal copper side will be shining.
33. Open the valve of the chamber slowly to avoid grids being blown away.
34. The carbon thickness on EM grids varies between vendors and also between batches of the same vendor. If application of medium power for 15 s results in the breakage of several meshes on the grids, use lower power for a longer time to ensure that the carbon coating remains intact.
35. The plasma cleaned grids should be used within 1 h after plasma cleaning.
36. A shorter or longer time of incubation will result in slightly less or more of the protein sample being deposited on the grid surface.
37. Be careful not to over blot the sample to dryness, and not to lose the grid from the tweezers.

38. It is possible to accelerate drying of the grids by using a heating lamp. However, this may lead to over-drying and positive staining.
39. When a drop of the protein condensate is applied on a EM grid and blotted, the strong surface tension results in breaking the round protein droplets and what remains on the grid are several protein clusters, not the intact droplet.
40. The observed cluster size is highly dependent on the concentration of the proteins used.
41. Ensure that the Vitrobot cryo-plunger tweezers (FEI, Ted Pella 47000-500) are sharp by examining under an optical microscope. Also check that they are not bent.
42. Link to a youtube movie showing the standard procedure of vitrification using a Vitrobot apparatus: https://www.youtube.com/watch?v=QML_KMQbOMc.
43. Both liquid N₂ and liquid ethane can cause severe burns on the hands/body and blindness if splashed in the eye. Always wear lab coat, gloves and glasses when handling either of them.
44. Replenish the liquid N₂ in the foam cryo cup as needed during the whole procedure of freezing the sample.
45. Make sure the pipet tip does not touch the grid, only the liquid droplet should make contact with the surface of the grid.
46. In cryo-EM, the grid is normally covered with carbon support and in the carbon structure there are empty holes with a certain size. Unfortunately, most of the protein is absorbed by carbon support and only a small portion distributes into the holes of the grid.
47. The blot force is normally specific to individual Vitrobot machines. The blot force and blot time is adjusted dependent on many factors, such as grid type, protein concentration and buffer conditions. Beware that a more positive blot force increases the force. A good

strategy is to prepare many grids, for example 12, with the same protein sample and buffer, keeping the loading volume and blot time constant and only varying the blotting force (for example, -1 to 4). One grid with nice thin vitreous ice would be sufficient to allow data collection.

48. The grids at this point can either be stored in labeled cryo-boxes in a liquid N₂ storage dewar or be clipped and inserted into the autoloader cassette and screened.
49. Link to a youtube movie showing the standard procedure for grid clipping:
<https://www.youtube.com/watch?v=3VjOtrZ-76I>
50. 'Autoloader inventory' is a procedure in which the 'gripper' of the autoloader works as a robotic arm to grip each slot in the cassette and to sense whether a grid is present. If a grid is present, the cassette status is then uploaded into the microscope.
51. The grid map can be saved as a record of ice thickness of the different blotting conditions used during vitrification.
52. Many vitrified grids will show a gradient of ice thickness.
53. The preliminary 2D classes are a standard evaluation of whether the sample on a particular grid displays a preferred orientation and if it is homogeneous. 2D classes showing some secondary structural features is a good indication that high resolution could be achieved with the sample. The best grid is saved and used for data collection on the high-voltage (300kV) TEM Titan Krios.
54. Selected squares for data collection should be free from cracks, crystalline ice and contamination. The Quantifoil grid used here has a total of 300 meshes, equivalent to 300 squares.

55. The raw data is the ‘movie’ in which 47 frames are collected at 0.15 s per frame, resulting in one exposure of 7 s. For a standard preliminary screening dataset ~1,000 movies are generally required, which are collected overnight. The number of movies is highly dependent on the sample.
56. The development of a direct electron detector (DED) allows direct collection of movies and together with recent software, such as MotionCor [24], has dramatically improved the data quality.
57. Similar to the disruption of the round protein droplets as in negative-stain EM upon blotting, the droplets are also disrupted due to strong surface tension upon freezing. Only one layer of the 2-dimensional protein network is preserved in vitreous ice.
58. Analysis in MotionCor performed by dividing each frame into 5*5 patches and the stage movement estimated for each patch. For more customized parameters, please refer to the MotionCor online manual (<https://emcore.ucsf.edu/ucsf-software>) [24].
59. Power spectrum refers to the two-dimensional image of individual micrographs in reciprocal space. If the micrograph displays a high signal around 3.6 Å, this reflects the diffraction of ice-crystals, suggesting that the vitreous ice had melted. Moreover, a bad quality image will have lost the high frequency “thick rings”.
60. The accurate estimation of contrast transfer function (CTF) parameters is important for both the initial evaluation of micrograph quality and subsequent structure determination. In the software CTFFIND4 (<https://grigoriefflab.umassmed.edu/ctffind4>) this is now automated.
61. D4 symmetry was used here because of the D4 symmetry in Rubisco.

Figure Legends

Figure 1. Hexadecameric Rubisco, CcmM and Rubisco-M35 complex.

- (a) Structure of hexadecameric Rubisco (PDB: 6HBC). Left and right panel, side view and end view, respectively. RbcL subunits (gold and khaki) of the four anti-parallel RbcL dimers are shown in surface representation. RbcS subunits (green) bound between the anti-parallel RbcL dimers are shown in ribbon representation.
- (b) Schematics of full-length CcmM protein (M58) and the shorter isoform (M35) of the cyanobacterium *Synechococcus elongatus* PCC7942S. γ CA, γ -carbonic anhydrase; SSUL, Rubisco small subunit like domain.
- (c) Crystal structure of the oxidized SSUL1 domain of *S. elongatus* (PDB: 6HBA).
- (d) Model of the *Se*Rubisco-*Se*M35 complex showing a hypothetical alternating arrangement of SSUL domains. *Se*Rubisco is shown as in (a) and SSUL domains (purple) are shown in ribbon representation.

Figure 2. Condensate formation of *Se*Rubisco and *Se*M35 monitored by turbidity assay.

- (a) Kinetics of condensate formation by *Se*Rubisco (0.25 μ M) and M35_{red} or M35_{ox} (2 μ M) monitored by absorbance wavelength at 340 nm at 25 °C. *Se*Rubisco and M35 alone serve as controls.
- (b) Apparent binding affinity (K_D^{app}) of M35_{red} and M35_{ox} to *Se*Rubisco. Condensate formation was measured as in (a) at increasing concentrations of M35. Turbidity values reached after 10 min are plotted as mean \pm s.d.

Figure 3. Condensate formation of *Se*Rubisco and *Se*M35 monitored by optical microscopy.

- (a) *Se*Rubisco-*Se*M35 condensates generated by mixing unlabeled *Se*Rubisco (0.25 μ M) and unlabeled M35 (1 μ M) and imaged by bright-field microscopy at 25 °C.
- (b) – (c) Unlabeled *Se*Rubisco and *Se*M35 were mixed at a molar ratio of 10:1 with the respective labeled proteins. *Se*Rubisco was labeled with Alexa 532 (green) and *Se*M35 labeled with NanoTemper 650 (red). The proteins were imaged by fluorescence microscopy at 25 °C either in combination (b) or alone (c). Droplets containing both *Se*Rubisco and *Se*M35 are indicated by the merged fluorescence image.

Figure 4. *Se*Rubisco-M35 condensate screened by transmission electron microscopy.

- (a) – (b) Representative micrographs of *Se*Rubisco-*Se*M35 condensates obtained by negative stain EM (a) and cryo-EM (b).

Figure 5. Single particle analysis of *Se*Rubisco-*Se*M35 complex by cryo-EM.

- (a) 2D class averages of complexes showing extra density of SSUL bound to Rubisco in some classes. The bound SSUL is indicated by white arrow heads.
- (b) 3D classes of complex shown in (a). Particle percentage in each class is indicated.
- (c) Density map of the 3D reconstruction from cryo-EM and single particle analysis of final 2RbcL-2RbcS-SSUL units at \sim 2.8 Å resolution (left). Colors as in Figure 1. A zoom-in view of the structural interface which is formed by two salt bridges, residues Arg251 and Arg252 of the SSUL domain of M35 with Asp93 of RbcS and Asp76 of RbcL, respectively (right). Colors as in Figure 1.

References

1. Kerfeld CA, Aussignargues C, Zarzycki J, Cai F, Sutter M (2018) Bacterial microcompartments. *Nat Rev Microbiol* 16(5):277-290. doi:10.1038/nrmicro.2018.10
2. Greening C, Lithgow T (2020) Formation and function of bacterial organelles. *Nat. Rev. Microbiol.* doi:10.1038/s41579-020-0413-0
3. Yeates TO, Thompson MC, Bobik TA (2011) The protein shells of bacterial microcompartment organelles. *Curr Opin Struct Biol* 21(2):223-231. doi:10.1016/j.sbi.2011.01.006
4. Espie GS, Kimber MS (2011) Carboxysomes: cyanobacterial Rubisco comes in small packages. *Photosynth. Res.* 109(1-3):7-20. doi:10.1007/s11120-011-9656-y
5. Rae BD, Long BM, Badger MR, Price GD (2013) Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO₂ fixation in cyanobacteria and some proteobacteria. *Microbiol Mol Biol Rev* 77(3):357-379. doi:10.1128/mnbr.00061-12
6. Kerfeld CA, Melnicki MR (2016) Assembly, function and evolution of cyanobacterial carboxysomes. *Curr Opin Plant Biol* 31:66-75. doi:10.1016/j.pbi.2016.03.009
7. Badger MR, Price GD (2003) CO₂ concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J Exp Bot* 54(383):609-622. doi:10.1093/jxb/erg076
8. Cai F, Menon BB, Cannon GC, Curry KJ, Shively JM, Heinhorst S (2009) The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO₂ leakage barrier. *PLoS One* 4(10):e7521. doi:10.1371/journal.pone.0007521
9. Kaplan A (2017) On the cradle of CCM research: discovery, development, and challenges ahead. *J Exp Bot* 68(14):3785-3796. doi:10.1093/jxb/erx122
10. Maurino VG, Peterhansel C (2010) Photorespiration: current status and approaches for metabolic engineering. *Curr Opin Plant Biol* 13(3):249-256. doi:10.1016/j.pbi.2010.01.006
11. Andersson I, Backlund A (2008) Structure and function of Rubisco. *Plant Physiol Biochem* 46(3):275-291. doi:10.1016/j.plaphy.2008.01.001
12. Tabita FR, Satagopan S, Hanson TE, Kreel NE, Scott SS (2008) Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J Exp Bot* 59(7):1515-1524. doi:10.1093/jxb/erm361
13. Oltrogge LM, Chaijarasphong T, Chen AW, Bolin ER, Marqusee S, Savage DF (2020) Multivalent interactions between CsoS2 and Rubisco mediate α -carboxysome formation. *Nat Struct Mol Biol* 27(3):281-287. doi:10.1038/s41594-020-0387-7

14. Wang H, Yan X, Aigner H, Bracher A, Nguyen ND, Hee WY, Long BM, Price GD, Hartl FU, Hayer-Hartl M (2019) Rubisco condensate formation by CcmM in β -carboxysome biogenesis. *Nature* 566(7742):131-135. doi:10.1038/s41586-019-0880-5
15. Ryan P, Forrester TJB, Wroblewski C, Kenney TMG, Kitova EN, Klassen JS, Kimber MS (2019) The small RbcS-like domains of the beta-carboxysome structural protein CcmM bind RubisCO at a site distinct from that binding the RbcS subunit. *J Biol Chem* 294(8):2593-2603. doi:10.1074/jbc.RA118.006330
16. Flecken M, Wang H, Popilka L, Hartl FU, Bracher A, Hayer-Hartl M (2020) Dual functions of a Rubisco activase in metabolic repair and recruitment to carboxysomes. *Cell* 183(2):457-473.e420. doi:10.1016/j.cell.2020.09.010
17. Banani SF, Lee HO, Hyman AA, Rosen MK (2017) Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* 18(5):285-298. doi:10.1038/nrm.2017.7
18. Alberti S, Saha S, Woodruff JB, Franzmann TM, Wang J, Hyman AA (2018) A user's guide for phase separation assays with purified proteins. *J Mol Biol* 430(23):4806-4820. doi:10.1016/j.jmb.2018.06.038
19. Cuevas-Velazquez CL, Dinneny JR (2018) Organization out of disorder: liquid-liquid phase separation in plants. *Curr Opin Plant Biol* 45(Pt A):68-74. doi:10.1016/j.pbi.2018.05.005
20. Babinchak WM, Surewicz WK (2020) Liquid-liquid phase separation and its mechanistic role in pathological protein aggregation. *J Mol Biol* 432(7):1910-1925. doi:10.1016/j.jmb.2020.03.004
21. Peña KL, Castel SE, de Araujo C, Espie GS, Kimber MS (2010) Structural basis of the oxidative activation of the carboxysomal gamma-carbonic anhydrase, CcmM. *Proc Natl Acad Sci USA* 107(6):2455-2460. doi:10.1073/pnas.0910866107
22. Chen AH, Robinson-Mosher A, Savage DF, Silver PA, Polka JK (2013) The bacterial carbon-fixing organelle is formed by shell envelopment of preassembled cargo. *PLoS One* 8(9):e76127. doi:10.1371/journal.pone.0076127
23. Price GD, Coleman JR, Badger MR (1992) Association of carbonic anhydrase activity with carboxysomes isolated from the cyanobacterium *Synechococcus* PCC7942. *Plant Physiol* 100(2):784-793. doi:10.1104/pp.100.2.784
24. Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, Agard DA (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* 14(4):331-332. doi:10.1038/nmeth.4193
25. Rohou A, Grigorieff N (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* 192(2):216-221. doi:10.1016/j.jsb.2015.08.008
26. Scheres SH (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* 180(3):519-530. doi:10.1016/j.jsb.2012.09.006

27. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605-1612. doi:10.1002/jcc.20084
28. Saschenbrecker S, Bracher A, Rao KV, Rao BV, Hartl FU, Hayer-Hartl M (2007) Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell* 129(6):1189-1200. doi:10.1016/j.cell.2007.04.025
29. Liu C, Young AL, Starling-Windhof A, Bracher A, Saschenbrecker S, Rao BV, Rao KV, Berninghausen O, Mielke T, Hartl FU (2010) Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* 463(7278):197-202. doi:10.1038/nature08651
30. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9(7):676-682. doi:10.1038/nmeth.2019
31. Carisey A, Stroud M, Tsang R, Ballestrem C (2011) Fluorescence recovery after photobleaching. In: Wells CM, Parsons M (eds) *Cell Migration: Developmental Methods and Protocols*. Humana Press, Totowa, NJ, pp 387-402. doi:10.1007/978-1-61779-207-6_26
32. Scarff CA, Fuller MJG, Thompson RF, Iadanza MG (2018) Variations on negative stain electron microscopy methods: Tools for tackling challenging systems. *J Vis Exp* (132):e57199. doi:10.3791/57199
33. Gallagher JR, Kim AJ, Gulati NM, Harris AK (2019) Negative stain transmission electron microscopy of molecular complexes for image analysis by 2D class averaging. *Curr Protoc Microbiol* 54(1):e90. doi:10.1002/cpmc.90
34. Mastrorade DN (2005) Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* 152(1):36-51. doi:10.1016/j.jsb.2005.07.007
35. Danev R, Yanagisawa H, Kikkawa M (2019) Cryo-electron microscopy methodology: Current aspects and future directions. *Trends Biochem Sci* 44(10):837-848. doi:10.1016/j.tibs.2019.04.008
36. Cheng Y, Grigorieff N, Penczek PA, Walz T (2015) A primer to single-particle cryo-electron microscopy. *Cell* 161(3):438-449. doi:10.1016/j.cell.2015.03.050
37. Passmore LA, Russo CJ (2016) Specimen preparation for high-resolution cryo-EM. *Methods Enzymol* 579:51-86. doi:10.1016/bs.mie.2016.04.011
38. Serna M (2019) Hands on methods for high resolution cryo-electron microscopy structures of heterogeneous macromolecular complexes. *Front Mol Biosci* 6(33). doi:10.3389/fmolb.2019.00033

39. Stark H, Chari A (2016) Sample preparation of biological macromolecular assemblies for the determination of high-resolution structures by cryo-electron microscopy. *Microscopy (Oxf)* 65(1):23-34. doi:10.1093/jmicro/dfv367

40. Cabra V, Samsó M (2015) Do's and don'ts of cryo-electron microscopy: A primer on sample preparation and high quality data collection for macromolecular 3D reconstruction. *J Vis Exp* (95):e52311. doi:10.3791/52311

41. Drulyte I, Johnson RM, Hesketh EL, Hurdiss DL, Scarff CA, Porav SA, Ranson NA, Muench SP, Thompson RF (2018) Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Crystallogr D Struct Biol* 74(Pt 6):560-571. doi:10.1107/s2059798318006496

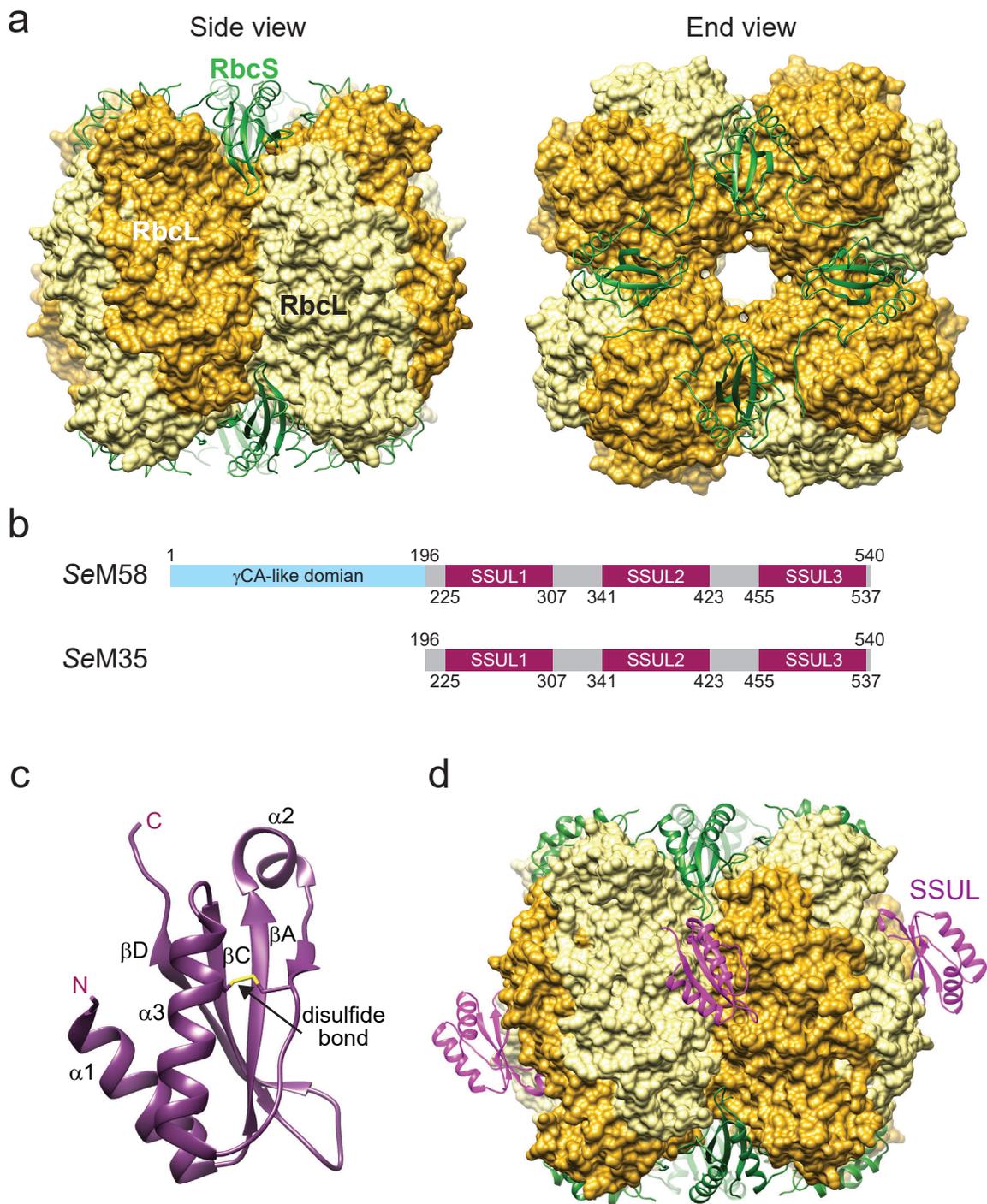


Fig. 2

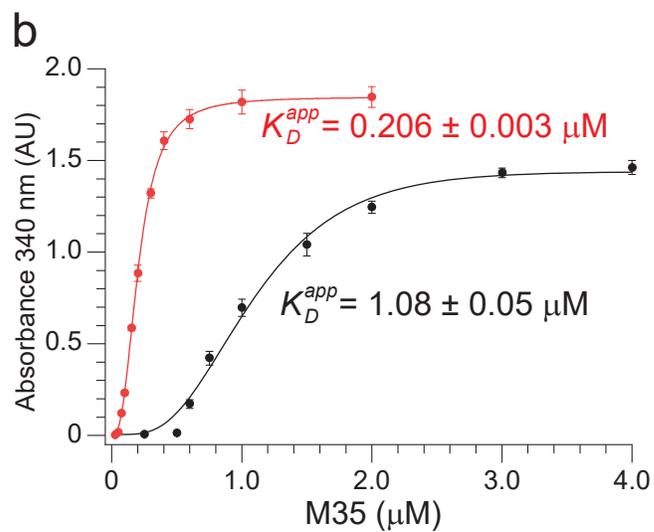
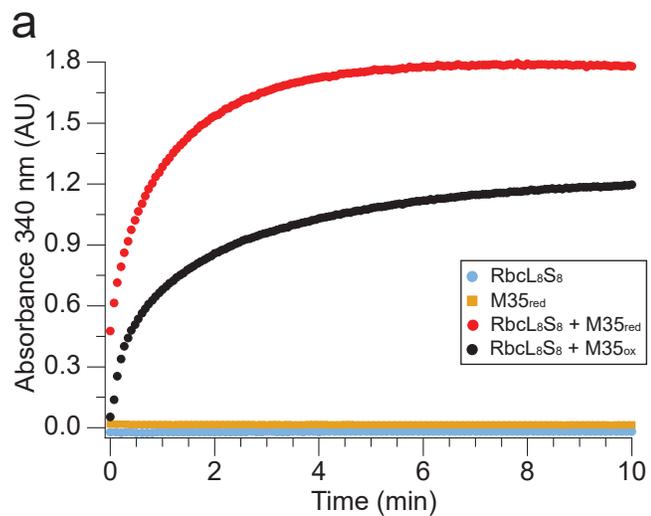


Fig. 3

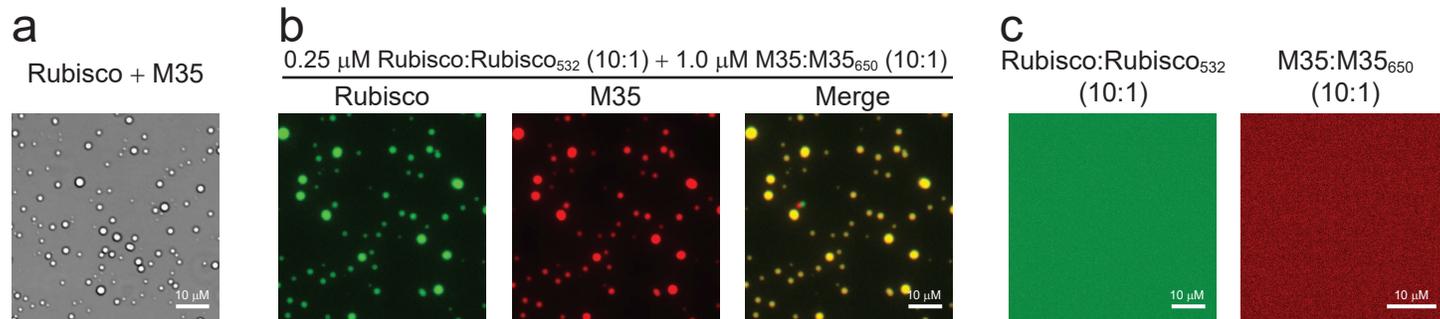


Fig. 4

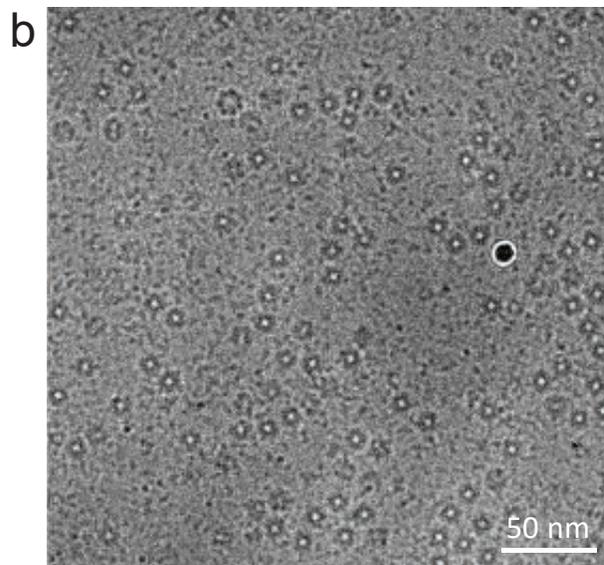
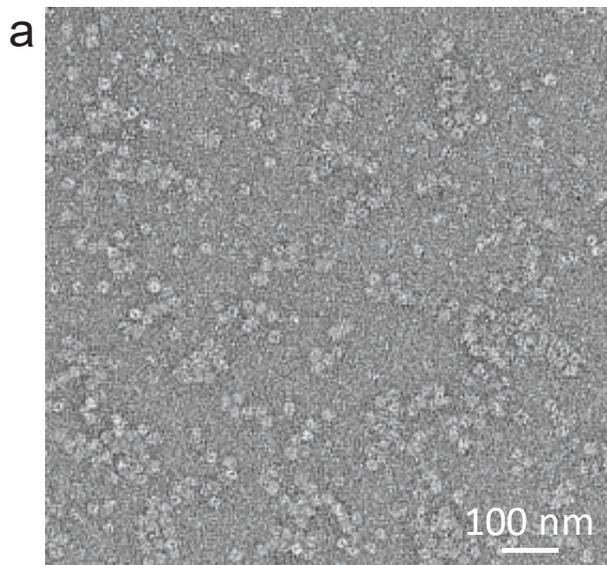
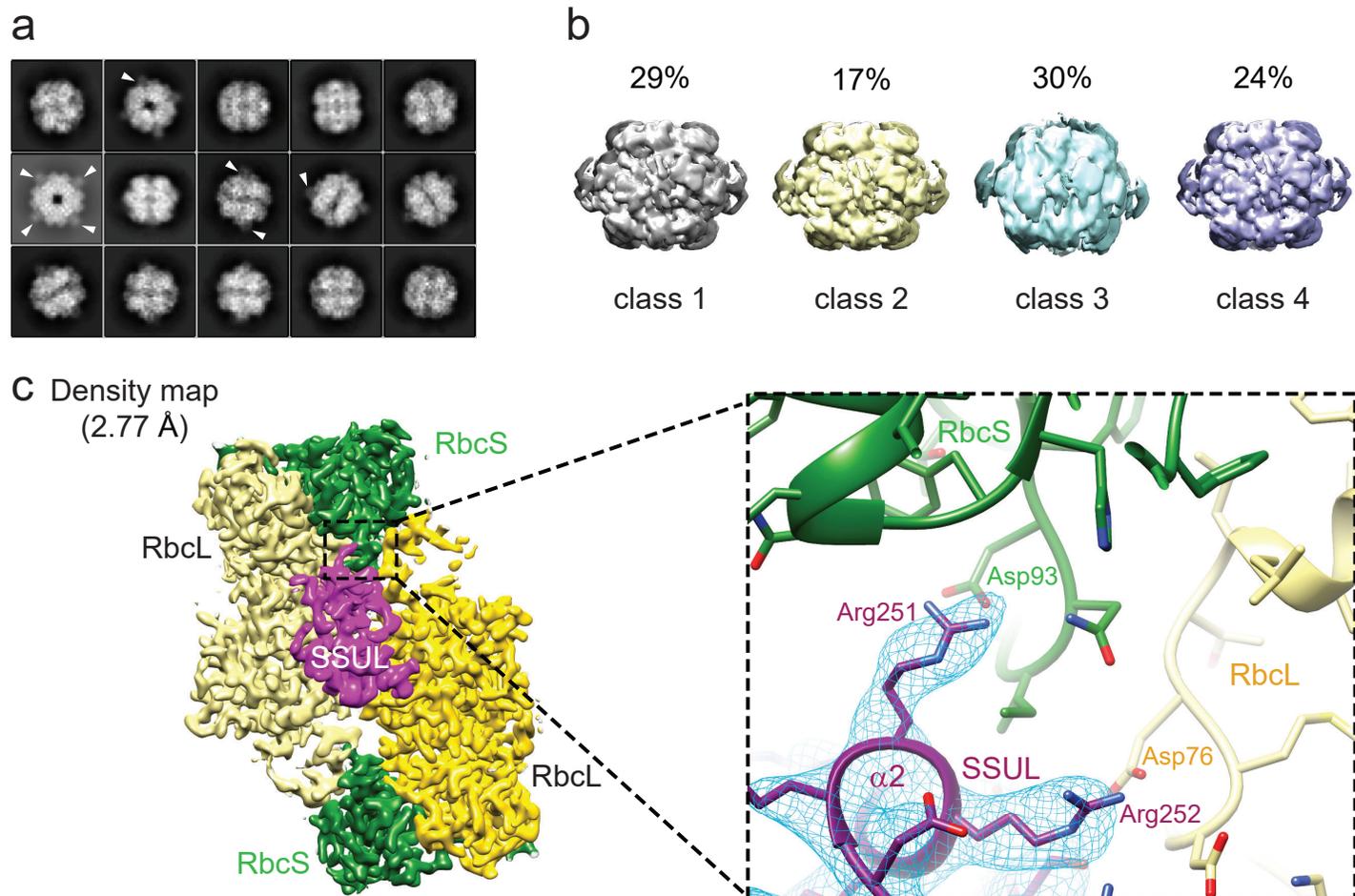


Fig. 5



4.5 Structure and conformational cycle of a bacteriophage-encoded chaperonin

RESEARCH ARTICLE

Structure and conformational cycle of a bacteriophage-encoded chaperonin

Andreas Bracher*, Simanta S. Paul[‡], Huping Wang, Nadine Wischnewski, F. Ulrich Hartl, Manajit Hayer-Hartl[‡]*

Department of Cellular Biochemistry, Max-Planck-Institute of Biochemistry, Martinsried, Germany

[‡] Current address: Department of Physics, University of Alberta, Edmonton, Canada* bracher@biochem.mpg.de (AB); mhartl@biochem.mpg.de (MH-H)

OPEN ACCESS

Citation: Bracher A, Paul SS, Wang H, Wischnewski N, Hartl FU, Hayer-Hartl M (2020) Structure and conformational cycle of a bacteriophage-encoded chaperonin. PLoS ONE 15(4): e0230090. <https://doi.org/10.1371/journal.pone.0230090>

Editor: Kornelius Zeth, Roskilde Universitet, DENMARK

Received: February 24, 2020

Accepted: April 7, 2020

Published: April 27, 2020

Copyright: © 2020 Bracher et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Crystallographic structure factors and model coordinates for crystal forms I and II of ϕ EL•ATP-BeF_x have been deposited to wwPDB under accession numbers 6TMT and 6TMU, respectively. Electron density maps for apo- ϕ EL, ϕ EL•ADP and ϕ EL•ATP γ S have been deposited to EMDB under accession numbers 10528, 10529 and 10530, respectively. The corresponding model coordinates have been deposited to wwPDB under accession codes 6TMV, 6TMW and 6TMZ.

Abstract

Chaperonins are ubiquitous molecular chaperones found in all domains of life. They form ring-shaped complexes that assist in the folding of substrate proteins in an ATP-dependent reaction cycle. Key to the folding cycle is the transient encapsulation of substrate proteins by the chaperonin. Here we present a structural and functional characterization of the chaperonin gp146 (ϕ EL) from the phage EL of *Pseudomonas aeruginosa*. ϕ EL, an evolutionarily distant homolog of bacterial GroEL, is active in ATP hydrolysis and prevents the aggregation of denatured protein in a nucleotide-dependent manner. However, ϕ EL failed to refold the encapsulation-dependent model substrate rhodanese and did not interact with *E. coli* GroES, the lid-shaped co-chaperone of GroEL. ϕ EL forms tetradecameric double-ring complexes, which dissociate into single rings in the presence of ATP. Crystal structures of ϕ EL (at 3.54 and 4.03 Å) in presence of ATP•BeF_x revealed two distinct single-ring conformational states, both with open access to the ring cavity. One state showed uniform ATP-bound subunit conformations (symmetric state), whereas the second combined distinct ATP- and ADP-bound subunit conformations (asymmetric state). Cryo-electron microscopy of apo- ϕ EL revealed a double-ring structure composed of rings in the asymmetric state (3.45 Å resolution). We propose that the phage chaperonin undergoes nucleotide-dependent conformational switching between double- and single rings and functions in aggregation prevention without substrate protein encapsulation. Thus, ϕ EL may represent an evolutionarily more ancient chaperonin prior to acquisition of the encapsulation mechanism.

Introduction

Chaperonins are large double-ring complexes that mediate protein folding in an ATP-dependent mechanism in all domains of life [1]. Two major groups of chaperonins exist, which transiently encapsulate non-native substrate protein (SP) for folding to proceed in an aggregation-free environment: Group I chaperonins occur in eubacteria and organelles of prokaryotic origin, mitochondria and chloroplasts (GroEL, Hsp60 and Cpn60, respectively). These chaperonins have 7-membered rings. The group II chaperonins in the cytosol of archaea and eukaryotes (thermosome and TRiC/CCT, respectively) have 8-mer rings. The subunits of both

Funding: This work was supported by the Deutsche Forschungsgemeinschaft (DFG) grant SFB1035 to MH-H and FUH. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

groups share a conserved three-domain architecture composed of an equatorial domain harboring the ATP binding site, an intermediate domain, which communicates nucleotide-dependent conformational changes, and a flexible apical domain for SP interaction. The equatorial domains mediate most of the inter-subunit contacts within and between the rings, whereas the apical domains form the ring opening. The group I chaperonins function together with a heptameric lid-shaped co-chaperone (GroES in bacteria, Hsp10 in mitochondria and Cpn10/20 in chloroplasts), which binds to the ends of the chaperonin cylinder and closes the folding chamber. The paradigm for the group I chaperonin mechanism is the GroEL/ES system of *Escherichia coli* [2]. GroEL and Hsp60 may transiently dissociate into single rings during their functional cycle [3, 4]. The group II chaperonins function independently of GroES-like co-chaperones and instead use helical protrusions of their apical domains as a built-in, iris-like closing mechanism [5]. Group I and II chaperonins also differ in their ring-ring contacts, with group I exhibiting staggered interactions, i.e. each subunit in one ring interacts with two subunits in the opposing ring, whereas the subunits in group II chaperonins interact in a one to one fashion.

Evolutionarily more distant chaperonin homologs were discovered in the genomes of bacteriophages. While many phages encode only a GroES homolog that cooperates with the host GroEL [6], some have both GroEL and GroES homologs and few encode only a GroEL-like chaperonin [7]. The latter group includes the protein gp146 from the bacteriophage EL of *Pseudomonas aeruginosa* [8, 9], henceforth referred to as ϕ EL. ϕ EL shares only 22% sequence identity with GroEL, equivalent to the evolutionary distance between group I and II chaperonins. Similar to group I chaperonins, ϕ EL was reported to form a double-ring complex of 7-mer rings [9]. ϕ EL was found to be expressed as early as 15 min after bacterial infection by phage, and was found to be associated with another phage protein of 32 kDa, gp188. This cell wall endolysin is unstable at 37°C [9]. ϕ EL prevented gp188 aggregation and maintained the protein in a functional state independent of host GroES [9–11]. A low resolution cryo-electron microscopy (cryo-EM) structure reported a spherical heptamer for ADP-bound ϕ EL, which was suggested to represent the SP-encapsulating state [10].

Here, we performed a functional and structural characterization of the chaperonin from the phage EL of *P. aeruginosa* to better understand the ϕ EL mechanism. Our results show that ϕ EL functions in preventing protein aggregation but is unable to mediate the folding of rhodanese (Rho), an encapsulation-dependent model SP. In the absence of nucleotide, ϕ EL forms mainly double-ring complexes, whereas single-rings are populated in the presence of ATP and physiological salt concentration. Our structural analysis by crystallography and cryo-EM failed to provide evidence of the previously reported spherical heptamer structure. The domain structures of ϕ EL differ from those of GroEL by numerous insertions and deletions, which alter the putative substrate-binding cleft and the contacts at the ring-ring interface. Structural elements that might act as a built-in lid were not present in ϕ EL, consistent with an evolutionarily more ancient chaperone mechanism independent of substrate encapsulation.

Materials and methods

Materials

Chemicals, enzymes and reagents were purchased from Merck unless otherwise noted.

Protein production

All protein purification steps were performed at 4°C. Protein concentrations in the final preparations were determined by measurement of absorbance at 280 nm. Purified protein samples were concentrated by ultrafiltration and snap-frozen in liquid nitrogen for storage at -80°C.

Gene product 146 (gp146 also known as ϕ EL) from *Pseudomonas* bacteriophage EL encoded by the plasmid pET22b-phi-GroEL was expressed in *E. coli* BL21(DE3) cells. Cells were grown in LB medium containing 100 mg L⁻¹ ampicillin at 37°C to an optical density of 0.5 at 595 nm. Expression of ϕ EL was induced with 1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) for 3 h at 30°C. Cells were harvested and re-suspended in ice-cold buffer PA (50 mM Tris-HCl pH 7.5). The cells were lysed by the French press method on ice using an Emulsi-flex C5 apparatus (Avestin, Ottawa, Canada). After removal of cell debris by centrifugation at 120,000 g for 45 min, the supernatant was applied to a 50 mL Source 30Q column (GE Healthcare) equilibrated in buffer PA. The column was washed with 3 column volumes (CV) buffer PA containing 50 mM NaCl, followed by a linear salt gradient in buffer PA (50–500 mM NaCl, 10 CV). The eluate fractions were analyzed by SDS-PAGE. Fractions containing ϕ EL (eluting at 17–21.5 mS cm⁻¹) were merged and transferred into buffer PB (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 mM DTT). Next, the material was passed over a 20 mL Heparin Sepharose Fast-Flow column (GE Healthcare), which was eluted with a gradient from 0 to 500 mM NaCl in buffer PB (10 CV). Finally, the concentrated fractions containing ϕ EL (eluting at 8–15 mS cm⁻¹) were subjected to size exclusion chromatography (SEC) on Sephacryl S-400 (GE Healthcare) in buffer, 20 mM MOPS-NaOH pH 7.2, 100 mM NaCl and 10% (v/v) glycerol.

Size-exclusion chromatography coupled to multi-angle static light scattering (SEC-MALS)

Purified ϕ EL at 2 g L⁻¹ was analyzed using static and dynamic light scattering by auto-injection of the sample onto a SEC column (5 μ m, 4.6x300 mm column, Wyatt Technology, product # WTC-030N5) at a flow rate of 0.35 mL min⁻¹ in buffers EM50 (20 mM HEPES-NaOH pH 7.5, 50 mM KCl, 4 mM Mg acetate) or EM100 (20 mM HEPES-NaOH pH 7.5, 100 mM KCl, 2 mM Mg acetate) at 25°C in the presence or absence of nucleotide (2 mM). The column was in line with the following detectors: a variable UV absorbance detector set at 280 nm (Agilent 1100 series), the DAWN EOS MALS detector (Wyatt Technology, 690 nm laser) and the Opti-lab rEXTM refractive index detector (Wyatt Technology, 690 nm laser) [12]. Molecular masses were calculated using the ASTRA software (Wyatt Technology) with the dn/dc value set to 0.185 mL g⁻¹. Bovine serum albumin (Thermo) was used as the calibration standard.

ATP hydrolysis

The ATPase activity of ϕ EL (200 nM tetradecamer) at different ATP concentrations was detected by absorbance at 340 nm wavelength in low salt (LS) buffer, 50 mM Tris-HCl pH 7.5, 10 mM KCl and 10 mM MgCl₂, or high salt (HS) buffer, 50 mM Tris-HCl pH 7.5, 100 mM KCl and 10 mM MgCl₂, using a NADH-coupled enzymatic assay (1 mM phosphoenolpyruvate, 20 U mL⁻¹ pyruvate kinase, 30 U mL⁻¹ lactate dehydrogenase and 0.25 mM NADH) [13]. ATPase activity assays of ϕ EL and GroEL (0.2 μ M tetradecamer) in the absence or presence of GroES (0.4 μ M heptamer) or 0.8 μ M denatured DM-MBP (diluted from 6M GuHCl, final ~20 mM) were performed in LS or HS buffer and in presence of 1 mM ATP. The ATPase activity of ϕ EL (0.2 μ M tetradecamer) was also measured with increasing concentrations of GroES. The assay temperature was 25°C.

Rhodanese prevention of aggregation and refolding

Rhodanese (Rho; 150 μ M) was denatured in 6M guanidinium-HCl (GuHCl)/10 mM DTT for 60 min at 25°C and diluted 300-fold into buffer AP (20 mM MOPS-KOH pH 7.4, 20 mM KCl, and 5 mM MgCl₂) containing 1 mM nucleotide in the absence or presence of ϕ EL or GroEL (0.5 μ M tetradecamer). Aggregation was monitored by measuring turbidity at 320 nm.

Rho refolding assays were performed as described previously with minor modifications [14]. GuHCl-denatured Rho was diluted 200-fold to a final concentration of 0.5 μM into buffer, 20 mM Tris-HCl pH 7.5, 100 mM KCl and 5 mM MgCl_2 , either lacking chaperone or containing ϕEL , $\phi\text{EL}/\text{GroES}$, GroEL or GroEL/GroES. The concentrations of GroEL and ϕEL were 1 μM (tetradecamer) and GroES 2 μM (heptamer). Refolding was initiated upon addition of ATP (5 mM). When indicated, chaperonin action was stopped by 1,2-cyclohexylenedinitri- lotetraacetic acid (CDTA) (10 mM). Enzymatic assays of Rho were performed as previously described [14]. Spontaneous refolding of Rho was inefficient (< 10% yield) due to aggregation.

Crystallization

ϕEL at 19.1 g L^{-1} in buffer, 20 mM MOPS-NaOH pH 7.3, 100 mM KCl and 2 mM Mg-acetate, was crystallized by the sitting-drop vapor diffusion method by mixing 100 nl sample with 100 nl reservoir solution using the robotics setup at the Crystallization Facility of the Max Planck Institute of Biochemistry. The drops were equilibrated against 150 μl reservoir solution at 16°C. Crystals of ϕEL were obtained with the Complex crystallization screen [15, 16].

Crystal form I was obtained in presence of 2 mM ATP, 5 mM BeSO_4 and 20 mM NaF with a reservoir solution containing 5% (w/v) PEG-4000, 0.2 M Na-acetate and 0.1 M Na_3 -citrate pH 5.5.

Crystal form II was obtained in presence of 2 mM ATP, 5 mM BeSO_4 and 20 mM NaF with a reservoir solution containing 8% (w/v) PEG-6000, 0.15 M NaCl and 0.1 M Tris-HCl pH 8.0.

For vitrification, the crystals were sequentially incubated in reservoir solution containing additionally 12.5 and 25% (v/v) glycerol for 15 min each and were then rapidly cooled in liquid nitrogen.

Diffraction data collection and processing

The X-ray diffraction data were collected by the oscillation method at beamline ID30B of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. All data were integrated and scaled with XDS [17]. Pointless [18], Aimless [19] and Ctruncate [20], as implemented in the CCP4i graphical user interface (GUI) [21], were used for data reduction.

Crystal structure solution and refinement

The space group symmetry and size of the asymmetric unit suggested that crystal forms I and II contained single-ring heptamers at 65% solvent content, which is within the expected range for protein crystals (~75–40% solvent) [22]. Analysis of the self-rotation function calculated with Molrep [23] indicated the presence of seven-fold non-crystallographic symmetry (NCS) consistent with the presence of single-ring heptamers in the crystal lattice. For solving the structure of crystal form II by molecular replacement, the cryo-EM density for the heptadecameric $\phi\text{EL}\cdot\text{ATP}$ at 6.8 Å resolution (EMDB entry EMD-6492, [10]) was segmented with Chimera [24] and the density for a single-ring heptamer extracted. With this density as a search model, a plausible molecular replacement solution was obtained. The density modification program Resolve [25] was used to extend the phases beyond 6.8 Å employing seven-fold NCS averaging and refinement. After B-factor sharpening, the resulting experimental electron density revealed features of secondary structure sufficient for manual model building with Coot [26]. ϕEL crystal form I was solved by molecular replacement using Molrep [23] with these coordinates as a search model. To identify the nucleotides bound to subunits of ϕEL , omit maps were calculated after applying random coordinate shifts to the preliminary model with Pdbset (to suppress model bias) and refinement with Refmac5 [27], which revealed density consistent with ADP-bound to chains B, D and F in crystal form II, and with ATP/ADP $\cdot\text{BeF}_x$

in all other chains. Since ATP and ADP•BeF_x cannot be distinguished at the resolution, and since ATP was added to the crystallization mix, ATP was included into the model. The models were refined with Refmac5 using local NCS restraints and translation-libration-screw (TLS) parametrization of B-factors [27]. Residues 1 and 553–558 were disordered in all chains of the final models. Furthermore, no interpretable density was observed for loop residues 290–294 in crystal form I and in chains A, C, E and G of crystal form II. Residues facing solvent channels with disordered sidechains were truncated after C-β.

Crystallographic structure factors and model coordinates for crystal forms I and II of ϕEL•ATP-BeF_x have been deposited to wwPDB under accession numbers 6TMT and 6TMU, respectively.

Cryo-electron microscopy and single particle analysis

The samples were prepared by mixing equal volumes of ϕEL stock solution (2.25 g L⁻¹) in buffer, 20 mM HEPES-NaOH pH 7.5, 100 mM NaCl and 1 mM EDTA, and dilution buffer (20 mM HEPES-NaOH pH 7.5 and 8 mM MgCl₂) containing either no nucleotide or 4 mM ADP or ATPγS. The dilution buffer with ADP also contained 0.08% (w/v) n-octyl-β-D-glucopyranoside. Subsequently the mixture was incubated at room temperature for 5 min. Holey carbon supported copper-grids (Quantifoil R2/1 300 mesh) were plasma-cleaned for 30 s (Harrick Plasma) immediately before use. All cryo-grids were prepared using a Vitrobot Mark 4 (FEI) by applying sample (5 μL) to a plasma-cleaned grid at 25°C and 100% humidity, then semi-automatically blotted for 2 s and plunge-frozen in liquid ethane.

The grids with ϕEL•ADP or ϕEL•ATPγS were analyzed on a Talos Arctica (FEI) transmission electron microscope (TEM) at 200 kV. Frames were recorded with a Falcon 3EC direct detector (FEI) operated in movie mode at 0.05 s per frame, at a pixel size of 1.997 Å and 2.019 s total exposure, with an estimated cumulative dose of 42–43 e⁻Å⁻². EPU (FEI) software was used for automated data collection. MotionCor2 [28] was employed to correct the movie stacks for beam-induced motion and dose weighting. Ctffind4.1 [29] was used to estimate the defocus. For generating templates used in auto-picking, 6,668 single particles of ϕEL•ADP were picked manually and subjected to 2D classification with RELION 3.0 [30, 31]. The eight largest 2D classes were selected and used in Gautomatch (<http://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch>) as templates for automated particle picking.

The grid with apo-ϕEL was analyzed on a Titan Krios (FEI) TEM at 300 kV with a pixel size of 1.09 Å. Data were collected with a K3 direct detector (Gatan) recording 50 frames per movie during 5.992 s total exposure with an estimated cumulative dose of 77.6 e⁻Å⁻². SerialEM software was used for automated data collection [32, 33]. Motion correction, dose weighting, defocus analysis and particle picking were carried out automatically during data collection using the Focus software package [34]. Movies with large drift, exhibiting ice diffraction or poor resolution (> 5 Å) in the power spectra were immediately discarded.

In absence of nucleotide and in presence of ADP, ϕEL had a tendency to associate into large aggregates. Images with thick aggregates were removed after visual inspection. RELION 3.0 was used for further data processing [31]. The complete data sets went through two rounds of 2D classification to remove contaminations or false positive particles. Using RELION 3.0, an initial model was generated from the remaining particles and used as a reference map for symmetry-free 3D refinement. The aligned particles were subjected to 3D classification. The particles from the largest class were used for further 3D refinement and post-processing including mask application and B-factor sharpening. The resulting EM density maps were inspected with Chimera [24], and C2 (apo-ϕEL) and D7 (ϕEL•ADP and ϕEL•ATPγS) particle symmetry were detected. Re-processing of the particles using the same protocol as above, but

with application of symmetry restraints, yielded improved maps for manual model building with Coot [26]. First, the apical, intermediate and equatorial domains of ϕ EL from the form-I crystal structure were separately placed into density by rigid-body real-space fitting. After manual adjustment of the coordinates to the density, the models were refined with Refmac5 in reciprocal space, using jelly-body and NCS restraints [27].

Cryo-EM density maps for apo- ϕ EL, ϕ EL•ADP and ϕ EL•ATP γ S have been deposited to EMDB under accession numbers 10528, 10529 and 10530, respectively. The corresponding model coordinates have been deposited to wwPDB under accession codes 6TMV, 6TMW and 6TMX.

Structure analysis

The quality of the structural models was analyzed with the program Molprobrity [35]. Coordinates were aligned with Lsqkab and Lsqman [36]. Figures were generated with the programs Pymol (<http://www.pymol.org>), Chimera [24] and ESPript [37].

Results

Oligomeric state of ϕ EL

ϕ EL was expressed as a soluble protein in *E. coli* and purified (S1A Fig). To determine the oligomeric state of ϕ EL at different ionic strength and in the absence or presence of nucleotide (at 25°C), we subjected ϕ EL to size exclusion chromatography combined with multi-angle light scattering (SEC-MALS). At 100 mM KCl, ϕ EL fractionated at a molecular weight (MW) of ~770 kDa (Fig 1A), indicating a high population of double-ring complexes (theoretical MW ~863 kDa). A similar MW (~725 kDa) was observed in the presence of ADP (Fig 1B). In contrast, ϕ EL in the presence of ATP shifted to a homogeneous population of complexes with a MW of ~409 kDa, close to the MW of the single-ring heptamer (theoretical MW ~431 kDa) (Fig 1C). This suggests that binding of ATP to ϕ EL in the presence of close to physiological salt concentration destabilizes the inter-ring contacts in the double ring structure. However, at a lower salt concentration of 50 mM KCl, this destabilization was not observed and mainly double-ring complexes were populated, as for apo and ADP-bound ϕ EL (Fig 1A–1C). These results suggests that under conditions of ongoing ATP-binding and hydrolysis (at 100 mM KCl), ϕ EL may cycle between double-ring and single-ring complexes.

ATPase activity of ϕ EL

ATP hydrolysis by group I and II chaperonins exhibits positive cooperativity within rings, with higher ATP occupancy triggering ATP hydrolysis (Hill coefficient for GroEL ~2.8) [38]. In addition, negative cooperativity between the rings results in a reduced hydrolysis activity at still higher ATP concentration [38]. At 25°C and 10 mM KCl, ϕ EL hydrolyzed ATP with near Michaelis-Menten kinetics, reaching a maximal turnover number of 558 ± 28 ATP min⁻¹ per tetradecamer, which is ~8-fold higher than that of *E. coli* GroEL [4, 38]. A Hill coefficient of 1.21 ± 0.05 and an apparent K_M of 0.68 ± 0.07 mM ATP were determined (Fig 2A). At 100 mM KCl, we measured a similar maximal turnover number (576 ± 7 ATP min⁻¹ per tetradecamer), a Hill coefficient of 1.61 ± 0.18 and a ~10-fold lower K_M of 0.072 ± 0.008 mM ATP (Fig 2B). Thus, at the near physiological salt concentration the affinity of ϕ EL for ATP is increased with weak positive cooperativity for ATP hydrolysis. No evidence for negative cooperativity in ATP hydrolysis was detected. The higher ATPase activity and lower Hill coefficient of ϕ EL compared to GroEL are consistent with a reduced level of allosteric coordination within and between rings of ϕ EL.

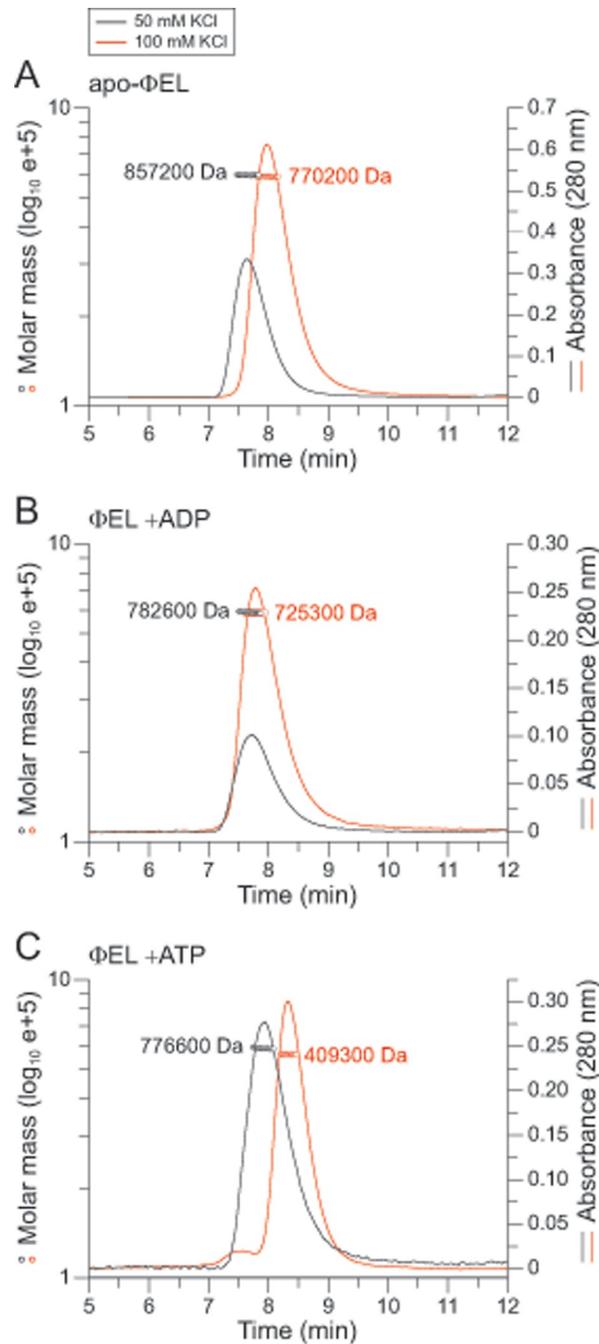


Fig 1. Oligomeric state of ϕ EL in solution. (A-C) SEC-MALS analysis of ϕ EL in absence of nucleotide (A), in presence of 2 mM ADP (B) or 2 mM ATP (C). The chromatographic absorbance traces at 280 nm wavelength are shown. The molecular mass determined for the protein peaks by static light scattering is indicated. The black and red traces were recorded in presence of 50 and 100 mM salt, respectively.

<https://doi.org/10.1371/journal.pone.0230090.g001>

No interaction of ϕ EL with *E. coli* GroES

The co-chaperone GroES (2-fold molar excess) inhibited the GroEL ATPase activity by ~50% both at 10 and 100 mM KCl (Fig 3A and 3B) [39]. In contrast, we observed no effect of *E. coli* GroES on the ATPase activity of ϕ EL (at 1 mM ATP) independent of salt concentration (Fig 3A

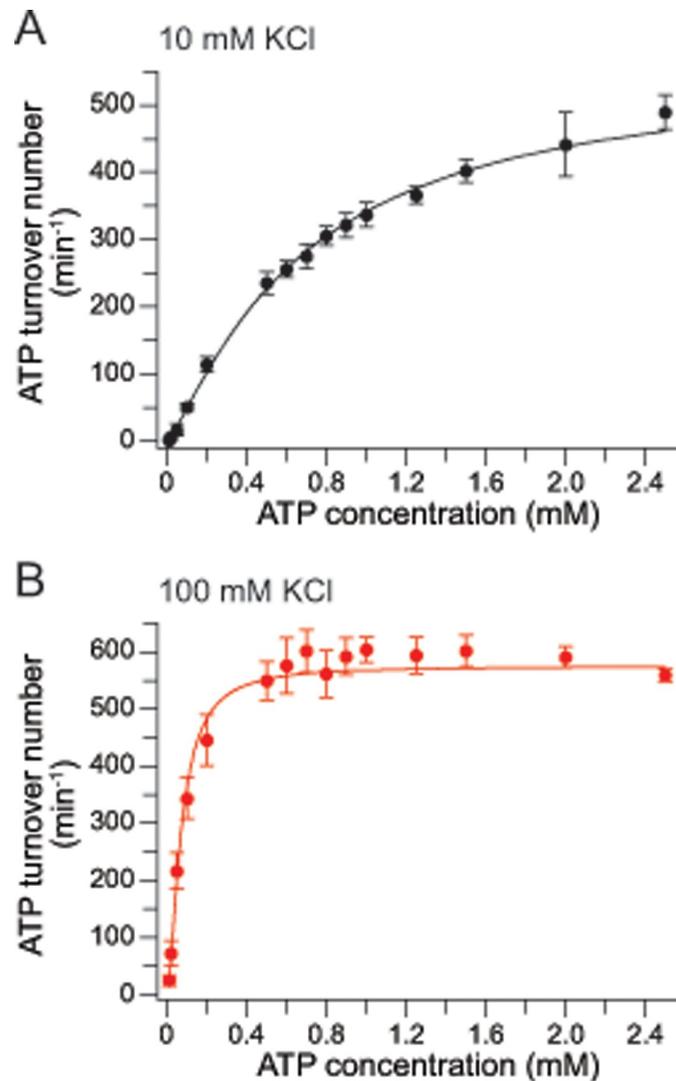


Fig 2. Concentration dependence of ATP hydrolysis by ϕ EL. (A, B) The curves show the concentration dependence of ATP hydrolysis by ϕ EL in presence of 10 (A) and 100 mM KCl (B). Shown are the averages of three experiments. Error bars indicate standard deviations. The red lines represent the Hill curve fittings of the data.

<https://doi.org/10.1371/journal.pone.0230090.g002>

and 3B). Note that at a concentration of 1 mM ATP, the ATPase rate of ϕ EL measured at 10 mM KCl is ~50% lower than in the presence of 100 mM KCl. Increasing GroES up to 8-fold excess over ϕ EL also showed no effect (S1B Fig), suggesting the absence of a functional interaction. *E. coli* GroES shares 61% amino acid identity (90% similarity) with the GroES of the host bacterium of phage EL. Specifically, the mobile loop residues of GroES that mediate binding to the apical domains of GroEL are completely conserved between *E. coli* GroES and *P. aeruginosa* GroES (S1C Fig). To further analyze a possible interaction of GroES with ϕ EL, we took advantage of the fact that the mobile loops of GroES become protected against degradation by proteinase K (PK) upon complex formation with GroEL in the presence of ADP (Fig 3C, lanes 1, 2, 9 and 10) [40]. In contrast, no protection of GroES was observed in the presence of ϕ EL (Fig 3C, lanes 5 and 6), suggesting absence of binding. Note that PK cleaved ϕ EL into fragments of ~45 and ~20 kDa (Fig 3C, lanes 3 and 4), while GroEL is largely PK resistant (Fig 3C, lanes 7 and 8), except for cleavage of the flexible 16 C-terminal residues of the GroEL subunits [41].

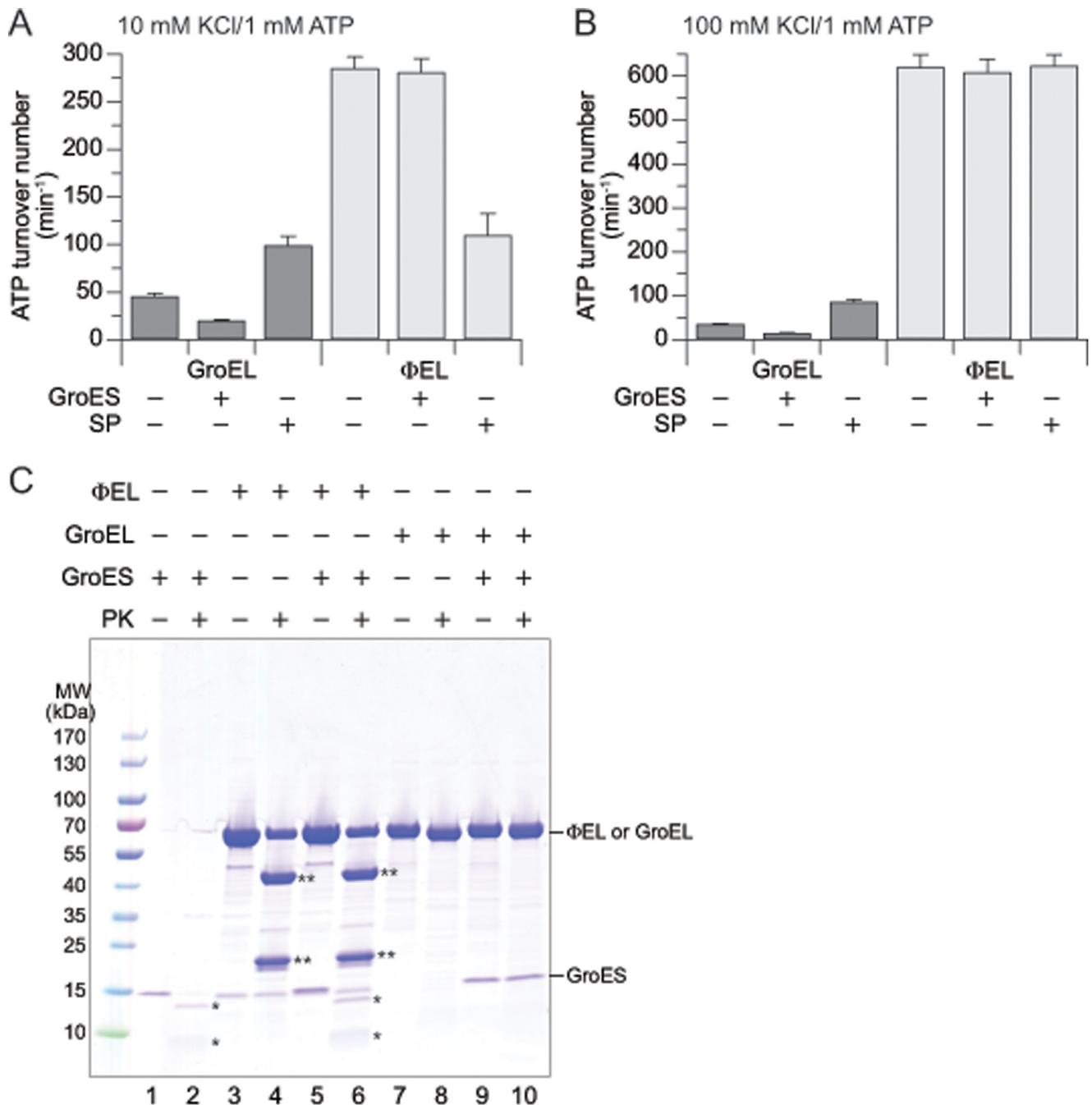


Fig 3. Effect of GroES on GroEL and ϕ EL. (A, B) ATPase activity of GroEL and ϕ EL in presence and absence of a two-fold excess of GroES or the model substrate DM-MBP. The buffer contained 10 mM (A) or 100 mM KCl (B) and the ATP concentration was 1 mM. The bar graph shows the averages from at least three experiments; error bars indicate standard deviations. (C) Proteinase K (PK) sensitivity of GroES alone, and of ϕ EL and GroEL in presence and absence of GroES. The experiment was performed in buffer 20 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM MgCl₂ and 0.2 mM ADP [42]. The concentrations of GroEL and ϕ EL were 1.5 μ M; GroES was at 1.0 μ M. After 3 min incubation with 0.2 g L⁻¹ PK at 25°C, the protease reaction was stopped by addition of phenylmethylsulfonyl fluoride (1 mM). The mixtures were analyzed by SDS-PAGE. *, proteolytic fragments of GroES; **, proteolytic fragments of ϕ EL.

<https://doi.org/10.1371/journal.pone.0230090.g003>

We next analyzed the effect of SP on the ATPase of ϕ EL. We used as model SP the double mutant of maltose binding protein (DM-MBP), which folds slowly ($t_{1/2} \sim 35$ min at 25°C) [43] in the absence of chaperonin and has a low aggregation propensity, thus allowing us to

perform ATPase measurements under SP saturation of chaperonin [44]. Binding of non-native DM-MBP stimulated the GroEL ATPase by ~2-fold independent of salt concentration (Fig 3A and 3B). Interestingly at 10 mM KCl, DM-MBP inhibited the ϕ EL ATPase by ~60% (Fig 3A). In contrast, no inhibition of ϕ EL ATPase was observed at 100 mM KCl (Fig 3B). Thus, the effect of SP on the ATPase appears to depend on whether ϕ EL populates double-ring complexes (low salt) or single-rings (high salt) (Fig 1C).

In summary, GroEL and ϕ EL differ substantially with regard to their interaction with co-chaperone (*E. coli* GroES) and SP.

Chaperone activity of ϕ EL

Next, we investigated the molecular chaperone activity of ϕ EL. We first tested the ability of ϕ EL to prevent aggregation of the model chaperonin SP rhodanese (Rho; ~30 kDa). This protein has a high propensity to aggregate upon dilution from denaturant into buffer, but folds efficiently upon transient encapsulation in the GroEL-GroES cage [42, 45–48]. Aggregation was monitored spectrophotometrically by measuring turbidity at 320 nm. A time-dependent aggregation of Rho was observed upon dilution from denaturant to a final concentration of 0.5 μ M (Fig 4A and 4B). Note that Rho aggregation was independent of the presence of nucleotide (Fig 4A and 4B). GroEL at a 1:1 molar ratio, in the absence of nucleotide, completely prevented the aggregation of Rho in the turbidity assay. In contrast, ϕ EL failed to prevent Rho aggregation (Fig 4A). However, ϕ EL inhibited Rho aggregation with high efficiency in the presence of ATP or ADP (Fig 4A and 4B). In the case of GroEL, aggregation was efficiently prevented in the presence of ADP (Fig 4B), whereas addition of ATP reduced the ability of GroEL to bind non-native Rho [45] (Fig 4A). Thus, in contrast to GroEL, SP binding to ϕ EL requires presence of nucleotide.

Rho activity assays were performed to determine whether aggregation prevention by ϕ EL was coupled to productive folding. Control reactions showed only ~5% recovery of Rho enzyme activity during spontaneous folding (Fig 4C), due to Rho aggregation. Efficient refolding was obtained by GroEL in a manner dependent on GroES and ATP [42, 45]. In contrast, only ~8% of Rho enzyme activity was recovered in the presence of ϕ EL and ATP (Fig 4C). Addition of GroES was without effect (Fig 4C), consistent with the absence of a functional interaction of ϕ EL with *E. coli* GroES (Figs 3 and S1B).

To investigate the fate of Rho during attempted refolding with ϕ EL and ATP, we analyzed the reaction after 30 min by SEC in the presence of ATP, followed by SDS-PAGE and immunoblotting with anti-Rho. ϕ EL fractionated in an asymmetric distribution (with the peak in fraction 8), consistent with cycling between double- and single-ring complexes (S2A Fig). Rho fractionated with a higher mass than ϕ EL (with the peak in fraction 7), indicating that it formed a high molecular weight aggregate as it dissociated from ϕ EL during gel filtration (S2A Fig). Rho refolding reactions with GroEL and GroES were analyzed as a control. In the absence of ATP, Rho co-fractionated with GroEL (S2B Fig), whereas in the presence of ATP most of the Rho fractionated as the monomeric native protein (S2C and S2D Fig).

In conclusion, ϕ EL can bind non-native Rho in the presence of nucleotide thereby preventing aggregation. However, ATP-dependent cycling of Rho fails to promote productive folding, independent of the presence of GroES. Apparently, during cycling Rho is transiently released in an unfolded state, explaining its aggregation during gel filtration when rebinding to ϕ EL is precluded.

Crystal structures of ϕ EL in presence of ATP•BeF_x

To obtain insight into the structural features underlying the ATP-dependent double- to single-ring transition in ϕ EL, we tried to determine the crystal structures of apo- ϕ EL and of ϕ EL in

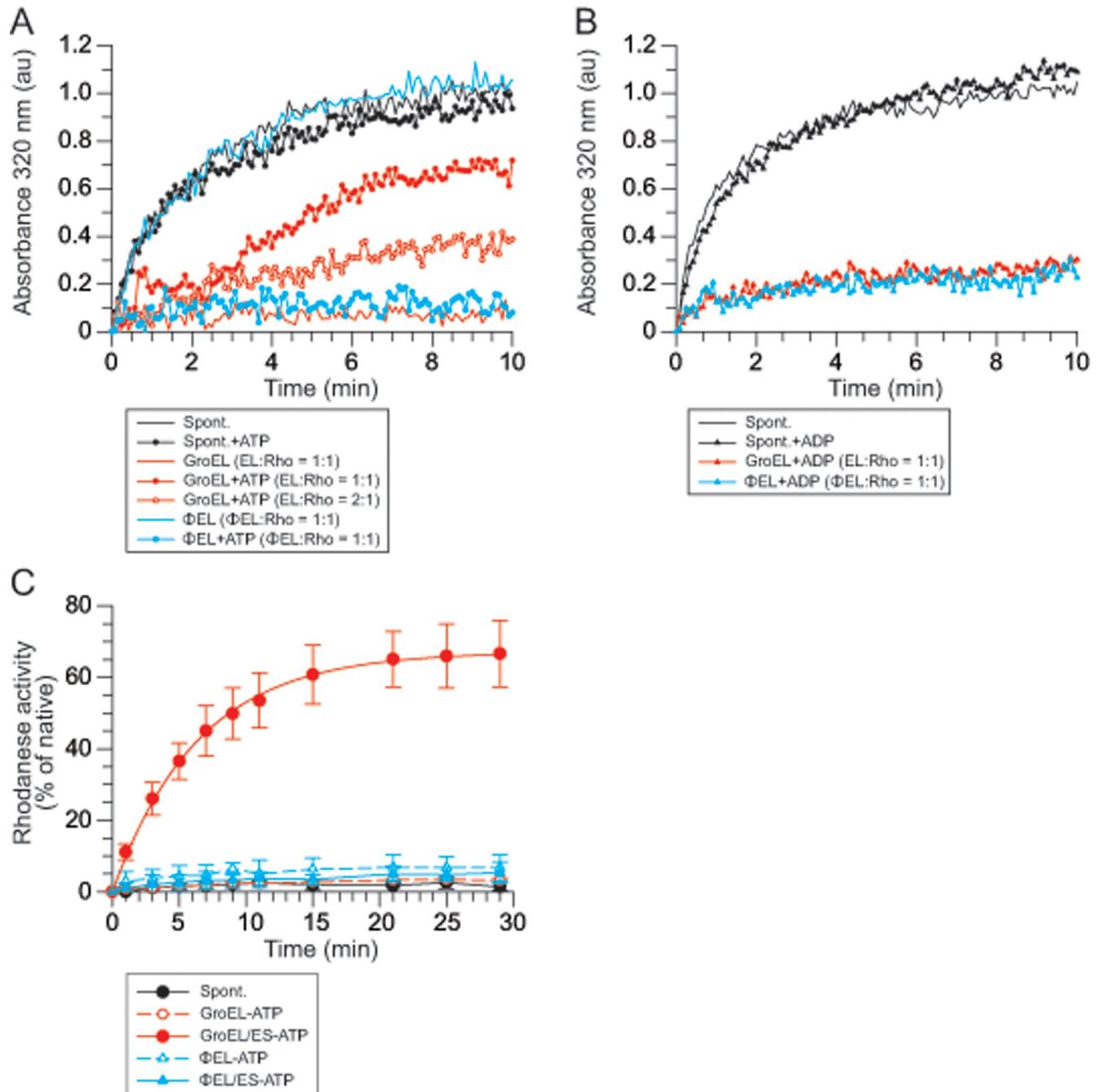


Fig 4. Molecular chaperone activity of ϕ EL. (A, B) Aggregation prevention activity of ϕ EL and GroEL in presence and absence of ATP (A) or ADP (B). Rhodanese (Rho) aggregation was monitored by turbidity assay at 320 nm wavelength. The results of representative experiments are shown. (C) Rho refolding in presence of ATP. The chaperones GroEL, GroES and ϕ EL were present when indicated. After quenching ATP hydrolysis by addition of CDTA at the indicated time points, Rho enzyme activity was determined. The averages from three experiments are shown, error bars indicate standard deviations.

<https://doi.org/10.1371/journal.pone.0230090.g004>

presence of ADP or ATP•BeF_x (ADP•BeF_x is a mimic of bound ATP prior to hydrolysis). However, only crystals of ϕ EL with ATP•BeF_x diffracted below 7 Å resolution. Two crystal forms were identified that diffracted to 4.03 and 3.54 Å resolution (Table 1). The structure was solved by molecular replacement at 6.8 Å resolution using the cryo-EM map of the ϕ EL-ATP

tetradecamer as a search model (EMD-6492) [10]. The seven-fold non-crystallographic symmetry was then employed to extend the phases to 3.54 Å, resulting in an electron density map of sufficient quality to build an atomic model. The other crystal form was solved by molecular replacement using this model. Both structures consist of heptameric single-ring complexes of ϕ EL (Fig 5A and 5B).

As anticipated, the ϕ EL subunits displayed the typical three-domain architecture of chaperonins, composed of an equatorial domain (residues 2–130 and 425–552), an intermediate domain (residues 131–188 and 388–424), and an apical domain (residues 189–387) (Fig 6A). These domains formed rigid-body units in the heptamer rings, which tend to move *en bloc* when comparing subunits (S3A–S3D Fig). The apical domains appeared to be the most mobile units, as judged from their comparatively poor electron density when not stabilized by crystal

Table 1. Crystallographic data collection and refinement statistics.

| Crystal form | I | II |
|---|--|--|
| PDB code | 6TMT | 6TMU |
| Data Collection | | |
| Nucleotide | ATP•BeF _x | ATP•BeF _x |
| Diffraction source | ESRF beamline ID30B | ESRF beamline ID30B |
| Space group | <i>P</i> ₂ ₁ ₂ ₁ | <i>P</i> ₂ ₁ ₂ ₁ |
| Cell dimensions | | |
| a, b, c (Å) | 137.9, 149.4, 268.8 | 145.5, 151.5, 261.3 |
| α, β, γ (°) | 90, 90, 90 | 90, 90, 90 |
| Wavelength (Å) | 0.85000 | 0.85000 |
| Resolution (Å) | 48.98–4.03 (4.17–4.03)* | 49.47–3.54 (3.62–3.54) |
| <i>R</i> _{merge} | 0.344 (1.891) | 0.111 (1.119) |
| <i>R</i> _{p.i.m.} | 0.099 (0.528) | 0.056 (0.557) |
| <i>I</i> / <i>σI</i> | 6.9 (1.6) | 7.9 (1.3) |
| Completeness (%) | 99.9 (100) | 99.8 (99.9) |
| Redundancy | 12.9 (13.7) | 4.9 (4.9) |
| Refinement | | |
| Resolution (Å) | 30–4.03 | 30–3.54 |
| No. reflections | 44298 | 67159 |
| <i>R</i> _{work} / <i>R</i> _{free} | 0.2562 / 0.2991 | 0.2504 / 0.2747 |
| No. atoms | | |
| Protein | 29155 | 29093 |
| Nucleotides | 224 | 213 |
| Average <i>B</i> -factors | | |
| Protein (Å ²) | 175 | 176 |
| Nucleotides (Å ²) | 112 | 151 |
| R.m.s. deviations | | |
| Bond length (Å) | 0.007 | 0.005 |
| Bond angles (°) | 1.067 | 0.967 |
| Ramachandran plot | | |
| Favored (%) | 94.5 | 95.2 |
| Allowed (%) | 5.3 | 4.7 |
| Outliers (%) | 0.2 | 0.1 |

* Values in parenthesis are for outer shell.

<https://doi.org/10.1371/journal.pone.0230090.t001>

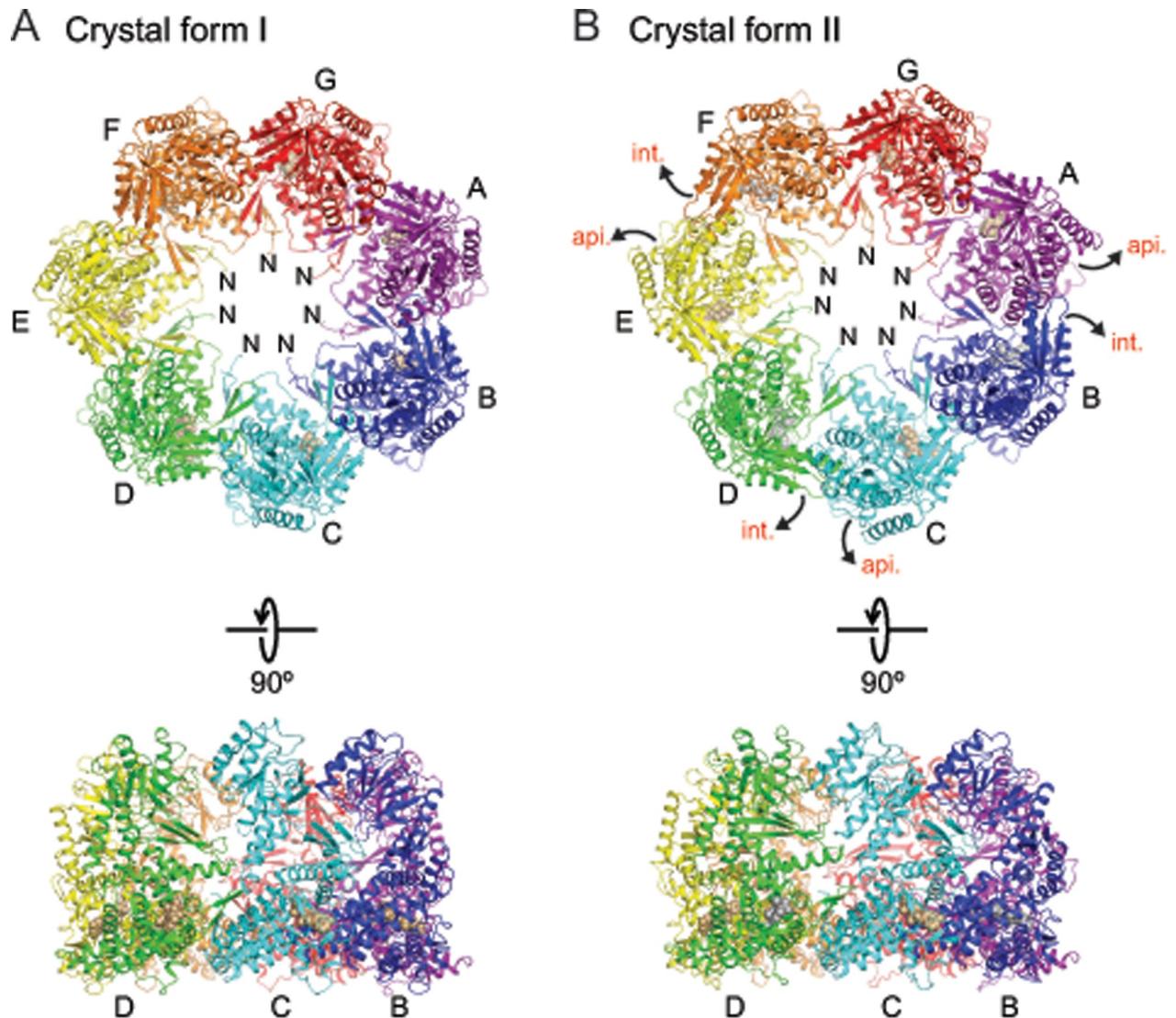


Fig 5. Crystal structures of ϕ EL. (A) Ribbon representation of the ϕ EL complex in crystal form I. Perpendicular views of the ϕ EL complex are shown. The subunits are shown in rainbow colors. Bound ATP and Mg^{2+} is colored beige and shown in space-filling mode. Subunit chain identifiers and N-termini are indicated. (B) Structure of the ϕ EL complex in crystal form II, using the same representation style. Bound ADP is colored silver and shown in space-filling mode. Domain movements in ϕ EL subunits compared to crystal form I are indicated by curved arrows.

<https://doi.org/10.1371/journal.pone.0230090.g005>

contacts. The equatorial domains were well-defined, as they form the majority of the contacts between adjacent subunits, similar to known chaperonin structures.

Structure of ϕ EL subunits in crystal form I. The ϕ EL heptamer in crystal form I (4.03 Å resolution) consisted of subunits in closely similar conformations (conformation I, r.m.s.d. range for C α positions 0.28–1.36 Å) (S3A Fig). In conformation I, the tip of the protruding α K- α L helical hairpin of the apical domain contacts the equatorial domain (Fig 6A). The α K- α L helical hairpin makes additional contacts to the intermediate domain (Fig 6A). These contacts appear to stabilize the orientation of the apical domain. No such contact is found in group I chaperonin structures (group II chaperonins do not contain this helical hairpin). There were no detectable contacts between apical domains unless forced by crystal packing.

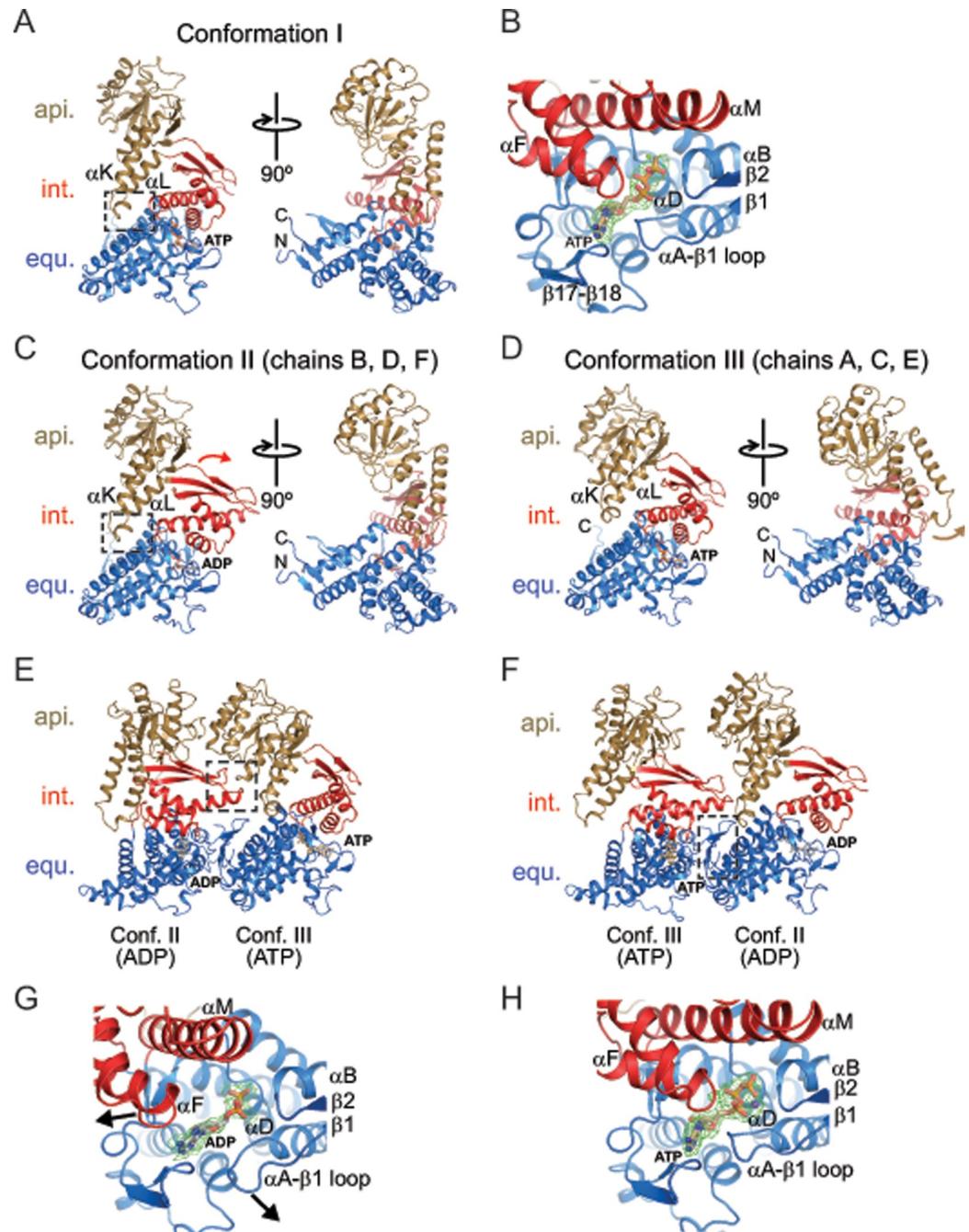


Fig 6. Conformations of ϕ EL subunits observed in the crystal structures. (A) Ribbon representation of ϕ EL subunit conformation I as observed in chain A of crystal form I. Perpendicular views of ϕ EL are shown. Apical, intermediate and equatorial domain are shown in gold, red and blue, respectively. Bound ATP is shown in stick representation. The intramolecular contact between apical and equatorial domain is highlighted by a box. The α K- α L hairpin is indicated. N- and C-termini are indicated. (B) Zoom-in on the bound ATP and Mg^{2+} in chain A of crystal form I (conformation I). The representation is equivalent to panel A. Unbiased *F_o-F_c* omit density at 2.85σ for the nucleotide is shown in green as meshwork. Secondary structure elements involved in contacts to the nucleotide are indicated. (C) Ribbon representation of ϕ EL subunit conformation II as observed in chain D of crystal form II. The rotation of the intermediate domain (red) compared to conformation I is indicated by a curved arrow. The intramolecular contact between apical and equatorial domain is highlighted by a box. Bound ADP is shown in stick representation. (D) Ribbon representation of ϕ EL subunit conformation III as observed in chain A of crystal form II. The rotation of the apical domain (gold) compared to conformations I and II is indicated by a curved arrow. Bound ATP is shown in stick representation. (E) Contact between subunits in conformations II and III, as observed between chains E and D in crystal form II. The characteristic intermediate-apical domain intermolecular

contact is highlighted by a box. (F) Contact between subunits in conformations III and II, as observed between chains B and A in crystal form II. The contacts are almost exclusively formed between the equatorial domains (boxed area). (G) Zoom-in on the bound ADP in chain D of crystal form II (conformation II). Unbiased *F_o-F_c* omit density at 1.7 σ for the nucleotide is shown in green as meshwork. Movement of the intermediate domain and remodeling of the α A- β 1 loop, respectively, are indicated by arrows. (H) Zoom-in on the bound ATP and Mg²⁺ in chain A of crystal form II (conformation III). Unbiased *F_o-F_c* omit density at 2.85 σ for the nucleotide is shown in green as meshwork.

<https://doi.org/10.1371/journal.pone.0230090.g006>

All subunits as judged from the electron density had either ATP or ADP•BeF_x bound in the equatorial domain. For simplicity, we modelled the bound nucleotide as ATP. ATP was cradled by the α A- β 2 loop (residues 30–33), the N-terminal ends of helix α D (residues 86–90) and helix α N1 (residues 428–430), helix α O (residues 474 and 478) and residues 504–506 and 519–521 (Fig 6B). Thr145 and Gln149 at the C-terminal end of α F in the intermediate domain approached the nucleobase, but did not appear to make full contact (Fig 6B). Intermediate domain helix α M with the catalytic residue Asp412 (equivalent to Asp398 in *E. coli* GroEL) was at ~6.5 Å distance to the γ -phosphate of ATP, consistent with a conformation poised for ATP hydrolysis (Fig 6B).

Structure of ϕ EL subunits in crystal form II. In contrast to crystal form I, the subunits in crystal form II exhibited considerable conformational differences (r.m.s.d. range 0.60–4.43 Å), and two new states could be assigned (conformations II and III) (Figs 6A and S3A–S3D). Subunits with conformation II (chains B, D, F) alternate with subunits in conformation III (chains A, C, E) around the ring, with the remaining chain G having conformation I (S3A–S3D Fig). Superposition of the subunits in conformations II and III showed that the apical domain can pivot by ~22° around the joint with the intermediate domain (as determined with the program DynDom [49]) (S3E Fig). Notably, in four of the subunits (chains B, D, F and G), the tip of the α K- α L helical hairpin protruding from the apical domain is oriented as in crystal form I (Figs 6C and S3B–S3D). In chains A, C and E (conformation III) the apical domain is re-oriented, allowing residues 290–294 to contact the tip of the intermediate domain of the adjacent subunit (chains B, D and F; conformation II) (Fig 6D and 6E). This re-orientation of the apical domains of chains A, C and E also results in the α K- α L helical hairpin to point outwards into the solvent (Figs 6D and S3C). The apical-intermediate domain intra-ring contact also requires a 19–21° outward rotation of the intermediate domains in chains B, D and F (Figs 6C and S3B). All other inter-subunit contacts are limited to the equatorial domains (Fig 6F).

All subunits in crystal form II contained electron density for a bound nucleotide (Fig 6G and 6H). The orientation of the intermediate domain correlated with the identity of the bound nucleotide: The subunits with outward-oriented intermediate domain (chains B, D and F) contained weak nucleotide density, consistent with partial occupancy by ADP (Fig 6G). The β -phosphate of ADP was coordinated by the amides at the N-terminal end of helix α D; additional density for BeF_x was not detectable. The nucleotide electron density for the subunits with inward-oriented intermediate domain (chains A, C, E and G) was consistent with nucleoside triphosphate, modelled as ATP and Mg²⁺ (Fig 6H). Compared to crystal form I, the catalytic residue Asp412 was more distant from the γ -phosphate (~8.5 Å) in these subunits. The nucleotide binding pattern is thus alternating in the ring, with the exception of the adjacent subunits G and A, which both harbor ATP and Mg²⁺. Interestingly, in two of the ADP-bound subunits (chains B and D), the tip of the α A- β 2 loop, which cradles the nucleotide in the ATP-bound subunits, was remodeled and flipped away from ADP (compare Fig 6B and 6G). This conformational change might facilitate dissociation of ADP from the chaperonin.

Cryo-EM structure of ϕ EL in absence of nucleotide

To obtain the solution structure of apo- ϕ EL, we next performed cryo-EM and single particle analysis. The raw micrographs and the particle 2D class averages suggested the presence of

double-ring structures with approximate seven-fold rotational symmetry (S4A and S4B Fig). The particles appeared to have a strong tendency to associate via their apical domains (S4A Fig). Symmetry-free 3D class averaging indicated a tetradecameric double-ring structure with C2 symmetry (S4C Fig). Refinement of the particles resulted in a density map at 3.45 Å resolution (Figs 7A and S4D), which allowed construction of an atomic model (Fig 7B and Table 2). The contacts between subunits were well defined in the density (Fig 7A–7C). Similar to the crystal structures, the apical domains were the most mobile elements of the structure (S4E Fig). Apo- ϕ EL has an open structure with a ring opening diameter at the apical domains of ~65 Å (Fig 7A). This is in contrast to the previously reported 9 Å cryo-EM map of apo- ϕ EL (EMD-6494) in which the ring opening is reduced to ~27 Å due to rearrangement of the apical domains [10]. Overall each ring of the apo- ϕ EL structure reported here is similar to the ϕ EL heptamer of crystal form II (r.m.s.d. C α positions 2.4 Å, versus 4.2 Å with crystal form I) in that the subunits alternate between conformations II (chains B, D, F) and III (chains A., C, E) and one subunit (chain G) is in conformation I (S5A–S5D Fig). Similar conformations of alternating subunits were recently observed by cryo-EM for the apo-state of the chaperonin of *Pseudomonas* phage OBP [50]. In all subunits of ϕ EL the tip of the α A- β 2 loop in the equatorial domain is moved away from the nucleotide binding site as in chains B and D of crystal form II (Fig 6G). Other nucleotide binding elements such as the P-loop (residues 79–86) and the β 17- β 18 hairpin (residues 503–512) are also moved outward (~1.5 Å) in the apo structure. These local conformational changes might facilitate ATP-binding.

Similar to group II chaperonins and unlike the staggered subunit contacts in group I, the ring-ring interface of apo- ϕ EL is formed by two-fold symmetrical contacts between subunits. This results in the burial of 489–537 Å² of accessible surface area (Fig 7C and 7D). For comparison, the buried surface area between subunits within heptamer rings is 1218–1401 Å². The inter-ring contacts involve residues 492–494 in one subunit forming van-der-Waals contacts with residues 460* and 461* of the subunit in the opposite ring (Fig 7D). The negative dipole at the C-terminal end of helix α N2 is positioned close to His492*. Residues Asn465 and Asn465* form a symmetrical contact; Glu466 makes polar contacts with Arg469* and the backbone at Gly462*. Moreover, there is a symmetrical van-der-Waals contact with the neighboring subunit in the opposite ring, involving residues Lys486 (Fig 7D). None of these contacts are observed in group I and II chaperonins due to the high divergence of sequences.

Cryo-EM structures of ϕ EL in presence of ADP or ATP γ S

To further elucidate the allosteric cycle of ϕ EL, we determined the solution structure of ϕ EL in the presence of ADP by single-particle cryo-EM. ϕ EL formed double-ring structures with seven-fold rotational symmetry (S6A and S6B Fig). Symmetry-free 3D class averaging indicated an open double-ring structure with nearly uniform rings, and thus D7 symmetry was applied, resulting in a refined cryo-EM density map at 5.9 Å resolution (Figs 8A and S6C). Note that the resolution is simply limited by the experimental setup. No particle class with closed single rings was found, in contrast to a previous report showing that ϕ EL•ADP (EMD-6493) forms single rings resembling a hollow sphere [10] (S6B Fig). We used the domain structures from the 3.54 Å crystal structure to rigid-body fit and refine a pseudo-atomic molecular model of the ADP-bound ϕ EL double-ring (Fig 8A and Table 2). In all the subunits, the α K- α L helical hairpin in the apical domain contacts the equatorial domain and the intermediate domain is in the inward-rotated conformation (subunit conformation I). The ring-ring interface was similar to that of apo- ϕ EL (r.m.s.d. 1.4 Å for C α atoms of all 14 equatorial domains) (Fig 8B).

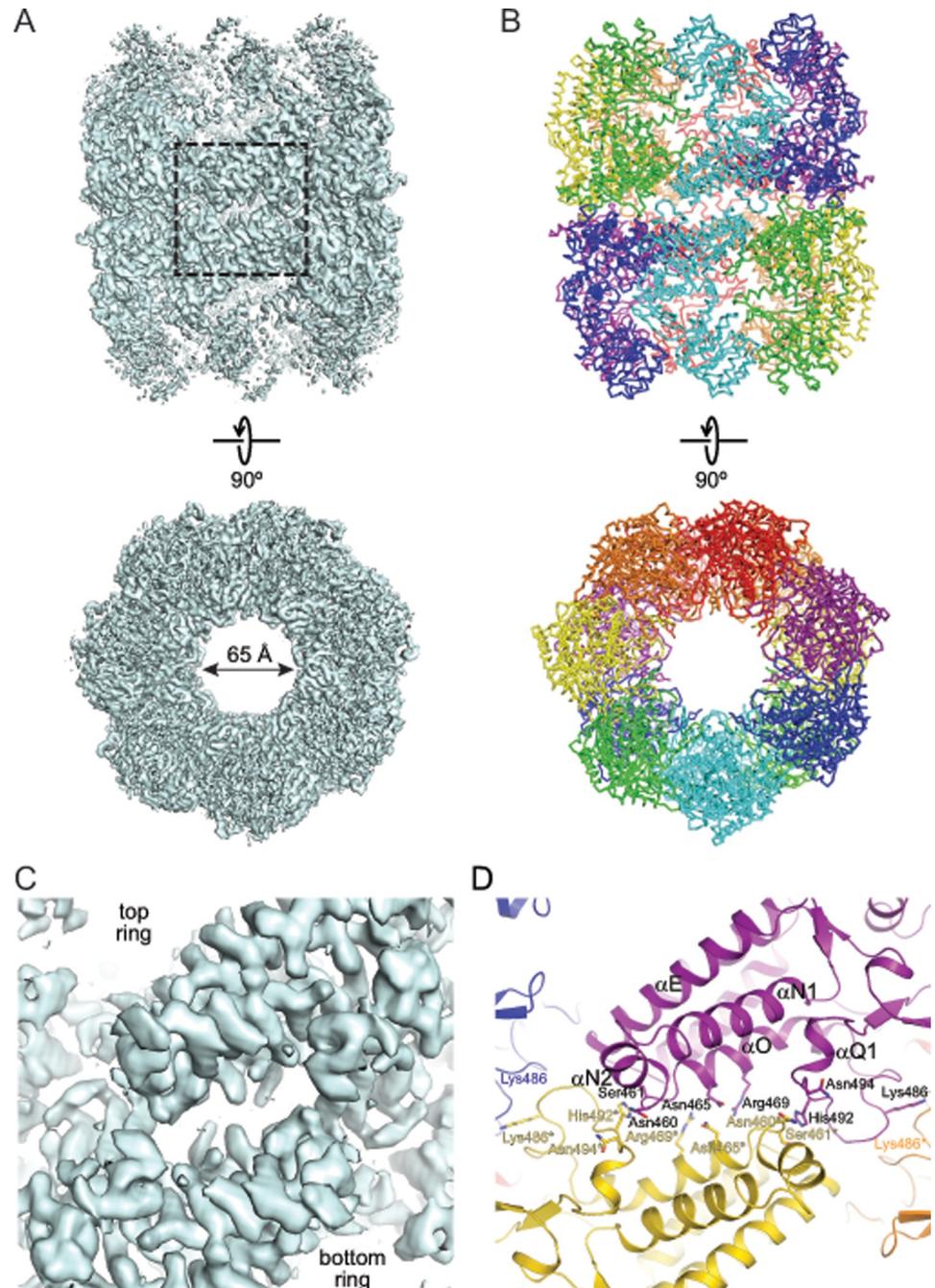


Fig 7. Cryo-EM structure of apo- ϕ EL. (A) Cryo-EM density map of apo- ϕ EL at 3.45 Å resolution. Perpendicular views of an isocontour surface at 4.8σ is shown. The top view is along the two-fold symmetry axis. The box highlights the inter-subunit interface at the equator of the complex. (B) Atomic model of apo- ϕ EL. Backbone traces of the subunits are shown. Symmetry-equivalent subunits are shown in the same color. (C, D) Zoom on the symmetrical inter-subunit interface at the equator of the complex. Panel C shows the cryo-EM density. Panel D the model in ribbon representation. Contact sidechains are indicated and shown in stick representation (Glu466 is hidden behind the α O helix ribbon in this projection). Secondary structure elements participating in the interactions.

<https://doi.org/10.1371/journal.pone.0230090.g007>

To mimic the conformational changes upon ATP-binding in solution, we also determined the cryo-EM structure of ϕ EL in the presence of the slowly hydrolyzing ATP analog, ATP γ S.

Table 2. Cryo-EM data collection and model statistics.

| ϕ EL complex | Apo | ADP | ATP γ S |
|---|--------------|--------------|----------------|
| EMDB number | 10528 | 10529 | 10530 |
| PDB code | 6TMV | 6TMW | 6TMX |
| Data Collection | | | |
| Magnification | 81,000 | 73,000 | 73,000 |
| Voltage (kV) | 300 | 200 | 200 |
| Electron exposure (e ⁻ Å ⁻²) | 78 | 43 | 42 |
| Defocus range (μm) | -1.5 to -3.0 | -1.5 to -4.0 | -1.5 to -3.5 |
| Pixel size (Å) | 1.090 | 1.997 | 1.997 |
| Symmetry imposed | C2 | D7 | D7 |
| Auto-picked features | 690,973 | 429,456 | 283,285 |
| Initial particle number | 356,175 | 408,474 | 137,262 |
| Final particle number | 178,107 | 170,957 | 52,885 |
| Map resolution (Å) | 3.45 | 5.9 | 5.8 |
| Map sharpening B factor (Å ²) | -122 | -200 | -200 |
| Model | | | |
| Resolution range (Å) | 209–3.45 | 192–5.9 | 192–5.8 |
| Average Fourier shell correlation | 0.824 | 0.925 | 0.875 |
| No. atoms | | | |
| Protein | 58106 | 58058 | 58226 |
| Ligands | – | 378 | 462 |
| Average B-factors | | | |
| Protein (Å ²) | 209 | 286 | 298 |
| Ligands (Å ²) | – | 120 | 150 |
| R.m.s. deviations | | | |
| Bond length (Å) | 0.007 | 0.010 | 0.008 |
| Bond angles (°) | 1.131 | 1.433 | 1.114 |
| Ramachandran plot | | | |
| Favored (%) | 96.1 | 94.2 | 96.1 |
| Allowed (%) | 3.9 | 5.6 | 3.9 |
| Outliers (%) | 0 | 0.2 | 0 |

<https://doi.org/10.1371/journal.pone.0230090.t002>

Double-ring structures were again observed (S6D and S6E Fig) and symmetry-free 3D class averaging indicated nearly uniform rings. Applying D7 symmetry resulted in a refined cryo-EM density map at 5.8 Å resolution (Figs 8C and S6F). The subunits assumed conformation I and were slightly tilted outward compared to ϕ EL•ADP (Fig 8D). Interestingly, we find a rotation at the ring-ring interface of 2.9° with a vertical displacement of 2.9 Å (Fig 8D). This re-orientation would break the contacts between Asn465 and Glu466 with Asn465* and Arg469* / Gly462*, respectively (Fig 7D). Moreover, the symmetrical van-der-Waals contact at the Lys486 residues can no longer form. Thus, binding of ATP γ S weakens the ring-ring interface and this effect may be more pronounced in the presence of ATP, consistent with the finding of mainly single-rings in solution (Fig 1C).

Comparison of ϕ EL with GroEL

Compared to GroEL of *E. coli*, ϕ EL has numerous sequence insertions and deletions (S7 Fig). Only the nucleotide binding pocket is highly conserved. The apical domains of ϕ EL differ in orientation from the apical domains of GroEL in both the open and GroES bound states (Fig

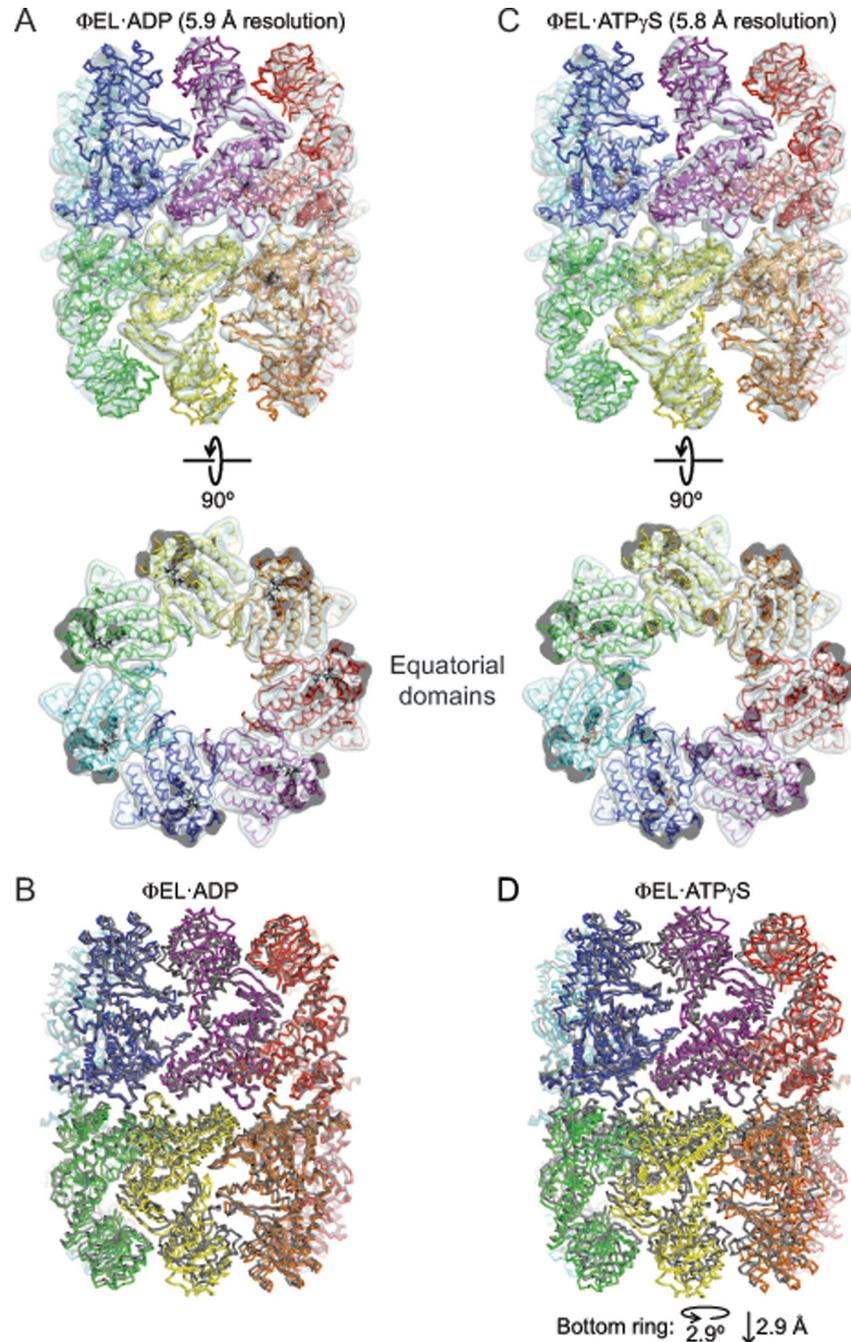


Fig 8. CryoEM structures of ϕ EL in complex with ADP or ATP γ S. (A) Superposition of the cryo-EM density map and the structural model of the ϕ EL•ADP complex. The pseudo-atomic model is shown as backbone trace, with the subunits colored individually. ADP is shown in stick representation. The cryo-EM density map is shown as an isocontour surface at 4.5 σ . Two perpendicular views are shown. In the bottom view, only the equatorial domain section is shown to demonstrate the quality of the fit. (B) Superposition of the ϕ EL•ADP complex with apo- ϕ EL, showing the unaltered ring-ring interface. Backbone traces are shown. The ϕ EL•ADP complex is shown in rainbow colors, and apo- ϕ EL in grey. (C) Superposition of the cryo-EM density map and the structural model of the ϕ EL•ATP γ S complex. The representation is the same as in panel A. (D) Superposition of the ϕ EL•ATP γ S complex with apo- ϕ EL, showing the changes at the ring-ring interface. Backbone traces are shown. The ϕ EL•ATP γ S complex is shown in rainbow colors, and apo- ϕ EL in grey.

<https://doi.org/10.1371/journal.pone.0230090.g008>

9A–9C), but share the same secondary structure topology, i.e. the β -sandwich core structure, α -helices α H, α I and α J at the surface and the helical hairpin formed by helices α K and α L (Fig 9D and 9E). However, the length of helices α H, α J and α K and of the connecting surface loops differ substantially, resulting in a re-orientation of helices and re-modelling of long surface loops, such as residues 191–215 and 306–331, respectively (Fig 9D, boxed areas). The groove between helices α H and α I, which forms the binding site for substrate and GroES in GroEL, is more narrow and less deep in ϕ EL. Accommodation of a helix or a β -hairpin element from SP would require structural remodeling of this site in ϕ EL. In crystal form I, the putative SP binding site is solvent exposed, consistent with the ability of ϕ EL to bind SP in the presence of ATP (Fig 4A). In contrast, in the apo- ϕ EL (and crystal form II), helices α H and α I of chains A, C and E are partially occluded by the respective adjacent apical domain. The contact between the tip of the α K- α L helical hairpin to the equatorial domain in ϕ EL is absent in GroEL, and the α K- α L connection is elongated and re-modelled in ϕ EL (Fig 9A–9C). The intermediate domains of GroEL and ϕ EL are quite similar (Fig 9F and 9G), consistent with a conserved function in coupling domain movements with changes of the nucleotide status of the equatorial domain.

The equatorial domains of ϕ EL and GroEL differ mainly at the ring-ring interface contacts, with ϕ EL making 1:1 subunit interactions and GroEL a 1:2 staggered interaction (Fig 9H and 9I). Helix α D is shortened in ϕ EL by one turn compared to GroEL and helix α N is elongated by two turns (α N1), followed by insertion of a short helix, α N2. Helix α Q in GroEL (residues 462–471) is replaced by a loop connection in ϕ EL (residues 479–503). This loop exhibits structural plasticity in the crystal structures of ϕ EL single-ring.

Inter-domain salt bridges within subunits (equatorial domain D83 to apical domain K327) and between subunits (intermediate domain R197 to apical domain E386), which are important in allosteric regulation of GroEL [38], are not conserved in ϕ EL.

Discussion

Our structural and functional analysis of the chaperonin ϕ EL from the bacteriophage EL of *P. aeruginosa* revealed that the protein is ATPase active and functions in aggregation prevention of denatured proteins. ϕ EL forms tetradecameric double ring complexes, which dissociate into a population of single rings in the presence of ATP and at physiological salt concentration. In contrast, the recently observed dissociation of GroEL into single rings occurs only transiently during the reaction cycle [4]. The nucleotide bound single-ring complexes in the crystal structures of ϕ EL closely resembled the individual rings of the double-ring complexes analyzed by cryo-EM. We could not confirm the existence of a sphere-like single-ring structure proposed to function in SP encapsulation [10]. Our functional data rather suggests that the chaperone mechanism of ϕ EL is encapsulation independent, and represents an evolutionary precursor of the more complex encapsulation mechanisms used by the group I and II chaperonins [2, 5].

Based on our structural and functional analysis, we propose the following hypothetical model for the chaperonin cycle of ϕ EL in SP binding and release, coupled to transitions between single- and double-ring complexes. Unlike group I and II chaperonins, the ability of ϕ EL to prevent protein aggregation, as tested with Rho as a model SP, was strictly nucleotide-dependent. In the apo-state, ϕ EL is a double ring with six of the seven subunits per ring assuming alternating states (conformations II and III) and one subunit in conformation I (Fig 10). The putative SP binding sites in the apical domains of subunits A, C and E are partially occluded. This form acquires competence in SP-binding upon binding of ATP, whereupon the apical domains of all seven subunits per ring are shifted to the same state (conformation I) and the equatorial domains are poised for ATP hydrolysis (Fig 10). ATP-binding weakens the

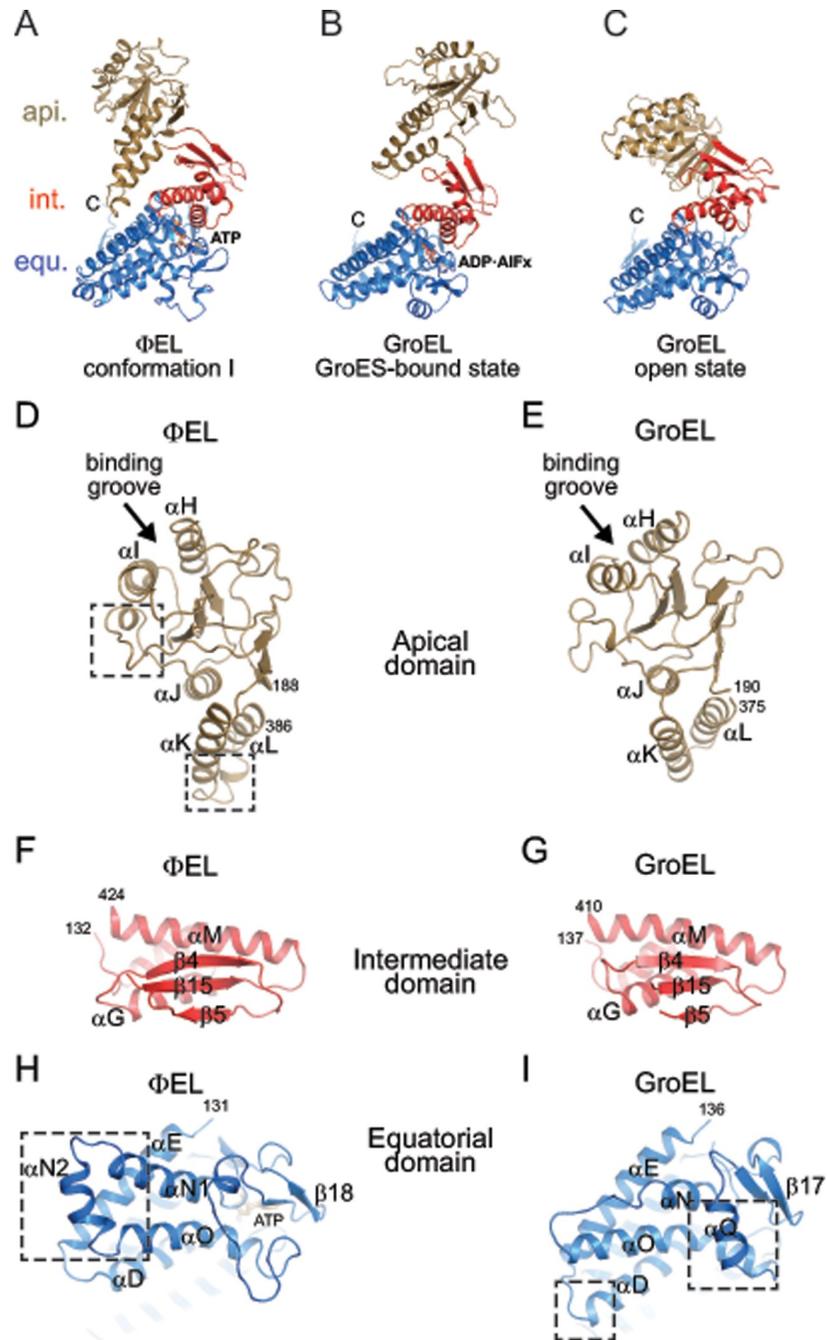


Fig 9. Comparison of the structures of ϕ EL and GroEL. (A) Subunit structure of ϕ EL (conformation I). The representation is the same as in Fig 6. (B, C) Structures of GroEL subunits from the *cis* (B) and *trans* rings (C) of the symmetric GroEL:GroES₂ complex (PDB 1PCQ) [51]. The same representation style as in panel A. (D, E) Structures of the apical domains in ϕ EL (D) and GroEL (E). Ribbon representations are shown. Remodeled regions in ϕ EL discussed in the text are highlighted in boxes. Chain termini and α -helices are indicated. Arrows point to the proposed substrate binding site in group I chaperonins. For GroEL, the PDB dataset 1XCK (apo, open state) [52] was used. (F, G) Structures of the intermediate domains in ϕ EL (F) and GroEL (G). Chain termini and selected secondary structure elements are indicated. (H, I) Structures of the equatorial domains in ϕ EL (H) and GroEL (I). Ribbon representations are shown. A view from the ring-ring interface is shown. Large insertions in the respective structure involved in inter-ring contacts are highlighted in boxes.

<https://doi.org/10.1371/journal.pone.0230090.g009>

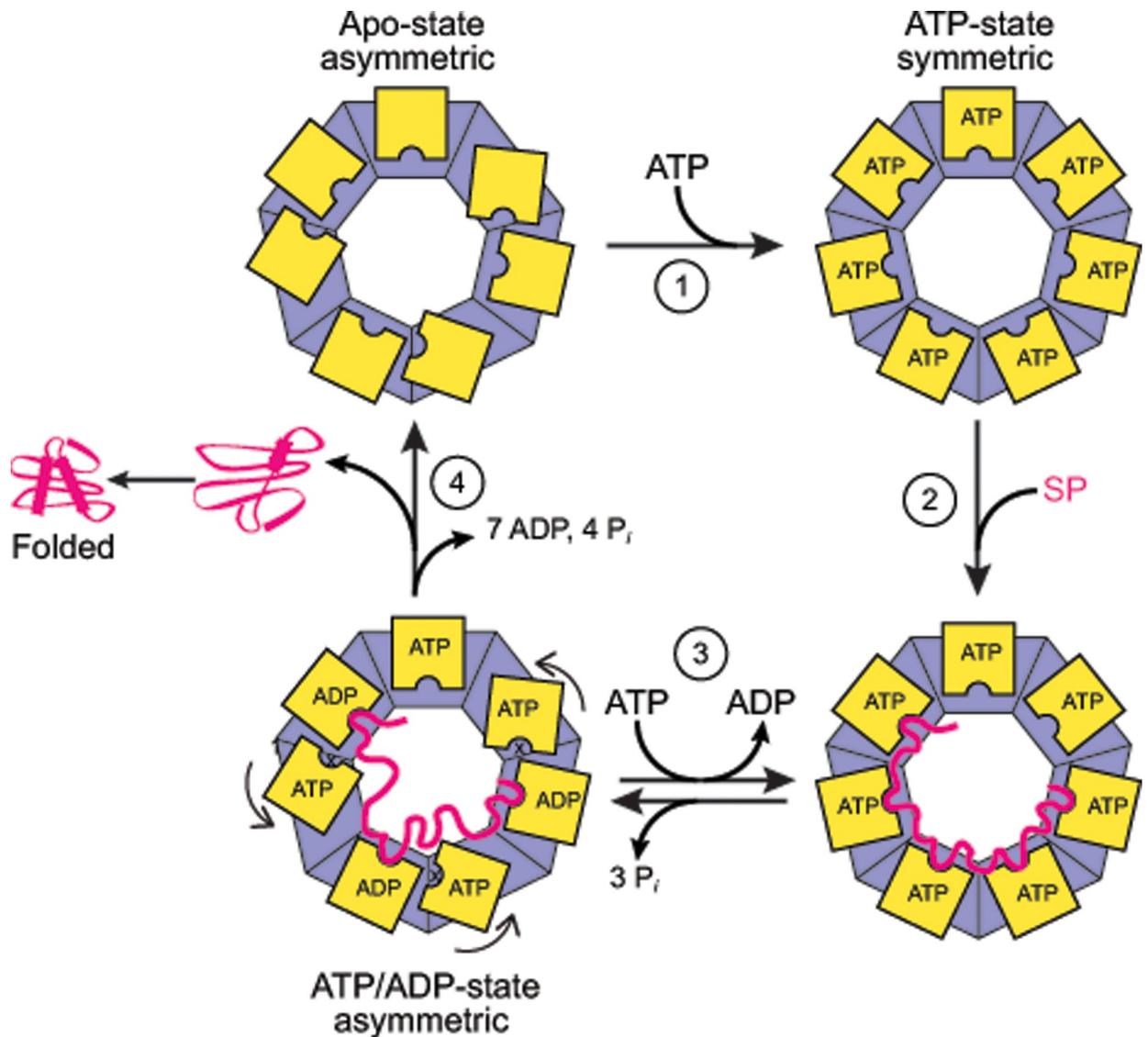


Fig 10. Hypothetical model for the ATPase and substrate interaction cycle of ϕ EL. ϕ EL is shown schematically as a top view with the apical domains in yellow. The double-ring apo-state with asymmetric apical domain orientation has only low affinity for SP. It is converted by ATP binding (1) to the single-ring characterized by symmetric apical domain topology and high binding affinity for SP (2). ATP hydrolysis in three alternating subunits (3) may result in partial SP release, with ATP hydrolysis in the remaining four subunits (4) completing SP release and folding.

<https://doi.org/10.1371/journal.pone.0230090.g010>

ring-ring interface, resulting in single-ring formation at physiological salt concentration. SPs with a lower requirement for the availability of binding sites on adjacent subunits may also bind to the apo-state of ϕ EL. ATP hydrolysis appears to occur in two stages: in the first stage, ATP is hydrolyzed in three alternating subunits, resulting in a ring conformation closely similar to the apo-state. This step may weaken the interaction with bound SP, perhaps allowing partial folding of some SPs. In the second stage, the remaining four subunits hydrolyze their ATP, followed by ADP dissociation generating the double-ring apo-state. This step presumably results in complete SP release for folding, as the apo-state is not SP-binding competent. Folding failed in the case of Rho, presumably because rebinding to ϕ EL is faster than folding, consistent with the dependence of this slow folding protein on the encapsulation mechanism

provided by GroEL/ES. We note that Rho is not a natural substrate of ϕ EL and it would be important in future studies to investigate the mechanism of ϕ EL in folding its cognate substrate proteins. Our structural analysis did not reveal an encapsulating state for ϕ EL. Moreover, ϕ EL did not functionally interact with GroES and does not contain a built-in-lid extension in its apical domains. However, the existence of a distant GroES homolog in the phage EL genome or an unrelated co-factor cannot be ruled out at this point. Distantly related GroES homologs have been identified in other phage genomes [53].

Only some bacteriophage genomes encode their own chaperonin, presumably to assist in the folding of an essential phage protein(s) that cannot utilize (or does so only partially) the host chaperonin system for folding [10, 54]. Alternatively, ϕ EL may prevent the folding of a phage protein up to a point when its function is required. For example, the putative ϕ EL substrate gp188 is a cell wall endolysin [9] that should only function once a large number of phage particles have been produced by the host cell. ϕ EL may stabilize gp188 in a non-native, inactive state until ATP levels may become depleted at the peak of phage production, resulting in concerted release from ϕ EL and activation. It is also possible that additional regulatory factors play a role in triggering gp188 release. Future investigations will have to test the feasibility of such scenarios.

Supporting information

S1 Fig. SDS-PAGE analysis of the ϕ EL preparation. (A) A Coomassie Blue stained gel is shown. Purified GroEL and GroES were also analyzed. The positions of molecular weight markers are indicated. (B) Absence of effect of GroES on the ϕ EL ATPase. ATPase measurements were performed with increasing molar excess of GroES heptamer over ϕ EL tetradecamer, as indicated. (C) Alignment of amino acid sequences of GroES from *E. coli* and *P. aeruginosa*. The mobile loop contact residues of *E. coli* GroES with GroEL in the crystal structure of the GroEL: GroES complex (pdb code 1pcq) are indicated by asterisks (cyan). TT, beta turn. (EPS)

S2 Fig. Size-exclusion chromatography ϕ EL-rhodanese reactions. (A) ϕ EL was incubated with denatured Rho in the presence of ATP for 30 min as in Fig 4C. The reaction was then analyzed by SEC with ATP in the column buffer, followed by Ponceau staining (top) and anti-Rho immunoblotting (bottom). (B and C) Rho folding reactions with GroEL/GroES in the absence of ATP (B) and in the presence of ATP (C). (D) Native Rho was analyzed as a control in the absence of chaperonin. (EPS)

S3 Fig. Subunit conformers found in ϕ EL crystal structures. (A) Superposition of subunit conformers found in crystal form I. Perpendicular views of backbone traces are shown. The subunit colors are the same as in Fig 5. Bound ligand is shown in stick representation. N- and C-termini are indicated. The increased heterogeneity in the apical domain orientations (top) is caused by contacts in the crystal lattice. (B-D) Superposition of subunit conformers found in crystal form II. The subunits are grouped according to their conformation: Panels B, C and D show subunits in conformation II, III and I, respectively. Curved arrows indicate domain reorientations relative to conformation I. (E) Domain movements in ϕ EL analyzed by DynDom. Two perpendicular views of a superposition of conformations II and III of ϕ EL are shown. The rotation axis is shown as a black arrow. The apical domain (api.), joint region (joint) and remainder of the subunit (rest) in conformation III are shown in red, magenta and blue, respectively. Conformation II is shown in grey. The reorientation of the intermediate domain was below the threshold in the DynDom analysis. (EPS)

S4 Fig. Cryo-EM analysis of apo- ϕ EL. (A) Motion-corrected and dose-weighted raw micrograph of apo- ϕ EL. For better contrast, the image was low pass filtered. The scale bar indicates 200 Å. (B) 2D class averages of apo- ϕ EL particles. (C) Symmetry-free (top) and C2-symmetry averaged (bottom) 3D classes. The fraction of particles and the estimated resolution of the 3D classes are indicated. The refined classes are boxed. (D) Gold-standard FSC corrected curve with unmasked map. The resolution was ~ 3.45 Å at the FSC cutoff of 0.143. (E) Local resolution map for the apo- ϕ EL structure. Local resolutions between 7 and 3 Å are represented as a rainbow color gradient from red to blue.

(EPS)

S5 Fig. Conformational analysis of the apo- ϕ EL structure. (A) Comparison of the subunits in the apo- ϕ EL complex with crystal form I. Superposed backbone traces of only the apical and intermediate domains in a single ring are shown for clarity. The subunit colors of apo- ϕ EL are the same as in Fig 7B. The crystal structure has grey color. Relative domain movements and conformation assignments are indicated. (B-D) Superposition of subunit conformers found in the cryo-EM structure of apo- ϕ EL. The subunits are grouped according to their conformation: Panels B, C and D show subunits in conformation II, III and I, respectively. The subunit colors are the same as in Fig 7B.

(EPS)

S6 Fig. Cryo-EM analysis of the complexes of ϕ EL with ADP and ATP γ S. (A, D) Motion-corrected and dose-weighted raw micrographs of ϕ EL•ADP (panel A) and ϕ EL•ATP γ S (D). The scale bar indicates 500 Å. (B, E) 2D class averages of ϕ EL•ADP (B) and ϕ EL•ATP γ S (E) particles. (C, F) Gold-standard FSC corrected curve with unmasked maps of ϕ EL•ADP (C) and ϕ EL•ATP γ S (F). The respective resolutions were ~ 5.9 Å and ~ 5.8 Å at the FSC cutoff of 0.143.

(EPS)

S7 Fig. Sequence alignment of ϕ EL and GroEL from *E. coli*. Secondary structure elements for ϕ EL and GroEL from *E. coli* are indicated above and below the sequences, respectively. Similar residues are shown in red and identical residues in white on a red background. Blue frames indicate homologous regions. The Uniprot accession codes for the sequences are: Q2Z0T5, chaperonin from *Pseudomonas* phage EL; P0A6F5, GroEL from *E. coli*.

(EPS)

S8 Fig. View onto the apical domains of GroEL and ϕ EL showing the binding groove for substrate protein and GroES mobile loop. Apical domains of GroEL (A) and ϕ EL (B) in ribbon representation (left) and surface representation (right) showing the distribution of hydrophobic and charged residues. In the GroEL apical domain, helices α I and α H are separated by a wide, deep cleft, revealing the hydrophobic underlying segment. In the ϕ EL apical domain, the respective helices contact each other, resulting in a shallow furrow. The underlying hydrophobic segment is occluded. Positively and negatively charged functional groups are indicated in blue and red, respectively, hydrophobic side chains in yellow and the rest of the surface in white.

(EPS)

Acknowledgments

We thank Lidia Kurochkina from the Russian Academy of Sciences, Moscow, Russia for providing us with the expression plasmid for ϕ EL. We thank the JSBG group at the ESRF synchrotron in Grenoble, France for maintaining the beamlines and help with data collection, and the

staff at MPIB crystallization facility for performing automated crystallization screening. Expert assistance by Tillman Schäfer and Daniel Bollschweiler from the MPIB cryo-EM facility is gratefully acknowledged.

Author Contributions

Conceptualization: F. Ulrich Hartl, Manajit Hayer-Hartl.

Data curation: Andreas Bracher, Simanta S. Paul, Huping Wang, Nadine Wischnewski, Manajit Hayer-Hartl.

Formal analysis: Andreas Bracher, Huping Wang, Nadine Wischnewski, Manajit Hayer-Hartl.

Investigation: Andreas Bracher, Simanta S. Paul, Nadine Wischnewski, Manajit Hayer-Hartl.

Project administration: Andreas Bracher, Manajit Hayer-Hartl.

Supervision: Andreas Bracher, F. Ulrich Hartl, Manajit Hayer-Hartl.

Validation: Andreas Bracher, Nadine Wischnewski, Manajit Hayer-Hartl.

Visualization: Andreas Bracher, Manajit Hayer-Hartl.

Writing – original draft: Andreas Bracher, F. Ulrich Hartl, Manajit Hayer-Hartl.

Writing – review & editing: Andreas Bracher, Simanta S. Paul, Huping Wang, Nadine Wischnewski, F. Ulrich Hartl, Manajit Hayer-Hartl.

References

1. Balchin D, Hayer-Hartl M, Hartl FU. *In vivo* aspects of protein folding and quality control. *Science*. 2016; 353: aac4354. <https://doi.org/10.1126/science.aac4354> PMID: 27365453
2. Hayer-Hartl M, Bracher A, Hartl FU. The GroEL–GroES chaperonin machine: A nano-cage for protein folding. *Trends Biochem Sci*. 2016; 41: 62–76. <https://doi.org/10.1016/j.tibs.2015.07.009> PMID: 26422689
3. Nielsen KL, Cowan NJ. A single ring is sufficient for productive chaperonin-mediated folding *in vivo*. *Mol Cell*. 1998; 2: 93–9. [https://doi.org/10.1016/s1097-2765\(00\)80117-3](https://doi.org/10.1016/s1097-2765(00)80117-3) PMID: 9702195
4. Yan X, Shi Q, Bracher A, Milicic G, Singh AK, Hartl FU, et al. GroEL ring separation and exchange in the chaperonin reaction. *Cell*. 2018; 172: 605–17. <https://doi.org/10.1016/j.cell.2017.12.010> PMID: 29336887
5. Lopez T, Dalton K, Frydman J. The mechanism and function of group II chaperonins. *J Mol Biol*. 2015; 427: 2919–30. <https://doi.org/10.1016/j.jmb.2015.04.013> PMID: 25936650
6. van der Vies SM, Gatenby AA, Georgopoulos C. Bacteriophage T4 encodes a co-chaperonin that can substitute for *Escherichia coli* GroES in protein folding. *Nature*. 1994; 368: 654–6. <https://doi.org/10.1038/368654a0> PMID: 7908418
7. Marine RL, Nasko DJ, Wray J, Polson SW, Wommack KE. Novel chaperonins are prevalent in the viroplankton and demonstrate links to viral biology and ecology. *ISME J*. 2017; 11: 2479–91. <https://doi.org/10.1038/ismej.2017.102> PMID: 28731469
8. Hertveldt K, Lavigne R, Pleteneva E, Sernova N, Kurochkina L, Korchevskii R, et al. Genome comparison of *Pseudomonas aeruginosa* large phages. *J Mol Biol*. 2005; 354: 536–45. <https://doi.org/10.1016/j.jmb.2005.08.075> PMID: 16256135
9. Kurochkina LP, Semenyuk PI, Orlov VN, Robben J, Sykilinda NN, Mesyanzhinov VV. Expression and functional characterization of the first bacteriophage-encoded chaperonin. *J Virol*. 2012; 86: 10103–11. <https://doi.org/10.1128/JVI.00940-12> PMID: 22787217
10. Molugu SK, Hildenbrand ZL, Morgan DG, Sherman MB, He L, Georgopoulos C, et al. Ring separation highlights the protein-folding mechanism used by the Phage EL-encoded chaperonin. *Structure*. 2016; 24: 537–46. <https://doi.org/10.1016/j.str.2016.02.006> PMID: 26996960

11. Tafoya DA, Hildenbrand ZL, Herrera N, Molugu SK, Mesyanzhinov VV, Miroshnikov KA, et al. Enzymatic characterization of a lysin encoded by bacteriophage EL. *Bacteriophage*. 2013; 3: e25449. <https://doi.org/10.4161/bact.25449> PMID: 24228221
12. Wyatt PJ. Light-scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta*. 1993; 272: 1–40. [https://doi.org/10.1016/0003-2670\(93\)80373-S](https://doi.org/10.1016/0003-2670(93)80373-S)
13. Poso D, Clarke AR, Burston SG. A kinetic analysis of the nucleotide-induced allosteric transitions in a single-ring mutant of GroEL. *J Mol Biol*. 2004; 338: 969–77. <https://doi.org/10.1016/j.jmb.2004.03.010> PMID: 15111060
14. Weber F, Hayer-Hartl M. Refolding of bovine mitochondrial rhodanese by chaperonins GroEL and GroES. *Methods Mol Biol*. 2000; 140: 117–26. <https://doi.org/10.1385/1-59259-061-6:117> PMID: 11484478
15. Radaev S, Li S, Sun PD. A survey of protein-protein complex crystallizations. *Acta Crystallogr D Biol Crystallogr*. 2006; 62: 605–12. <https://doi.org/10.1107/S0907444906011735> PMID: 16699187
16. Radaev S, Sun PD. Crystallization of protein-protein complexes. *J App Cryst*. 2002; 35: 674–6.
17. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr*. 2010; 66: 125–32. <https://doi.org/10.1107/S0907444909047337> PMID: 20124692
18. Evans P. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr*. 2006; 62: 72–82. <https://doi.org/10.1107/S0907444905036693> PMID: 16369096
19. Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr*. 2013; 69: 1204–14. <https://doi.org/10.1107/S0907444913000061> PMID: 23793146
20. French G, Wilson K. On the treatment of negative intensity observations. *Acta Cryst Sect A*. 1978; 34: 517–25. <https://doi.org/10.1107/S0567739478001114>
21. Potterton E, Briggs P, Turkenburg M, Dodson E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr*. 2003; 59: 1131–7. <https://doi.org/10.1107/s0907444903008126> PMID: 12832755
22. Matthews BW. Solvent content of protein crystals. *J Mol Biol*. 1968; 33: 491–7. [https://doi.org/10.1016/0022-2836\(68\)90205-2](https://doi.org/10.1016/0022-2836(68)90205-2) PMID: 5700707
23. Vagin AA, Isupov MN. Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr D Biol Crystallogr*. 2001; 57: 1451–6. <https://doi.org/10.1107/s0907444901012409> PMID: 11567159
24. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004; 25: 1605–12. <https://doi.org/10.1002/jcc.20084> PMID: 15264254
25. Terwilliger TC. Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr*. 2000; 56: 965–72. <https://doi.org/10.1107/S0907444900005072> PMID: 10944333
26. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004; 60: 2126–32. <https://doi.org/10.1107/S0907444904019158> PMID: 15572765
27. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr*. 2011; 67: 355–67. <https://doi.org/10.1107/S0907444911001314> PMID: 21460454
28. Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, Agard DA. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods*. 2017; 14: 331–2. <https://doi.org/10.1038/nmeth.4193> PMID: 28250466
29. Rohou A, Grigorieff N. CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol*. 2015; 192: 216–21. <https://doi.org/10.1016/j.jsb.2015.08.008> PMID: 26278980
30. Scheres SH. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol*. 2012; 180: 519–30. <https://doi.org/10.1016/j.jsb.2012.09.006> PMID: 23000701
31. Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJ, Lindahl E, et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife*. 2018; 7. <https://doi.org/10.7554/eLife.42166> PMID: 30412051
32. Schorb M, Haberbosch I, Hagen WJH, Schwab Y, Mastronarde DN. Software tools for automated transmission electron microscopy. *Nat Methods*. 2019; 16: 471–7. <https://doi.org/10.1038/s41592-019-0396-9> PMID: 31086343
33. Mastronarde DN. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol*. 2005; 152: 36–51. <https://doi.org/10.1016/j.jsb.2005.07.007> PMID: 16182563
34. Biyani N, Righetto RD, McLeod R, Caujolle-Bert D, Castano-Diez D, Goldie KN, et al. Focus: The interface between data collection and data processing in cryo-EM. *J Struct Biol*. 2017; 198: 124–33. <https://doi.org/10.1016/j.jsb.2017.03.007> PMID: 28344036

35. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2010; 66: 12–21. <https://doi.org/10.1107/S0907444909042073> PMID: 20057044
36. Kleywegt GT, Jones TA. A super position. *CCP4/ESF-EACBM Newsletter on Protein Crystallography*. 1994; 31: 9–14.
37. Gouet P, Courcelle E, Stuart DI, Metz F. ESPript: multiple sequence alignments in PostScript. *Bioinformatics*. 1999; 15: 305–8. <https://doi.org/10.1093/bioinformatics/15.4.305> PMID: 10320398
38. Gruber R, Horovitz A. Allosteric mechanisms in chaperonin machines. *Chem Rev*. 2016; 116: 6588–606. <https://doi.org/10.1021/acs.chemrev.5b00556> PMID: 26726755
39. Langer T, Pfeifer G, Martin J, Baumeister W, Hartl FU. Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein substrate within its central cavity. *EMBO J*. 1992; 11: 4757–65. PMID: 1361169
40. Martin J, Geromanos S, Tempst P, Hartl FU. Identification of nucleotide-binding regions in the chaperonin proteins GroEL and GroES. *Nature*. 1993; 366: 279–82. <https://doi.org/10.1038/366279a0> PMID: 7901771
41. Kawe M, Plückthun A. GroEL walks the fine line: the subtle balance of substrate and co-chaperonin binding by GroEL. A combinatorial investigation by design, selection and screening. *J Mol Biol*. 2006; 357: 411–26. <https://doi.org/10.1016/j.jmb.2005.12.005> PMID: 16427651
42. Hayer-Hartl MK, Weber F, Hartl FU. Mechanism of chaperonin action: GroES binding and release can drive GroEL-mediated protein folding in the absence of ATP hydrolysis. *EMBO J*. 1996; 15: 6111–21. PMID: 8947033
43. Tang YC, Chang HC, Roeben A, Wischniewski D, Wischniewski N, Kerner MJ, et al. Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. *Cell*. 2006; 125: 903–14. <https://doi.org/10.1016/j.cell.2006.04.027> PMID: 16751100
44. Gupta AJ, Haldar S, Milicic G, Hartl FU, Hayer-Hartl M. Active cage mechanism of chaperonin-assisted protein folding demonstrated at single-molecule level. *J Mol Biol*. 2014; 426: 2739–54. <https://doi.org/10.1016/j.jmb.2014.04.018> PMID: 24816391
45. Martin J, Langer T, Boteva R, Schramel A, Horwich AL, Hartl FU. Chaperonin-mediated protein folding at the surface of GroEL through a 'molten globule'-like intermediate. *Nature*. 1991; 352: 36–42. <https://doi.org/10.1038/352036a0> PMID: 1676490
46. Mayhew M, da Silva AC, Martin J, Erdjument-Bromage H, Tempst P, Hartl FU. Protein folding in the central cavity of the GroEL-GroES chaperonin complex. *Nature*. 1996; 379: 420–6. <https://doi.org/10.1038/379420a0> PMID: 8559246
47. Weissman JS, Rye HS, Fenton WA, Beechem JM, Horwich AL. Characterization of the active intermediate of a GroEL-GroES-mediated protein folding reaction. *Cell*. 1996; 84: 481–90. [https://doi.org/10.1016/s0092-8674\(00\)81293-3](https://doi.org/10.1016/s0092-8674(00)81293-3) PMID: 8608602
48. Brinker A, Pfeifer G, Kerner MJ, Naylor DJ, Hartl FU, Hayer-Hartl M. Dual function of protein confinement in chaperonin-assisted protein folding. *Cell*. 2001; 107: 223–33. [https://doi.org/10.1016/s0092-8674\(01\)00517-7](https://doi.org/10.1016/s0092-8674(01)00517-7) PMID: 11672529
49. Hayward S, Berendsen HJ. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins*. 1998; 30: 144–54. PMID: 9489922
50. Stanishneva-Konovalova TB, Semenyuk PI, Kurochkina LP, Pichkur EB, Vasilyev AL, Kovalchuk MV, et al. Cryo-EM reveals an asymmetry in a novel single-ring viral chaperonin. *J Struct Biol*. 2019; 107439. <https://doi.org/10.1016/j.jsb.2019.107439> PMID: 31870903
51. Chaudhry C, Farr GW, Todd MJ, Rye HS, Brunger AT, Adams PD, et al. Role of the gamma-phosphate of ATP in triggering protein folding by GroEL-GroES: function, structure and energetics. *EMBO J*. 2003; 22: 4877–87. <https://doi.org/10.1093/emboj/cdg477> PMID: 14517228
52. Bartolucci C, Lamba D, Grazulis S, Manakova E, Heumann H. Crystal structure of wild-type chaperonin GroEL. *J Mol Biol*. 2005; 354: 940–51. <https://doi.org/10.1016/j.jmb.2005.09.096> PMID: 16288915
53. Ang D, Keppel F, Klein G, Richardson A, Georgopoulos C. Genetic analysis of bacteriophage-encoded cochaperonins. *Annu Rev Genet*. 2000; 34: 439–56. <https://doi.org/10.1146/annurev.genet.34.1.439> PMID: 11092834
54. Semenyuk PI, Orlov VN, Kurochkina LP. Effect of chaperonin encoded by gene 146 on thermal aggregation of lytic proteins of bacteriophage EL *Pseudomonas aeruginosa*. *Biochemistry* 2015; 80: 172–9. <https://doi.org/10.1134/S0006297915020042> PMID: 25756532

5 Discussion

Many photosynthetic and autotrophic organisms use carbon concentrating mechanisms (CCMs) to improve the enzymatic performance of Rubisco, the key enzyme of carbon fixation in the CBB cycle of photosynthesis. To avoid the loss of Rubisco activity due to inhibition with tight binding sugar phosphates, Rca enzymes are employed for enzyme repair. Research over the past decades provided insight into the protein components of CCM microcompartments, and identified the Rca proteins of distinct Rubisco forms (Badger et al., 1998; Bracher et al., 2017; Rae et al., 2013). However, the detailed mechanisms of microcompartment biogenesis and form IB Rubisco reactivation remained obscure.

The present work revealed how the scaffold protein CcmM performs its essential function in β -carboxysome biogenesis and elucidated a conserved mechanism in form IB Rubisco reactivation in cyanobacteria and plants. CcmM interacts with Rubisco and CcaA multivalently, inducing protein condensate formation. The cyanobacterial Rca (Rca-like) co-assembles into this condensate. Our studies used a combination of biochemical and structural analysis tools to obtain mechanistic information at different scales. The properties of the phase-separated protein condensate were characterized by turbidity measurements and fluorescence microscopy. My collaborator Mirkko Flecken carried out traditional biochemistry assays of ATPase and carbon fixation rates and identified the function of Rca-like protein as a Rubisco activase. Cryo-EM with single-particle analysis (SPA) together with crystallography provided detailed structural information.

In the following sections I discuss the results of our study in the context of recent related publications.

5.1 The requirement for RbcS-containing Rubisco holoenzyme in CCM microcompartment biogenesis

M35, the short isoform of CcmM that consists only of several linked SSUL modules, has been known to interact with Rubisco and to be critical in β -carboxysome biogenesis for years (Kerfeld and Melnicki, 2016; Ludwig et al., 2000). Due to the sequence similarity of the SSUL modules and the small Rubisco subunit RbcS, it seemed plausible that the SSUL functions in linking Rubisco complexes by replacing RbcS. Surprisingly, our structure of the *Synechococcus elongatus* PCC7942 Rubisco-M35 complex in conjunction with functional data revealed that RbcS is required for the interaction of the SSUL module with Rubisco, arguing against the replacement model. Interestingly, a recent structure of *Halothiobacillus neapolitanus* Rubisco in complex with the repeated N-terminal domain peptide of CsoS2 (scaffold protein in α -carboxysome) also showed the participation of RbcS in the interaction (Oltrogge et al., 2020). Furthermore, the scaffolding component in the pyrenoid, EPYC1, was reported to interact with RbcS in the complex with its cognate Rubisco as well (Atkinson et al., 2019; He et al., 2020; Meyer et al., 2012) (Fig. 19). The conserved requirement for RbcS in the interaction with the scaffold protein probably ensures that only fully assembled Rubisco holoenzyme, but not the catalytically inactive RbcL₈ core particles or other Rubisco assembly intermediates, are incorporated into CCM microcompartments.

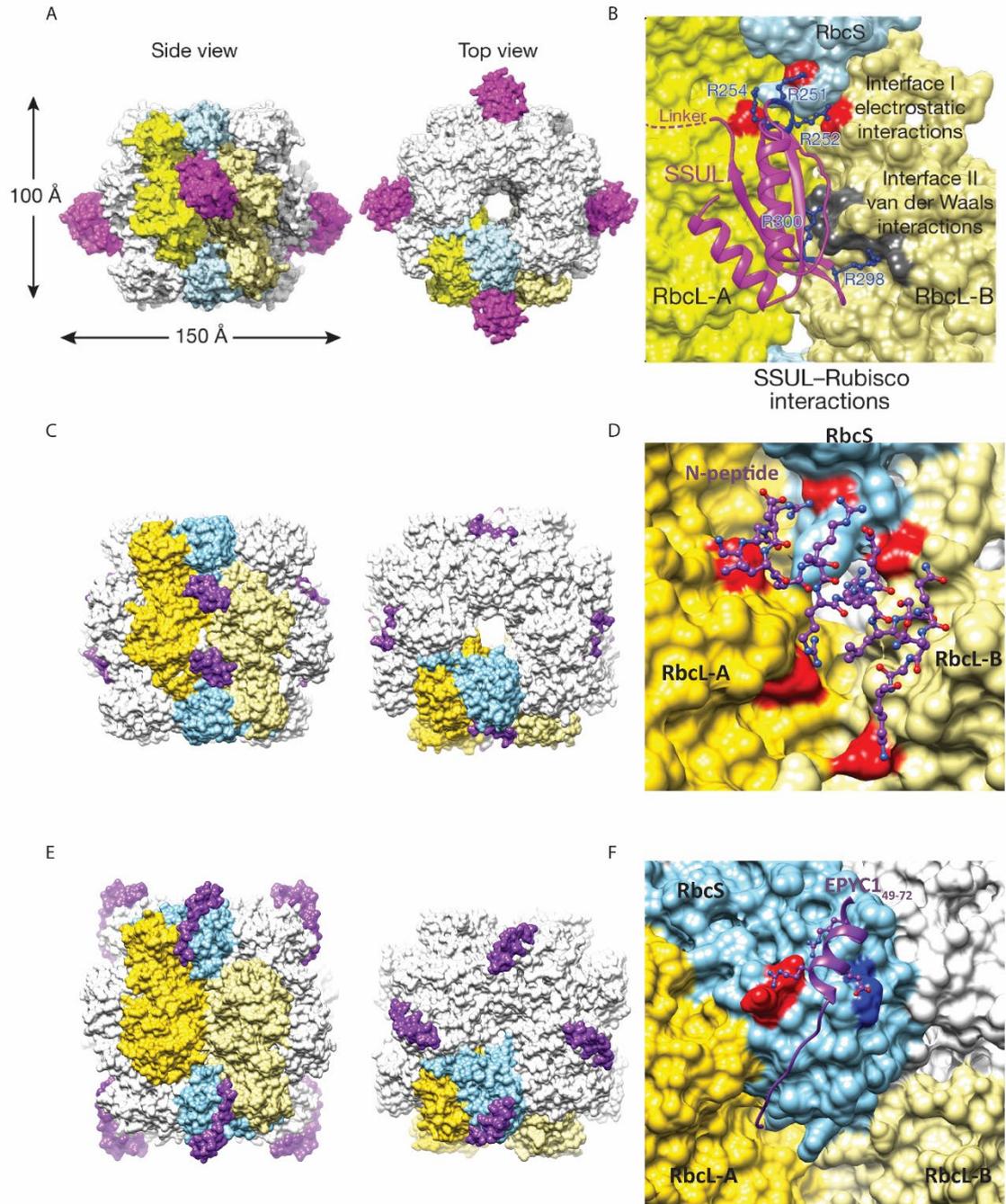


Figure 19: RbcS subunits participate in Rubisco-SSUL, Rubisco-N-CsoS2 and Rubisco-EPYC1₄₉₋₇₂ interactions. (A) Model of the Rubisco–SSUL complex showing a hypothetical alternating arrangement of four SSUL modules (magenta). Side and top views are shown. One 2RbcL–2RbcS–SSUL unit is colored. Gold and pale yellow, RbcL subunits in contact with SSUL (from adjacent anti-parallel RbcL dimers); pale blue, RbcS; magenta, SSUL. (B) Rubisco-SSUL interface. Charge interactions are highlighted. (C) Model of the Rubisco-N-CsoS2 complex (PDB: 6uew), with Rubisco subunits colored as in (A), CsoS2 in violet. (D) Rubisco-N-CsoS2 interfaces. The N-terminal repeated peptide of CsoS2 (with sequence GRDLARARREALSQQGK) is shown in stick representation. Charge interactions are highlighted. (E) Model of the Rubisco-EPYC1₄₉₋₇₂ complex (PDB: 7jfo), with Rubisco

subunits colored as in (A), EPYC1₄₉₋₇₂ in violet. (F) Rubisco- EPYC1₄₉₋₇₂ interfaces. Charge interactions are highlighted. Figure adapted from (Wang et al., 2019), (Oltrogge et al., 2020) and (He et al., 2020).

5.2 Phase separation in the CCM microcompartment

The scaffold proteins exist in both α - and β -carboxysomes as well as in the pyrenoid and they are essential for the function of the respective CCM microcompartment (Cameron et al., 2013; Gonzales et al., 2005; Mackinder et al., 2016). Each scaffold protein contains multiple Rubisco binding sites and can induce Rubisco condensate formation *in vitro*. The pyrenoid matrix is not a crystalline solid and instead behaves like a liquid (Freeman Rosenzweig et al., 2017). Our *in vitro* studies revealed that the Rubisco-M35 condensate also possesses dynamicity, as shown by fluorescence recovery after photobleaching (FRAP) (Wang et al., 2019).

The protein phase separation underlying the formation of the CCM microcompartments is distinct from the well-characterized liquid-like phase separation (LLPS) mediated by intrinsically disordered regions (IDR) (Lin et al., 2017; Uversky, 2017; Wang et al., 2018b). For Rubisco condensates formation, the scaffold protein interacts with Rubisco in a specific binding site (He et al., 2020; Oltrogge et al., 2020; Wang et al., 2019) (Fig. 19), while in the condensates formed with IDRs, the interactions are usually unspecific (Wang et al., 2018b). Furthermore, the Rubisco condensate is less dynamic with a half time of recovery after photobleaching of minutes compared to seconds for condensates mediated by IDRs (Molliex et al., 2015; Wang et al., 2018a).

5.3 The function of the SSUL modules in β -carboxysome biogenesis

The scaffold protein in β -carboxysomes, CcmM, interacts dynamically with Rubisco via its SSUL modules. The interaction is moreover salt and redox sensitive. The salt sensitivity presumably results from two critical salt bridges located at the protein-protein interface (Fig. 19B). The redox regulation is implemented by a pair of structurally juxtaposed cysteine residues, which are found in most SSUL modules. The affinity between SSUL and Rubisco is increased under reducing conditions such as in the cytosol, but decreased under oxidizing conditions as found in the carboxysome lumen. Hence, we proposed a redox regulation mechanism of β -carboxysome assembly and function. The β -carboxysome biogenesis follows an inside-out model, where the formation of the Rubisco condensate in the presence of CcmM would be the first step (Cameron et al., 2013). The reducing environment in the cytosol facilitates Rubisco being incorporated into the protein condensate. After the carboxysome shell has been assembled, cellular reducing agents cannot pass through the shell and the carboxysome lumen becomes oxidizing. Disulfide formation in SSUL under these conditions renders the interaction with Rubisco more dynamic. *In vivo* mutational analysis in carboxysomes (collaboration with the group of Prof. G. Dean Price) confirmed that disulfide formation is required for carboxysome and CCM function, although, the exact mechanism is not yet clear.

The Rca-like protein, located in the β -carboxysome, contains a C-terminal SSUL module in each of its 6 subunits. Our structural and biochemical data showed that the SSUL modules interact with Rubisco in a way similar to the SSUL modules of CcmM. It would be interesting to see if in α -carboxysomes the binding motif of CsoS2 is also employed to recruit additional factors.

5.4 The central role of M58 and M35 in β -carboxysome biogenesis

Both CcmM and CsoS2 have two isoforms. In both cases, the short isoform is sufficient to induce Rubisco condensate formation (Oltrogge et al., 2020; Wang et al., 2019). The requirement for the long isoform is not fully understood. Two CcmM isoforms share a structure of multiple linked SSUL modules. The long isoform has in addition a γ -CA like (γ -CAL) domain at the N-terminus. In *Synechococcus elongatus* PCC7942 and some other cyanobacteria species, this domain lost its CA activity due to mutations in the catalytic residues. We found that the γ -CAL domain forms a trimer and interacts with the functional β -CA, CcaA, to recruit CcaA into the Rubisco condensate. Furthermore, an interaction between M58 and CcmN has been reported. CcmN is hypothesized to recruit the β -carboxysome shell proteins (Kinney et al., 2012). Thus, M58 functions as the central scaffold protein, together with CcmN, to link the pre-carboxysome core containing Rubisco and CcaA with the outer shell (Fig. 20).

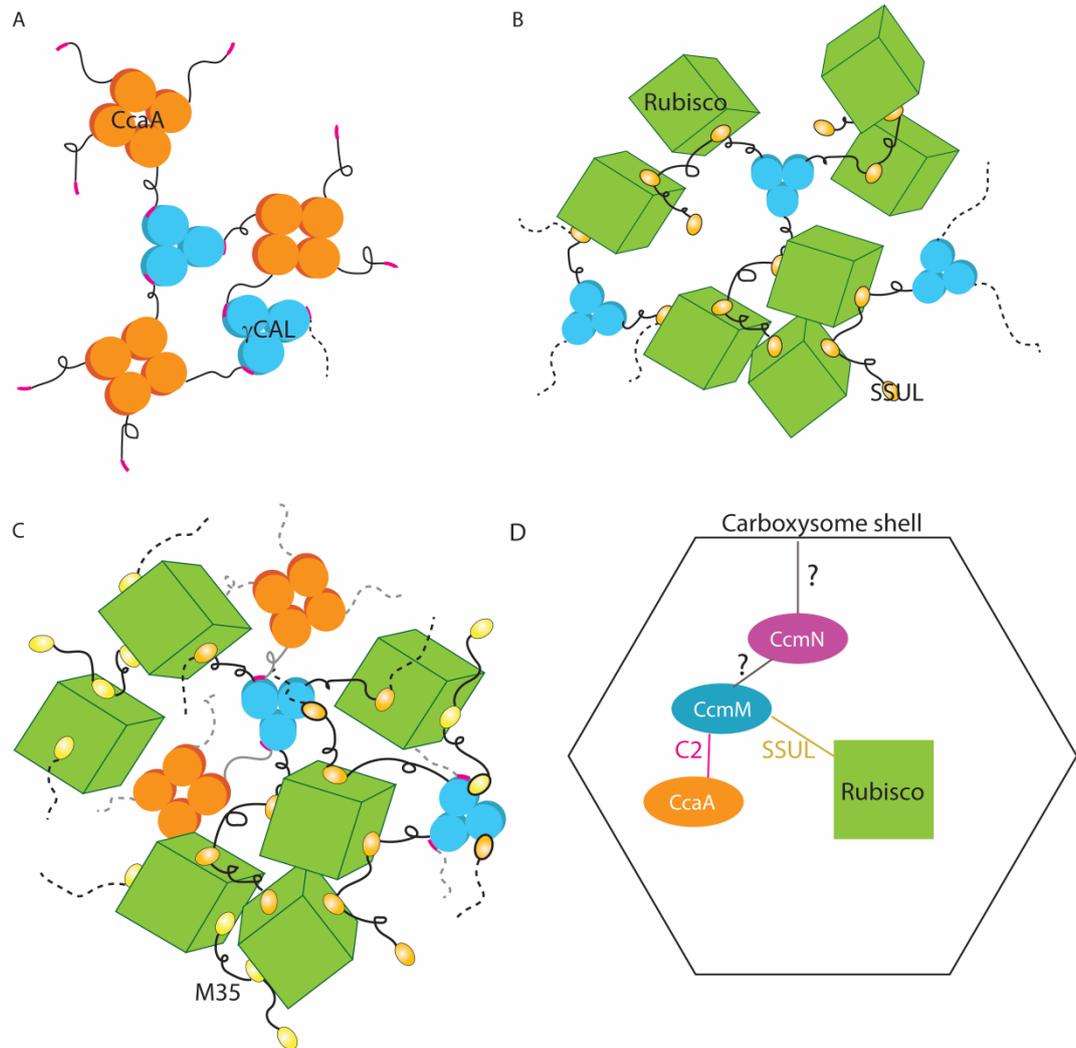


Figure 20: Role of M58 as central organizer of the pre-carboxysome condensate. (A) – (C) Versatility of M58-mediated protein interactions in multi-protein phase separation. Interaction of the trimeric γ CAL domains (blue) of M58 with the C2 sequence of tetrameric CcaA (orange). The C2 peptide is shown in pink (A). Interaction of trimeric M58 (blue) with Rubisco (green) mediated by the SSUL modules of M58 (yellow) present at high local concentration (B). A protein condensate including M35 (light-yellow), M58, Rubisco and CcaA (C). (D) A scheme of potential network of protein interaction in β -carboxysome biogenesis. Figure adapted from Zang et al. (manuscript under review).

M58 interacts with Rubisco at a much higher affinity than M35 due to the high multivalency of SSUL modules in trimeric M58. Surprisingly, we found that the SSUL modules besides interacting with Rubisco also make intermolecular interactions with the γ -CAL domains of neighboring M58 molecules, inducing M58 homo demixing. M58 homo

demixing is enhanced in oxidizing conditions, which might serve as another layer of redox regulation.

Why do cells need both M58 and M35, since M58 alone can induce both Rubisco and CcaA condensate formation? We speculate that the requirement of M35 is due to its sensitivity in response to redox conditions, with higher affinity for Rubisco in the reducing cytosol and lower affinity in the oxidizing environment of the carboxysome. *In vivo* mutagenesis indicated the importance of redox regulation in SSUL for the assembly and function of β -carboxysome (Wang et al., 2019). Consistent with this is the finding that *in vivo* M35 is more abundant than M58 (Long et al., 2007), highlighting the requirement for M35.

5.5 Form IB Rubisco reactivation

A previous study using hydrogen/deuterium exchange (H/DX) and chemical crosslinking mass spectrometry (MS) plus a low resolution cryo-EM map revealed that Rca of the proteobacteria *Rhodobacter sphaeroides* (*Rs*) reactivates *Rs*Rubisco through interaction with the long C-terminal extension of the RbcL subunits (Bhat et al., 2017a). However, green-type form IB RbcL subunits do not comprise such a long sequence at the C-terminus. In this study, we first characterized the Rca-like protein from *Nostoc* sp. PCC 7120 (*NosRca*). *In vitro*, *NosRca* reactivated CABP-inhibited cognate Rubisco. The reactivation only required the AAA+ domain. The physiological inhibitor of *Nos*Rubisco remains to be identified.

NosRca Δ C, the truncation variant containing only the AAA+ domain, provides us with a good opportunity to study the form IB Rubisco reactivation mechanism. The plant Rca forms a range of different oligomeric species and its interaction with cognate Rubisco is very transient (Stotz et al., 2011). Attempts to capture and stabilize a plant Rubisco-Rca complex have so far failed (data not shown). Serendipitously, *NosRca* Δ C is a well-behaved protein

that mainly forms stable hexamers. We solved the *Nos*Rubisco-*Nos*Rca Δ C complex structure in its working state by cryo-EM using a nucleotide substitution strategy and found that this Rca interacted via its central pore with the N-terminal tail of form IB RbcL instead of the C-terminal tail. By comparing the Rubisco substrate binding pocket in the closed conformation and the pocket in open conformation interacting with Rca, we found that the Rca-mediated Rubisco remodeling happens locally, involving four structural elements of RbcL: the 60s loop, loop 6 and both N- and C- terminal tails (Fig. 21). The N-terminal 13-residue peptide binds inside the AAA+ central pore and the C-terminus is pushed away by a protruding α 4- β 4 loop on the Rca *cis* side. Since the stabilization of 60s loop and loop 6 by N- or C-terminal tails, respectively, in the closed conformation, is removed, the two loops move toward solvent and adapt an open conformation, giving enough space for the release of bound sugar phosphate. Next, we asked if such a structural remodeling mechanism is conserved in the reactivation of plant form IB Rubisco. The sequence of the α 4- β 4 loop is conserved between cyanobacterial Rca-like proteins and the Rca of plants. The truncation of the N-terminal dodecapeptide in *N. tabacum* RbcL abolishes the reactivation by its cognate Rca. Independently, this sensitivity to N-terminal truncation has been shown for the activation of the Rubisco from *A. thaliana* (Ng et al., 2020). Thus, the local remodeling of the two terminal peptides serves as a conserved remodeling mechanism of Rubisco by its cognate Rca.

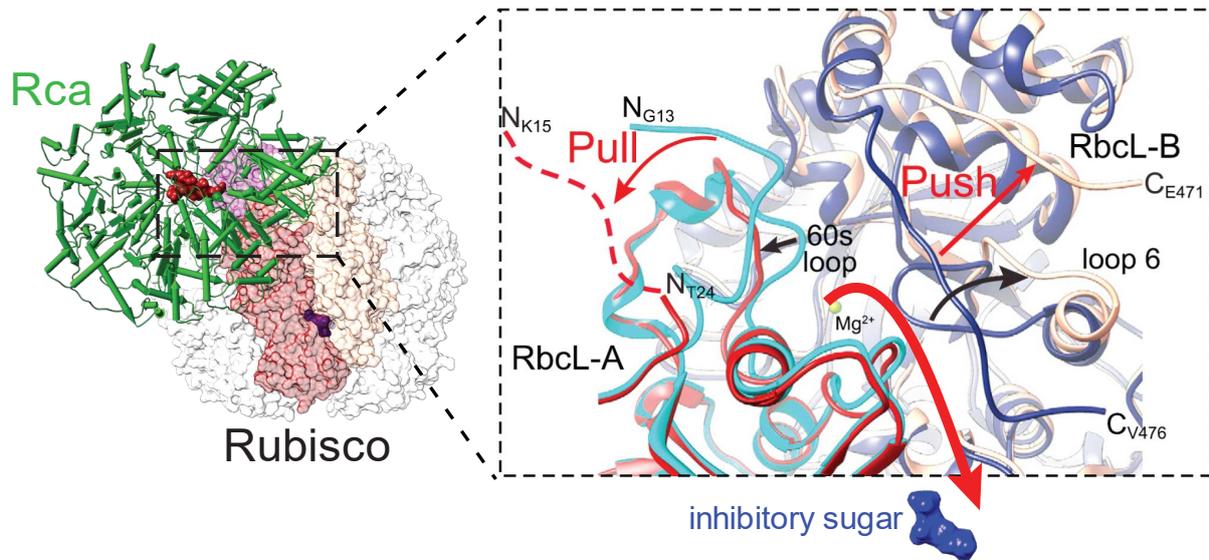


Figure 21: Mechanism of *NosRca* in Rubisco Reactivation. *NosRca* Δ C in complex with the cognate Rubisco (left). Rubisco is shown in surface representation. The antiparallel RbcL dimer and the RbcS subunits involved in the interaction with Rca are colored. The black-dashed region focuses on one Rubisco substrate binding pocket. A zoomed-in image with superposition of the binding pocket, in the closed (RbcL-A, light blue; RbcL-B, dark blue) and the *NosRca*-engaged open state (RbcL-A, red; RbcL-B, peach) is shown on the right. The remodeled regions, including the 60s loop of RbcL-A and loop 6 of RbcL-B as well as the pull and push actions of *NosRca* are indicated. Figure reproduced from (Flecken et al., 2020).

Despite of the local structural remodeling in RbcL, some aspects of plant Rca function remain obscure. The plant Rca proteins contain an additional N-terminal domain (NTD) in comparison to cyanobacterial Rca-like proteins, which is essential for plant Rubisco reactivation. An *in vitro* study suggested that the functional hexamer unit only requires one or two NTDs (Shivhare et al., 2019). How NTD functions in reactivation requires further study. Hopefully, our results on β -carboxysome biogenesis and form IB Rubisco reactivation will help in the ongoing efforts to optimize photosynthesis and biomass production.

6 References

- Ammelburg, M., Frickey, T., and Lupas, A.N. (2006). Classification of AAA+ proteins. *J Struct Biol* 156, 2-11.
- Anderson, L.E. (1971). Chloroplast and cytoplasmic enzymes. II. Pea leaf triose phosphate isomerases. *Biochim Biophys Acta* 235, 237-244.
- Andersson, I., and Backlund, A. (2008). Structure and function of Rubisco. *Plant Physiol Biochem* 46, 275-291.
- Atkinson, N., Velanis, C.N., Wunder, T., Clarke, D.J., Mueller-Cajar, O., and McCormick, A.J. (2019). The pyrenoidal linker protein EPYC1 phase separates with hybrid Arabidopsis–Chlamydomonas Rubisco through interactions with the algal Rubisco small subunit. *Journal of Experimental Botany* 70, 5271-5285.
- Badger, M.R., Andrews, T.J., Whitney, S.M., Ludwig, M., Yellowlees, D.C., Leggat, W., and Price, G.D. (1998). The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO₂-concentrating mechanisms in algae. *Canadian Journal of Botany-Revue Canadienne De Botanique* 76, 1052-1071.
- Badger, M.R., and Bek, E.J. (2008). Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO₂ acquisition by the CBB cycle. *Journal of experimental botany* 59, 1525-1541.
- Badger, M.R., and Price, G.D. (1994). The Role of Carbonic-Anhydrase in Photosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology* 45, 369-392.
- Badger, M.R., and Price, G.D. (2003). CO₂ concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *Journal of Experimental Botany* 54, 609-622.
- Bainbridge, G., Madgwick, P., Parmar, S., Mitchell, R., Paul, M., Pitts, J., Keys, A.J., and Parry, M.A.J. (1995). Engineering Rubisco to Change Its Catalytic Properties. *Journal of Experimental Botany* 46, 1269-1276.
- Bauwe, H. (2010). Recent developments in photorespiration research. *Biochem Soc Trans* 38, 677-682.
- Berry, J.A., Lorimer, G.H., Pierce, J., Seemann, J.R., Meek, J., and Freas, S. (1987). Isolation, identification, and synthesis of 2-carboxyarabinitol 1-phosphate, a diurnal regulator of ribulose-bisphosphate carboxylase activity. *Proc Natl Acad Sci U S A* 84, 734-738.
- Bhat, J.Y., Miličić, G., Thieulin-Pardo, G., Bracher, A., Maxwell, A., Ciniawsky, S., Mueller-Cajar, O., Engen, J.R., Hartl, F.U., Wendler, P., *et al.* (2017a). Mechanism of Enzyme Repair by the AAA+ Chaperone Rubisco Activase. *Molecular Cell* 67, 744-756.e746.
- Bhat, J.Y., Thieulin-Pardo, G., Hartl, F.U., and Hayer-Hartl, M. (2017b). Rubisco Activases: AAA+ Chaperones Adapted to Enzyme Repair. *Front Mol Biosci* 4, 20.
- Bloom, A.J., Burger, M., Rubio Asensio, J.S., and Cousins, A.B. (2010). Carbon dioxide enrichment inhibits nitrate assimilation in wheat and Arabidopsis. *Science* 328, 899-903.
- Botos, I., Melnikov, E.E., Cherry, S., Khalatova, A.G., Rasulova, F.S., Tropea, J.E., Maurizi, M.R., Rotanova, T.V., Gustchina, A., and Wlodawer, A. (2004). Crystal structure of the AAA+ α domain of E. coli Lon protease at 1.9 Å resolution. *Journal of structural biology* 146, 113-122.
- Bowes, G., Ogren, W.L., and Hageman, R.H. (1971). Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochem Biophys Res Commun* 45, 716-722.
- Bracher, A., Whitney, S.M., Hartl, F.U., and Hayer-Hartl, M. (2017). Biogenesis and Metabolic Maintenance of Rubisco. *Annu Rev Plant Biol* 68, 29-60.
- Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoegel, C., Gharakhani, J., Julicher, F., and Hyman, A.A. (2009). Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science* 324, 1729-1732.

Cameron, Jeffrey C., Wilson, Steven C., Bernstein, Susan L., and Kerfeld, Cheryl A. (2013). Biogenesis of a Bacterial Organelle: The Carboxysome Assembly Pathway. *Cell* 155, 1131-1140.

Carmo-Silva, A.E., and Salvucci, M.E. (2013). The Regulatory Properties of Rubisco Activase Differ among Species and Affect Photosynthetic Induction during Light Transitions. *Plant Physiology* 161, 1645.

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. *Journal of Molecular Biology* 428, 153-164.

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

d Moore, B., and Seemann, J.R. (1992). Metabolism of 2'-carboxyarabinitol in leaves. *Plant Physiology* 99, 1551-1555.

Davidi, D., Shamshoum, M., Guo, Z., Bar-On, Y.M., Prywes, N., Oz, A., Jablonska, J., Flamholz, A., Wernick, D.G., Antonovsky, N., *et al.* (2020). Highly active rubiscos discovered by systematic interrogation of natural sequence diversity. *EMBO J* 39, e104081.

Davies, J.M., Brunger, A.T., and Weis, W.I. (2008). Improved structures of full-length p97, an AAA ATPase: implications for mechanisms of nucleotide-dependent conformational change. *Structure* 16, 715-726.

Dean, M., and Annilo, T. (2005). Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. *Annu Rev Genomics Hum Genet* 6, 123-142.

Delwiche, C.F., and Palmer, J.D. (1996). Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. *Mol Biol Evol* 13, 873-882.

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., *et al.* (2019). DABs are inorganic carbon pumps found throughout prokaryotic phyla. *Nature Microbiology* 4, 2204-2215.

Dodd, A.N., Borland, A.M., Haslam, R.P., Griffiths, H., and Maxwell, K. (2002). Crassulacean acid metabolism: plastic, fantastic. *J Exp Bot* 53, 569-580.

Dong, Y., Zhang, S., Wu, Z., Li, X., Wang, W.L., Zhu, Y., Stoilova-McPhie, S., Lu, Y., Finley, D., and Mao, Y. (2019). Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. *Nature* 565, 49-55.

Duysens, L.N.M., Amesz, J., and Kamp, B.M. (1961). Two Photochemical Systems in Photosynthesis. *Nature* 190, 510-511.

Eberhard, S., Finazzi, G., and Wollman, F.A. (2008). The dynamics of photosynthesis. *Annu Rev Genet* 42, 463-515.

Edmondson, D.L., Kane, H.J., and Andrews, T.J. (1990). Substrate Isomerization Inhibits Ribulosebisphosphate Carboxylase-Oxygenase during Catalysis. *Febs Letters* 260, 62-66.

Eisenhut, M., Ruth, W., Haimovich, M., Bauwe, H., Kaplan, A., and Hagemann, M. (2008). The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiotically to plants. *Proceedings of the National Academy of Sciences* 105, 17199.

Ellis, R.J. (1979). The most abundant protein in the world. *Trends in Biochemical Sciences* 4, 241-244.

Erdmann, R., Wiebel, F.F., Flessau, A., Rytka, J., Beyer, A., Frohlich, K.U., and Kunau, W.H. (1991). PAS1, a yeast gene required for peroxisome biogenesis, encodes a member of a novel family of putative ATPases. *Cell* 64, 499-510.

Erzberger, J.P., and Berger, J.M. (2006). Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu Rev Biophys Biomol Struct* 35, 93-114.

Esau, B.D., Snyder, G.W., and Portis, A.R., Jr. (1996). Differential effects of N- and C-terminal deletions on the two activities of rubisco activase. *Arch Biochem Biophys* 326, 100-105.

Espie, G.S., and Kimber, M.S. (2011). Carboxysomes: cyanobacterial RubisCO comes in small packages. *Photosynthesis Research* 109, 7-20.

Fei, X., Bell, T.A., Jenni, S., Stinson, B.M., Baker, T.A., Harrison, S.C., and Sauer, R.T. (2020). Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate. *eLife* 9, e52774.

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.

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.e420.

Freeman Rosenzweig, E.S., Xu, B., Kuhn Cuellar, L., Martinez-Sanchez, A., Schaffer, M., Strauss, M., Cartwright, H.N., Ronceray, P., Pnitzko, J.M., Förster, F., *et al.* (2017). The Eukaryotic CO₂-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization. *Cell* 171, 148-162.e119.

Frolov, E.N., Kublanov, I.V., Toshchakov, S.V., Lunev, E.A., Pimenov, N.V., Bonch-Osmolovskaya, E.A., Lebedinsky, A.V., and Chernyh, N.A. (2019). Form III RubisCO-mediated transaldolase variant of the Calvin cycle in a chemolithoautotrophic bacterium. *Proceedings of the National Academy of Sciences* 116, 18638.

Fukuzawa, H., Suzuki, E., Komukai, Y., and Miyachi, S. (1992). A Gene Homologous to Chloroplast Carbonic-Anhydrase (Icfa) Is Essential to Photosynthetic Carbon-Dioxide Fixation by *Synechococcus Pcc7942*. *Proceedings of the National Academy of Sciences of the United States of America* 89, 4437-4441.

Geiger, D.R., and Servaites, J.C. (1994). Dynamics of Self-Regulation of Photosynthetic Carbon Metabolism. *Plant Physiology and Biochemistry* 32, 173-183.

Gibson, J.L., and Tabita, F.R. (1997). Analysis of the cbbXYZ operon in *Rhodobacter sphaeroides*. *Journal of bacteriology* 179, 663-669.

Gontero, B., and Salvucci, M.E. (2014). Regulation of photosynthetic carbon metabolism in aquatic and terrestrial organisms by Rubisco activase, redox-modulation and CP12. *Aquatic Botany* 118, 14-23.

Gonzales, A.D., Light, Y.K., Zhang, Z.D., Iqbal, T., Lane, T.W., and Martino, A. (2005). Proteomic analysis of the CO₂-concentrating mechanism in the open-ocean cyanobacterium *Synechococcus WH8102*. *Canadian Journal of Botany-Revue Canadienne De Botanique* 83, 735-745.

Gutteridge, S., Parry, M.A.J., Burton, S., Keys, A.J., Mudd, A., Feeney, J., Servaites, J.C., and Pierce, J. (1986). A Nocturnal Inhibitor of Carboxylation in Leaves. *Nature* 324, 274-276.

Hagemann, M., and Bauwe, H. (2016). Photorespiration and the potential to improve photosynthesis. *Curr Opin Chem Biol* 35, 109-116.

Hanson, P.I., and Whiteheart, S.W. (2005). AAA+ proteins: Have engine, will work. *Nature Reviews Molecular Cell Biology* 6, 519-529.

Harpel, M.R., Serspersu, E.H., Lamerdin, J.A., Huang, Z.H., Gage, D.A., and Hartman, F.C. (1995). Oxygenation mechanism of ribulose-bisphosphate carboxylase/oxygenase. Structure and origin of 2-carboxytetritol 1,4-bisphosphate, a novel O₂-dependent side product generated by a site-directed mutant. *Biochemistry* 34, 11296-11306.

Hasse, D., Larsson, A.M., and Andersson, I. (2015). Structure of *Arabidopsis thaliana* Rubisco activase. *Acta Crystallogr D Biol Crystallogr* 71, 800-808.

Hauser, T., Popilka, L., Hartl, F.U., and Hayer-Hartl, M. (2015). Role of auxiliary proteins in Rubisco biogenesis and function. *Nat Plants* 1, 15065.

Hayashi, N.R., Arai, H., Kodama, T., and Igarashi, Y. (1997). The Novel Genes, cbbQandcbbO, Located Downstream from the RubisCO Genes of *Pseudomonas hydrogenothermophila*, Affect the Conformational States and Activity of RubisCO. *Biochemical and biophysical research communications* 241, 565-569.

Hayer-Hartl, M., and Hartl, F.U. (2020). Chaperone Machineries of Rubisco - The Most Abundant Enzyme. *Trends Biochem Sci* 45, 748-763.

Hazra, S., Henderson, J.N., Liles, K., Hilton, M.T., and Wachter, R.M. (2015). Regulation of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase: PRODUCT INHIBITION, COOPERATIVITY, AND MAGNESIUM ACTIVATION*. *Journal of Biological Chemistry* *290*, 24222-24236.

He, S., Chou, H.-T., Matthies, D., Wunder, T., Meyer, M.T., Atkinson, N., Martinez-Sanchez, A., Jeffrey, P.D., Port, S.A., Patena, W., *et al.* (2020). The structural basis of Rubisco phase separation in the pyrenoid. *Nature Plants* *6*, 1480-1490.

Henderson, J.N., Kuriata, A.M., Fromme, R., Salvucci, M.E., and Wachter, R.M. (2011). Atomic resolution x-ray structure of the substrate recognition domain of higher plant ribulose-bisphosphate carboxylase/oxygenase (Rubisco) activase. *J Biol Chem* *286*, 35683-35688.

Hill, R., and Bendall, F.A.Y. (1960). Function of the Two Cytochrome Components in Chloroplasts: A Working Hypothesis. *Nature* *186*, 136-137.

Holbrook, G.P., Bowes, G., and Salvucci, M.E. (1989). Degradation of 2-carboxyarabinitol 1-phosphate by a specific chloroplast phosphatase. *Plant Physiol* *90*, 673-678.

Hovde, B.T., Deodato, C.R., Hunsperger, H.M., Ryken, S.A., Yost, W., Jha, R.K., Patterson, J., Monnat, R.J., Jr., Barlow, S.B., Starkenburg, S.R., *et al.* (2015). Genome Sequence and Transcriptome Analyses of *Chrysochromulina tobin*: Metabolic Tools for Enhanced Algal Fitness in the Prominent Order Prymnesiales (Haptophyceae). *PLoS Genet* *11*, e1005469.

Huang, F., Vasieva, O., Sun, Y., Faulkner, M., Dykes, G.F., Zhao, Z., and Liu, L.-N. (2019). Roles of RbcX in Carboxysome Biosynthesis in the Cyanobacterium *Synechococcus elongatus*; PCC7942. *Plant Physiology* *179*, 184.

Hyman, A.A., Weber, C.A., and Julicher, F. (2014). Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol* *30*, 39-58.

Iyer, L.M., Leipe, D.D., Koonin, E.V., and Aravind, L. (2004). Evolutionary history and higher order classification of AAA+ ATPases. *J Struct Biol* *146*, 11-31.

JAGENDORF, A.T. (1967). Acid-base transition and phosphorylation by chloroplasts. Paper presented at: Fed Proc.

Kane, H.J., Wilkin, J.M., Portis, A.R., and John Andrews, T. (1998). Potent inhibition of ribulose-bisphosphate carboxylase by an oxidized impurity in ribulose-1,5-bisphosphate. *Plant Physiol* *117*, 1059-1069.

Kannappan, B., and Gready, J.E. (2008). Redefinition of Rubisco Carboxylase Reaction Reveals Origin of Water for Hydration and New Roles for Active-Site Residues. *Journal of the American Chemical Society* *130*, 15063-15080.

Karkehabadi, S., Satagopan, S., Taylor, T.C., Spreitzer, R.J., and Andersson, I. (2007). Structural analysis of altered large-subunit loop-6/carboxy-terminus interactions that influence catalytic efficiency and CO₂/O₂ specificity of ribulose-1, 5-bisphosphate carboxylase/oxygenase. *Biochemistry* *46*, 11080-11089.

Karkehabadi, S., Taylor, T.C., and Andersson, I. (2003). Calcium supports loop closure but not catalysis in Rubisco. *Journal of molecular biology* *334*, 65-73.

Kelly, G.J., and Latzko, E. (1976). Inhibition of spinach-leaf phosphofructokinase by 2-phosphoglycollate. *FEBS Lett* *68*, 55-58.

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

Kerfeld, C.A., Sawaya, M.R., Tanaka, S., Nguyen, C.V., Phillips, M., Beeby, M., and Yeates, T.O. (2005). Protein structures forming the shell of primitive bacterial organelles. *Science* *309*, 936-938.

Keys, A.J., Major, I., and Parry, M.A. (1995). Is there another player in the game of Rubisco regulation? *Journal of Experimental Botany*, 1245-1251.

Kim, K., and Portis, A.R., Jr. (2004). Oxygen-dependent H₂O₂ production by Rubisco. *FEBS Lett* *571*, 124-128.

Kingston-Smith, A.H., Major, I., Parry, M.A., and Keys, A.J. (1992). Purification and properties of a phosphatase in French bean (*Phaseolus vulgaris* L.) leaves that hydrolyses 2'-carboxy-d-arabinitol 1-phosphate. *Biochemical Journal* **287**, 821-825.

Kinney, J.N., Axen, S.D., and Kerfeld, C.A. (2011). Comparative analysis of carboxysome shell proteins. *Photosynth Res* **109**, 21-32.

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*. *Journal of Biological Chemistry* **287**, 17729-17736.

Kitano, K., Maeda, N., Fukui, T., Atomi, H., Imanaka, T., and Miki, K. (2001). Crystal structure of a novel-type archaeal rubisco with pentagonal symmetry. *Structure* **9**, 473-481.

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.

Klughhammer, B., Sultemeyer, D., Badger, M.R., and Price, G.D. (1999). The involvement of NAD(P)H dehydrogenase subunits, NdhD3 and NdhF3, in high-affinity CO₂ uptake in *Synechococcus* sp. PCC7002 gives evidence for multiple NDH-1 complexes with specific roles in cyanobacteria. *Mol Microbiol* **32**, 1305-1315.

Kozaki, A., and Takeba, G. (1996). Photorespiration protects C₃ plants from photooxidation. *Nature* **384**, 557-560.

Kunau, W.H., Beyer, A., Franken, T., Gotte, K., Marzioch, M., Saidowsky, J., Skaletz-Rorowski, A., and Wiebel, F.F. (1993). Two complementary approaches to study peroxisome biogenesis in *Saccharomyces cerevisiae*: forward and reversed genetics. *Biochimie* **75**, 209-224.

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 Phytologist* **225**, 793-806.

Lee, S.Y., De La Torre, A., Yan, D., Kustu, S., Nixon, B.T., and Wemmer, D.E. (2003). Regulation of the transcriptional activator NtrC1: structural studies of the regulatory and AAA+ ATPase domains. *Genes Dev* **17**, 2552-2563.

Lenzen, C.U., Steinmann, D., Whiteheart, S.W., and Weis, W.I. (1998). Crystal structure of the hexamerization domain of N-ethylmaleimide-sensitive fusion protein. *Cell* **94**, 525-536.

Li, C., Salvucci, M.E., and Portis, A.R. (2005). Two Residues of Rubisco Activase Involved in Recognition of the Rubisco Substrate*. *Journal of Biological Chemistry* **280**, 24864-24869.

Li, L.A., Zianni, M.R., and Tabita, F.R. (1999). Inactivation of the monocistronic *rca* gene in *Anabaena variabilis* suggests a physiological ribulose bisphosphate carboxylase/oxygenase activase-like function in heterocystous cyanobacteria. *Plant Mol Biol* **40**, 467-478.

Lin, Y., Currie, S.L., and Rosen, M.K. (2017). Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. *Journal of Biological Chemistry* **292**, 19110-19120.

Loganathan, N., Tsai, Y.C., and Mueller-Cajar, O. (2016). Characterization of the heterooligomeric red-type rubisco activase from red algae. *Proc Natl Acad Sci U S A* **113**, 14019-14024.

Long, B.M., Badger, M.R., Whitney, S.M., and Price, G.D. (2007). Analysis of carboxysomes from *Synechococcus* PCC7942 reveals multiple Rubisco complexes with carboxysomal proteins CcmM and CcaA. *J Biol Chem* **282**, 29323-29335.

Long, B.M., Tucker, L., Badger, M.R., and Price, G.D. (2010). Functional Cyanobacterial β -Carboxysomes Have an Absolute Requirement for Both Long and Short Forms of the CcmM Protein. *Plant Physiology* **153**, 285-293.

Lorimer, G., Andrews, T., and Tolbert, N. (1973). Ribulose diphosphate oxygenase. II. Further proof of reaction products and mechanism of action. *Biochemistry* **12**, 18-23.

Lorimer, G.H., Badger, M.R., and Andrews, T.J. (1976). The activation of ribulose-1,5-bisphosphate carboxylase by carbon dioxide and magnesium ions. Equilibria, kinetics, a suggested mechanism, and physiological implications. *Biochemistry* *15*, 529-536.

Lorimer, G.H., and Miziorko, H.M. (1980). Carbamate formation on the epsilon.-amino group of a lysyl residue as the basis for the activation of ribulosebisphosphate carboxylase by carbon dioxide and magnesium (2+). *Biochemistry* *19*, 5321-5328.

Ludwig, M., Sültemeyer, D., and Price, G.D. (2000). Isolation of ccmKLMN genes from the marine cyanobacterium, *Synechococcus* sp. PCC7002 (Cyanophyceae), and evidence that CcmM is essential for carboxysome assembly. *Journal of Phycology* *36*, 1109-1119.

MacCready, J.S., Tran, L., Basalla, J.L., Hakim, P., and Vecchiarelli, A.G. (2021). The McdAB system positions alpha-carboxysomes in proteobacteria. *Mol Microbiol n/a*.

Mackinder, L.C.M., Meyer, M.T., Mettler-Altmann, T., Chen, V.K., Mitchell, M.C., Caspari, O., Freeman Rosenzweig, E.S., Pallesen, L., Reeves, G., Itakura, A., *et al.* (2016). A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. *Proceedings of the National Academy of Sciences* *113*, 5958.

Maeda, N., Kitano, K., Fukui, T., Ezaki, S., Atomi, H., Miki, K., and Imanaka, T. (1999). Ribulose bisphosphate carboxylase/oxygenase from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1 is composed solely of large subunits and forms a pentagonal structure. *Journal of Molecular Biology* *293*, 57-66.

Maeda, S., Badger, M.R., and Price, G.D. (2002). Novel gene products associated with NdhD3/D4-containing NDH-1 complexes are involved in photosynthetic CO₂ hydration in the cyanobacterium, *Synechococcus* sp. PCC7942. *Mol Microbiol* *43*, 425-435.

Maharana, S., Wang, J., Papadopoulos, D.K., Richter, D., Pozniakovsky, A., Poser, I., Bickle, M., Rizk, S., Guillen-Boixet, J., Franzmann, T.M., *et al.* (2018). RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* *360*, 918-921.

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. *Proceedings of the National Academy of Sciences* *113*, E5354.

Mangiapi, M., Usf, M., Brown, T.W., Chaput, D., Haller, E., Harmer, T.L., Hashemy, Z., Keeley, R., Leonard, J., Mancera, P., *et al.* (2017). Proteomic and Mutant Analysis of the CO₂ Concentrating Mechanism of Hydrothermal Vent Chemolithoautotroph *Thiomicrospira crunogena*. *J Bacteriol* *199*.

Mann, C.C. (1999). Genetic engineers aim to soup up crop photosynthesis (American Association for the Advancement of Science).

McCarty, R.E. (1966). Effect of a coupling factor and its antiserum on photophosphorylation and hydrogen ion transport. Paper presented at: Brookhaven symposia in biology.

Meyer, M., and Griffiths, H. (2013). Origins and diversity of eukaryotic CO₂-concentrating mechanisms: lessons for the future. *J Exp Bot* *64*, 769-786.

Meyer, M.T., Genkov, T., Skepper, J.N., Jouhet, J., Mitchell, M.C., Spreitzer, R.J., and Griffiths, H. (2012). Rubisco small-subunit α -helices control pyrenoid formation in *Chlamydomonas*. *Proceedings of the National Academy of Sciences* *109*, 19474-19479.

Mitchell, P. (1961). Coupling of Phosphorylation to Electron and Hydrogen Transfer by a Chemi-Osmotic type of Mechanism. *Nature* *191*, 144-148.

Miziorko, H.M., and Lorimer, G.H. (1983). Ribulose-1,5-bisphosphate carboxylase-oxygenase. *Annu Rev Biochem* *52*, 507-535.

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.

Moore, B.D., Sharkey, T.D., and Seemann, J.R. (1995). Intracellular localization of CA1P and CA1P phosphatase activity in leaves of *Phaseolus vulgaris* L. *Photosynth Res* *45*, 219-224.

Morita, E., Kuroiwa, H., Kuroiwa, T., and Nozaki, H. (1997). High localization of ribulose-1, 5-bisphosphate carboxylase/oxygenase in the pyrenoids of *Chlamydomonas reinhardtii* (chlorophyta), as revealed by cryofixation and immunogold electron microscopy. *Journal of phycology* 33, 68-72.

Moroney, J.V., and Mason, C.B. (1991). The Role of the Chloroplast in Inorganic Carbon Acquisition by *Chlamydomonas-Reinhardtii*. *Canadian Journal of Botany-Revue Canadienne De Botanique* 69, 1017-1024.

Mueller-Cajar, O., Stotz, M., Wendler, P., Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Structure and function of the AAA+ protein CbbX, a red-type Rubisco activase. *Nature* 479, 194-199.

Mueller-Cajar, O., and Whitney, S.M. (2008). Directing the evolution of Rubisco and Rubisco activase: first impressions of a new tool for photosynthesis research. *Photosynth Res* 98, 667-675.

Neuwald, A.F., Aravind, L., Spouge, J.L., and Koonin, E.V. (1999). AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* 9, 27-43.

Newman, J., Branden, C.I., and Jones, T.A. (1993). Structure determination and refinement of ribulose 1,5-bisphosphate carboxylase/oxygenase from *Synechococcus PCC6301*. *Acta Crystallogr D Biol Crystallogr* 49, 548-560.

Ng, J., Guo, Z., and Mueller-Cajar, O. (2020). Rubisco activase requires residues in the large subunit N terminus to remodel inhibited plant Rubisco. *Journal of Biological Chemistry* 295, 16427-16435.

NG, P. (1963). MECHANISM OF THE CARBOXYDISMUTASE REACTION. I. THE EFFECT OF PRELIMINARY INCUBATION OF SUBSTRATES, METAL ION AND ENZYME ON ACTIVITY. *Biochemische Zeitschrift* 338, 7-19.

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. *Nature Structural & Molecular Biology* 27, 281-287.

Omata, T., Price, G.D., Badger, M.R., Okamura, M., Gohta, S., and Ogawa, T. (1999). Identification of an ATP-binding cassette transporter involved in bicarbonate uptake in the cyanobacterium *Synechococcus* sp. strain PCC 7942. *Proc Natl Acad Sci U S A* 96, 13571-13576.

Osmond, B., Badger, M., Maxwell, K., Björkman, O., and Leegood, R. (1997). Too many photons: photorespiration, photoinhibition and photooxidation. *Trends in Plant Science* 2, 119-121.

Osmond, C., and Grace, S. (1995). Perspectives on photoinhibition and photorespiration in the field: quintessential inefficiencies of the light and dark reactions of photosynthesis? *Journal of Experimental Botany*, 1351-1362.

Parry, M.A., Andralojc, P.J., Lowe, H.M., and Keys, A.J. (1999). The localisation of 2-carboxy-D-arabinitol 1-phosphate and inhibition of Rubisco in leaves of *Phaseolus vulgaris* L. *FEBS Lett* 444, 106-110.

Parry, M.A., Keys, A.J., Madgwick, P.J., Carmo-Silva, A.E., and Andralojc, P.J. (2008). Rubisco regulation: a role for inhibitors. *J Exp Bot* 59, 1569-1580.

Pearce, F.G. (2006). Catalytic by-product formation and ligand binding by ribulose bisphosphate carboxylases from different phylogenies. *Biochem J* 399, 525-534.

Pearce, F.G., and Andrews, T.J. (2003). The relationship between side reactions and slow inhibition of ribulose-bisphosphate carboxylase revealed by a loop 6 mutant of the tobacco enzyme. *Journal of Biological Chemistry* 278, 32526-32536.

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.

Portis, A.R. (2003). Rubisco activase—Rubisco's catalytic chaperone. *Photosynthesis research* 75, 11-27.

Portis Jr, A.R., Li, C., Wang, D., and Salvucci, M.E. (2008). Regulation of Rubisco activase and its interaction with Rubisco. *Journal of experimental botany* 59, 1597-1604.

- Portis Jr, A.R., and Salvucci, M.E. (2002). The discovery of Rubisco activase—yet another story of serendipity. *Photosynthesis research* 73, 257-264.
- Price, G.D. (2011). Inorganic carbon transporters of the cyanobacterial CO₂ concentrating mechanism. *Photosynthesis Research* 109, 47-57.
- Price, G.D., Sültemeyer, D., Klughammer, B., Ludwig, M., and Badger, M.R. (1998). The functioning of the CO₂ concentrating mechanism in several cyanobacterial strains: a review of general physiological characteristics, genes, proteins, and recent advances. *Canadian Journal of Botany* 76, 973-1002.
- Price, G.D., Woodger, F.J., Badger, M.R., Howitt, S.M., and Tucker, L. (2004). Identification of a SulP-type bicarbonate transporter in marine cyanobacteria. *Proc Natl Acad Sci U S A* 101, 18228-18233.
- Prommeenate, P., Lennon, A.M., Markert, C., Hippler, M., and Nixon, P.J. (2004). Subunit composition of NDH-1 complexes of *Synechocystis* sp. PCC 6803: identification of two new ndh gene products with nuclear-encoded homologues in the chloroplast Ndh complex. *J Biol Chem* 279, 28165-28173.
- 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. *Microbiology and Molecular Biology Reviews* 77, 357-379.
- Raven, J.A. (2010). Inorganic carbon acquisition by eukaryotic algae: four current questions. *Photosynth Res* 106, 123-134.
- Robinson, S.P., and Portis, A.R. (1988). Release of the nocturnal inhibitor, carboxyarabinitol-1-phosphate, from ribulose biphosphate carboxylase/oxygenase by rubisco activase. *FEBS Letters* 233, 413-416.
- Robinson, S.P., and Portis, A.R. (1989a). Ribulose-1,5-bisphosphate carboxylase/oxygenase activase protein prevents the in vitro decline in activity of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Plant Physiol* 90, 968-971.
- Robinson, S.P., and Portis, A.R., Jr. (1989b). Adenosine triphosphate hydrolysis by purified rubisco activase. *Arch Biochem Biophys* 268, 93-99.
- Roll-Mecak, A., and Vale, R.D. (2008). Structural basis of microtubule severing by the hereditary spastic paraplegia protein spastin. *Nature* 451, 363-367.
- Ryan, P., Forrester, T.J., Wroblewski, C., Kenney, T.M., 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. *Journal of Biological Chemistry* 294, 2593-5195.
- Salvucci, M.E., and Anderson, J.C. (1987). Factors affecting the activation state and the level of total activity of ribulose biphosphate carboxylase in tobacco protoplasts. *Plant Physiol* 85, 66-71.
- Salvucci, M.E., Osteryoung, K.W., Crafts-Brandner, S.J., and Vierling, E. (2001). Exceptional sensitivity of Rubisco activase to thermal denaturation in vitro and in vivo. *Plant Physiol* 127, 1053-1064.
- Salvucci, M.E., Portis, A.R., Jr., and Ogren, W.L. (1985). A soluble chloroplast protein catalyzes ribulosebiphosphate carboxylase/oxygenase activation in vivo. *Photosynth Res* 7, 193-201.
- Satagopan, S., Chan, S., Perry, L.J., and Tabita, F.R. (2014). Structure-Function Studies with the Unique Hexameric Form II Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) from *Rhodospseudomonas palustris**. *Journal of Biological Chemistry* 289, 21433-21450.
- 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.
- Schneider, G., Lindqvist, Y., Brändén, C.I., and Lorimer, G. (1986). Three-dimensional structure of ribulose-1, 5-bisphosphate carboxylase/oxygenase from *Rhodospirillum rubrum* at 2.9 Å resolution. *The EMBO journal* 5, 3409-3415.
- Seemann, J.R., Berry, J.A., Freas, S.M., and Krump, M.A. (1985). Regulation of ribulose biphosphate carboxylase activity in vivo by a light-modulated inhibitor of catalysis. *Proc Natl Acad Sci U S A* 82, 8024-8028.
- Sharkey, T.D., Savitch, L.V., and Butz, N.D. (1991). Photometric method for routine determination of k_{cat} and carbamylation of rubisco. *Photosynthesis Research* 28, 41-48.

Sharwood, R.E. (2017). Engineering chloroplasts to improve Rubisco catalysis: prospects for translating improvements into food and fiber crops. *New Phytologist* *213*, 494-510.

Shen, J.B., and Ogren, W.L. (1992). Alteration of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase activities by site-directed mutagenesis. *Plant Physiol* *99*, 1201-1207.

Shibata, M., Ohkawa, H., Kaneko, T., Fukuzawa, H., Tabata, S., Kaplan, A., and Ogawa, T. (2001). Distinct constitutive and low-CO₂-induced CO₂ uptake systems in cyanobacteria: Genes involved and their phylogenetic relationship with homologous genes in other organisms. *Proceedings of the National Academy of Sciences of the United States of America* *98*, 11789-11794.

Shively, J., and English, R. (1991). The carboxysome, a prokaryotic organelle: a mini-review. *Canadian journal of botany* *69*, 957-962.

Shivhare, D., Ng, J., Tsai, Y.-C.C., and Mueller-Cajar, O. (2019). Probing the rice Rubisco–Rubisco activase interaction via subunit heterooligomerization. *Proceedings of the National Academy of Sciences* *116*, 24041.

Simonetta, K.R., Kazmirski, S.L., Goedken, E.R., Cantor, A.J., Kelch, B.A., McNally, R., Seyedin, S.N., Makino, D.L., O'Donnell, M., and Kuriyan, J. (2009). The mechanism of ATP-dependent primer-template recognition by a clamp loader complex. *Cell* *137*, 659-671.

Singleton, M.R., and Wigley, D.B. (2002). Modularity and specialization in superfamily 1 and 2 helicases. *J Bacteriol* *184*, 1819-1826.

Snider, J., and Houry, W.A. (2006). MoxR AAA+ ATPases: a novel family of molecular chaperones? *J Struct Biol* *156*, 200-209.

So, A.K., John-McKay, M., and Espie, G.S. (2002). Characterization of a mutant lacking carboxysomal carbonic anhydrase from the cyanobacterium *Synechocystis* PCC6803. *Planta* *214*, 456-467.

So, A.K.C., Espie, G.S., Williams, E.B., Shively, J.M., Heinhorst, S., and Cannon, G.C. (2004). A novel evolutionary lineage of carbonic anhydrase (epsilon class) is a component of the carboxysome shell. *Journal of Bacteriology* *186*, 623-630.

Somerville, C., Portis, A.R., and Ogren, W.L. (1982). A mutant of *Arabidopsis thaliana* which lacks activation of RuBP carboxylase in vivo. *Plant Physiology* *70*, 381-387.

Spreitzer, R.J. (2003). Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Arch Biochem Biophys* *414*, 141-149.

Stotz, M., Mueller-Cajar, O., Ciniawsky, S., Wendler, P., Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Structure of green-type Rubisco activase from tobacco. *Nat Struct Mol Biol* *18*, 1366-1370.

Strome, S., and Wood, W.B. (1983). Generation of asymmetry and segregation of germ-line granules in early *C. elegans* embryos. *Cell* *35*, 15-25.

Sutter, M., Roberts, E.W., Gonzalez, R.C., Bates, C., Dawoud, S., Landry, K., Cannon, G.C., Heinhorst, S., and Kerfeld, C.A. (2015a). Structural Characterization of a Newly Identified Component of alpha-Carboxysomes: The AAA+ Domain Protein CsoCbbQ. *Sci Rep* *5*, 16243.

Sutter, M., Roberts, E.W., Gonzalez, R.C., Bates, C., Dawoud, S., Landry, K., Cannon, G.C., Heinhorst, S., and Kerfeld, C.A. (2015b). Structural Characterization of a Newly Identified Component of α -Carboxysomes: The AAA+ Domain Protein CsoCbbQ. *Scientific Reports* *5*.

Tabita, F.R. (1999). Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective. *Photosynthesis Research* *60*, 1-28.

Tabita, F.R., Hanson, T.E., Li, H., Satagopan, S., Singh, J., and Chan, S. (2007). Function, Structure, and Evolution of the RubisCO-Like Proteins and Their RubisCO Homologs. *Microbiology and Molecular Biology Reviews* *71*, 576.

Tabita, F.R., Satagopan, S., Hanson, T.E., Kreel, N.E., and Scott, S.S. (2008). Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *Journal of Experimental Botany* *59*, 1515-1524.

Taler, D., Galperin, M., Benjamin, I., Cohen, Y., and Kenigsbuch, D. (2004). Plant *er* genes that encode photorespiratory enzymes confer resistance against disease. *Plant Cell* *16*, 172-184.

Tanaka, S., Kerfeld, C.A., Sawaya, M.R., Cai, F., Heinhorst, S., Cannon, G.C., and Yeates, T.O. (2008a). Atomic-Level Models of the Bacterial Carboxysome Shell. *Science* *319*, 1083.

Tanaka, S., Kerfeld, C.A., Sawaya, M.R., Cai, F., Heinhorst, S., Cannon, G.C., and Yeates, T.O. (2008b). Atomic-level models of the bacterial carboxysome shell. *Science* *319*, 1083-1086.

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.

Taylor, T.C., and Andersson, I. (1997). Structure of a product complex of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase. *Biochemistry* *36*, 4041-4046.

Toyokawa, C., Yamano, T., and Fukuzawa, H. (2020). Pyrenoid Starch Sheath Is Required for LCIB Localization and the CO₂-Concentrating Mechanism in Green Algae. *Plant Physiol* *182*, 1883-1893.

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.

Tsai, Y., Sawaya, M.R., and Yeates, T.O. (2009). Analysis of lattice-translocation disorder in the layered hexagonal structure of carboxysome shell protein CsoS1C. *Acta Crystallogr D Biol Crystallogr* *65*, 980-988.

Tsai, Y.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.

Tsai, Y.C.C., Ye, F.Z., 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 plus ATPase. *Proceedings of the National Academy of Sciences of the United States of America* *117*, 381-387.

Twomey, E.C., Ji, Z., Wales, T.E., Bodnar, N.O., Ficarro, S.B., Marto, J.A., Engen, J.R., and Rapoport, T.A. (2019). Substrate processing by the Cdc48 ATPase complex is initiated by ubiquitin unfolding. *Science* *365*, eaax1033.

Uversky, V.N. (2017). Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Current opinion in structural biology* *44*, 18-30.

van de Loo, F.J., and Salvucci, M.E. (1996). Activation of ribulose-1, 5-bisphosphate carboxylase/oxygenase (Rubisco) involves Rubisco activase Trp16. *Biochemistry* *35*, 8143-8148.

van Lun, M., van der Spoel, D., and Andersson, I. (2011). Subunit interface dynamics in hexadecameric rubisco. *J Mol Biol* *411*, 1083-1098.

Wachter, R.M., Salvucci, M.E., Carmo-Silva, A.E., Barta, C., Genkov, T., and Spreitzer, R.J. (2013). Activation of interspecies-hybrid Rubisco enzymes to assess different models for the Rubisco–Rubisco activase interaction. *Photosynthesis research* *117*, 557-566.

Wang, A., Conicella, A.E., Schmidt, H.B., Martin, E.W., Rhoads, S.N., Reeb, A.N., Nourse, A., Ramirez Montero, D., Ryan, V.H., Rohatgi, R., *et al.* (2018a). A single N-terminal phosphomimic disrupts TDP-43 polymerization, phase separation, and RNA splicing. *EMBO J* *37*, e97452.

Wang, D., and Portis, A.R. (2006). Increased Sensitivity of Oxidized Large Isoform of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase to ADP Inhibition Is Due to an Interaction between Its Carboxyl Extension and Nucleotide-binding Pocket*. *Journal of Biological Chemistry* *281*, 25241-25249.

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.

Wang, J., Choi, J.-M., Holehouse, A.S., Lee, H.O., Zhang, X., Jahnke, M., Maharana, S., Lemaitre, R., Pozniakovskiy, A., Drechsel, D., *et al.* (2018b). A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* *174*, 688-699.e616.

Wendler, P., Ciniawsky, S., Kock, M., and Kube, S. (2012). Structure and function of the AAA+ nucleotide binding pocket. *Biochim Biophys Acta* *1823*, 2-14.

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*. *Journal of Biological Chemistry* 289, 7973-7981.

Whittaker, C.A., and Hynes, R.O. (2002). Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol Biol Cell* 13, 3369-3387.

Wilson, R.H., and Hayer-Hartl, M. (2018). Complex Chaperone Dependence of Rubisco Biogenesis. *Biochemistry* 57, 3210-3216.

Wolf, N., Priess, J., and Hirsh, D. (1983). Segregation of germline granules in early embryos of *Caenorhabditis elegans*: an electron microscopic analysis. *J Embryol Exp Morphol* 73, 297-306.

Wu, J., Neimanis, S., and Heber, U. (1991). Photorespiration Is More Effective Than the Mehler Reaction in Protecting the Photosynthetic Apparatus against Photoinhibition. *Botanica Acta* 104, 283-291.

Wunder, T., Cheng, S.L.H., Lai, S.-K., Li, H.-Y., and Mueller-Cajar, O. (2018). The phase separation underlying the pyrenoid-based microalgal Rubisco supercharger. *Nature Communications* 9, 5076.

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.

Yeates, T.O., Thompson, M.C., and Bobik, T.A. (2011). The protein shells of bacterial microcompartment organelles. *Current opinion in structural biology* 21, 223-231.

Yu, J.W., Price, G.D., Song, L., and Badger, M.R. (1992). Isolation of a Putative Carboxysomal Carbonic Anhydrase Gene from the Cyanobacterium *Synechococcus* PCC7942. *Plant Physiol* 100, 794-800.

Zhang, N., and Portis, A.R., Jr. (1999). Mechanism of light regulation of Rubisco: a specific role for the larger Rubisco activase isoform involving reductive activation by thioredoxin-f. *Proc Natl Acad Sci U S A* 96, 9438-9443.

Zhang, N., Schurmann, P., and Portis, A.R., Jr. (2001). Characterization of the regulatory function of the 46-kDa isoform of Rubisco activase from *Arabidopsis*. *Photosynth Res* 68, 29-37.

Zhang, X.D., and Wigley, D.B. (2008). The 'glutamate switch' provides a link between ATPase activity and ligand binding in AAA plus proteins. *Nature Structural & Molecular Biology* 15, 1223-1227.

Zhou, Y., and Whitney, S. (2019). Directed Evolution of an Improved Rubisco; In Vitro Analyses to Decipher Fact from Fiction. *Int J Mol Sci* 20.

Zhu, G., and Jensen, R.G. (1991). Xylulose 1,5-Bisphosphate Synthesized by Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase during Catalysis Binds to Decarbamylated Enzyme. *Plant Physiol* 97, 1348-1353.