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

Mechanism of pro-β-carboxysome maturation by the CM-ApN hetero-complex

Xiaoyu Hong
aus
Xinyang, Henan, China
2026

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, 01.01.2026

Xiaoyu Hong

Dissertation eingereicht am: 20.01.2026

1. Gutachter: Prof. Dr. F. Ulrich Hartl

2. Gutachter: PD Dr. Bettina Bölter

Mündliche Prüfung am: 26.03.2026

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors, Dr. Manajit Hayer-Hartl and Prof. Dr. Franz-Ulrich Hartl, for giving me the opportunity to join their harmonious and collaborative laboratory in the Department of Cellular Biochemistry at the Max Planck Institute of Biochemistry (MPIB) and to learn how to do research. Over the past few years, I have been deeply moved and inspired by their profound thinking, creative ideas and honest approach to science. Their mentorship has not only shaped my research skills, but will also guide me to approach my career with a critical and thoughtful mindset.

I would also like to express my sincere gratitude to my collaborator, Dr. Kun Zang, a postdoctoral fellow in the department, for his valuable contributions to this project. Each discussion with him broadened my understanding of the project and refreshed my perspective. Additionally, I would like to thank Dr. Wei Xue, Dr. Alonso Izzat Carvajal Alvarez, and Zsölt Semperger for creating a friendly and supportive working atmosphere in the Rubisco research group.

I would also like to thank the technical staff of the department, Darija Pompino, Emmanuel Burghardt, Albert Ries, Massimo Bossi, Silvia Gärtner, Romy Lange and Nadine Wischnewski for their support and assistance in maintaining the various laboratories in the department.

I am also very grateful to Dr. Daniel Bollschweiler and Dr. Tillman Schäfer from the MPIB cryo-EM facility for their professional expertise and guidance. Their support and advice are greatly appreciated.

Special thanks go to Dr. Manajit Hayer-Hartl, Prof. Dr. Franz-Ulrich Hartl, Dr. Kun Zang, Dr. Andreas Bracher, and the members of my PhD committee: PD Dr. Bettina Bölter, Prof. Dr. Johannes Stigler, Prof. Dr. Jörg Nickelsen, Prof. Dr. Roland Beckmann, and Prof. Dr. Karl-Peter Hopfner. They provided insightful feedback and encouragement throughout my PhD studies.

Finally, I am eternally grateful to my family for their unwavering support and encouragement over the years. I would also like to thank my friends Dr. Jiale Du and Shuangshuang Niu for the happy time we spent together in Munich. Last but not least, I want to acknowledge myself, for being ordinary but brave and courageous enough to support myself through all the difficulties and dilemmas.

I believe that the people I met will be remembered, the memories we made will be cherished, and the knowledge I gained will lead me to a bright life and a fulfilling future.

Summary

Photosynthesis is a fundamental process utilized by plants, algae, and cyanobacteria that converts solar energy into chemical energy, which directly or indirectly fuels most living organisms on Earth. A key enzyme in this process is ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), which catalyzes the first major step of the Calvin-Benson-Bassham (CBB) cycle, mediating the fixation of inorganic carbon dioxide (CO_2) into organic carbon compounds. However, Rubisco has a low catalytic rate and is prone to side reactions with oxygen (O_2), leading to energetically wasteful photorespiration. To overcome this inefficiency, cyanobacteria have evolved a carboxysome-based CO_2-concentrating mechanism (CCM). Carboxysomes are bacterial microcompartments (BMCs), found in many chemoautotrophic proteobacteria and all cyanobacteria. They sequester the enzymes Rubisco and carbonic anhydrase (CA) surrounded by a selectively permeable proteinaceous shell that regulates the entry and exit of various substrates. The role of carboxysomes is to enhance the efficiency of photosynthesis by elevating the concentration of CO_2 in the vicinity of Rubisco. Recent structural and biochemical studies have contributed to our present understanding on the mechanism of β-carboxysome formation by the scaffolding protein CM, which recruits the enzymes Rubisco and CA into a biomolecular condensate – the so-called pro-β-carboxysome. However, the assembly of the pro-β-carboxysome prior to shell formation remains poorly defined.

In this study, we investigated the role of the shell adaptor protein ApN (also known as CcmN) which contains the so-called encapsulation peptide (EP) at its C-terminus. The EP interacts with several shell proteins and is required for carboxysome maturation. We succeeded in recombinantly purifying the ApN protein and showed that it is tetrameric in solution, as confirmed by size-exclusion chromatography coupled to multiangle static light scattering (SEC-MALS) analysis and cryo-electron microscopy (cryo-EM). Pull-down and sedimentation assays, and fluorescence microscopy-based condensate formation analysis indicated that $(ApN)_4$ does not interact with the pro-carboxysome core proteins, including Rubisco, CA and the scaffolding protein CM (also known as CcmM or M58). This suggested that ApN must be recruited to the pro-carboxysome biomolecular condensate via another mechanism. Interestingly, recombinant expression of ApN and the scaffolding protein CM in a bicistronic plasmid resulted in the formation of a hetero-tetrameric complex, consisting of three protomers of ApN and one protomer of CM, $(ApN)_3$:CM. Next, we conducted an *in vivo* mutational

analysis (in collaboration with Benedict M. Long, New South Wales and G. Dean Price, Australian Capital Territory), in which we generated a *ccmK2LMNO* knockout strain (ΔK2-O) in the cyanobacterium *Synechococcus elongatus* PCC 7942 (*Se*7942), and complemented it with the pSE4*ccmK2-O* plasmid carrying either wild-type ApN or ApN mutants. After analyzing the ultrastructures and growth rates of these strains, we found that only those containing a hetero-complex of CM and ApN could form functional carboxysomes and grow under atmospheric CO_2 concentration. These results confirmed that the formation of a hetero-complex between CM and ApN is essential for shell protein recruitment and carboxysome maturation. Fluorescence microscopy analysis revealed that the $(ApN)_3$:CM hetero-complex undergoes multiple, dynamic protein-protein interactions, including the interaction of C-terminal Rubisco small subunit-like (SSUL) modules of CM with the N-terminal γ-carbonic anhydrase-like (γCAL) domain of adjacent CM protomers, the interaction of SSUL modules with Rubisco, and the interaction of the C-terminal tail of $(CA)_4$ with the γCAL domain of CM. These interactions allow the $(ApN)_3$:CM hetero-complex to be reconstituted to the periphery of the pro-carboxysome condensate, forming a pro-β-carboxysome that is ready to recruit shell proteins and form the mature β-carboxysome. Our findings provide new insights into the biogenesis of these microcompartments, with implications for engineering synthetic CO_2-fixation systems into crop plants.

Table of contents

Glossary

Abbreviation	Description
ACI	*Acidobacterium* microcompartment
ADP	adenosine diphosphate
Å	angstrom
AlcDH	alcohol dehydrogenase
AldDH	aldehyde dehydrogenase
ApN	adaptor protein (also known as CcmN)
APS	ammonium persulfate
A. thaliana; *At*	*Arabidopsis thaliana*
ATP	adenosine triphosphate
ATPase	adenosine triphosphatase
BMC	bacterial microcompartment
BMC-H	BMC-hexameric
BMC-P	BMC-pentameric
BMC-T	BMC-trimeric
BMC-T^{ds}	double BMC-T
BMC-T^{s}	single BMC-T
1,3-BPG	1,3-Bisphosphoglycerate
BSA	bovine serum albumin
BSD2	bundle sheath defective-2
CA/CsoSCA	carbonic anhydrase
CAH3	carbonic anhydrase 3
γCAL	γ-carbonic anhydrase-like
CAM	crassulacean acid metabolism
CBB	Calvin-Benson-Bassham
CCM	CO_2-concentrating mechanism
C. elegans	*Caenorhabditis elegans*
CHAPSO	3-(3-Cholamidopropyl-dimethylammonio)-2-hydroxy-1-propanesulfonate
Chl *a*	chlorophyll *a*

C. reinhardtii; *Cr*	*Chlamydomonas reinhardtii*
Chl *b*	chlorophyll *b*
Chl *d*	chlorophyll *d*
Chl *f*	chlorophyll *f*
Ci	inorganic carbon
CM	scaffolding protein CcmM
CMC	critical micelle concentration
CO_2	carbon dioxide
CoA	coenzyme A
Cryo-EM	cryo-electron microscopy
Cryo-ET	cryo-electron tomography
C-terminal tail	C2
*Cy*7001	*Cyanobium* sp. PCC 7001
*Cy*CsoSCA	CsoSCA from *Cy*7001
Cyt b_6f	cytochrome b_6f
2D	two-dimensional
ddH_2O	double-distilled water
DTT	1,4-Dithiothreitol
E	eluate
e^-	electrons
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EFI-EST	Enzyme Function Initiative-Enzyme Similarity Tool
EFI-GNT	Enzyme Function Initiative-Genome Neighborhood Tool
EGFP	enhanced green fluorescent protein
EMD	Electron Microscopy Data Bank
EP	encapsulation peptide
EPYC1	Essential Pyrenoid Component 1
ETC	electron transport chain
F-ATPase	ATP synthase

Fd	ferredoxin
Fe^{2+}	ferrous iron
FNR	Ferredoxin-$NADP^+$ reductase
Formalin	formaldehyde
FR	far-red light
FRAP	fluorescence recovery after photobleaching
GFP	green fluorescent protein
GNNs	genome neighborhood networks
G3P	Glyceraldehyde-3-phosphate
GST	Glutathione S-transferase
h	hour(s)
H^+	protons
H_{10}	10 × His-tagged
*Ha*7418	*Halothece* sp. PCC 7418
HCl	hydrochloric acid
HCO_3^-	bicarbonate
HCR	high CO_2-requiring
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
His6ubi	His6-ubiquitin
His6-Usp2	His-tagged deubiquitylating enzyme Usp2
H. neapolitanus	*Halothiobacillus neapolitanus*
H. ochraceum	*Haliangium ochraceum*
H_2O	water
H_2O_2	hydrogen peroxide
H_2S	hydrogen sulfide
Hsp	heat shock protein
IDPs	intrinsically disordered proteins
IDRs	intrinsically disordered regions
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ipTM	interface predicted template modeling
ITC	isothermal titration calorimetry

3-KCABP or KCABP	3-keto-2-carboxyarabinitol-1,5-bisphosphate
KCl	potassium chloride
keV	kiloelectron volt
ΔK2-O	*ΔccmKLMNO*
KOH	potassium hydroxide
LLPS	liquid-liquid phase separation
MBP	maltose-binding protein
β-ME	β-mercaptoethanol
MES	2-(N-morpholino)ethanesulfonic acid
$Mg(OAc)_2$	magnesium acetate
min	minute(s)
MITH1	MIssing THylakoids 1
ms	milliseconds
NaCl	sodium chloride
$NADP^+$	nicotinamide adenine dinucleotide phosphate
NADPH	nicotinamide adenine dinucleotide phosphate
$Na_2S_2O_3$	sodium thiosulfate
Ni-NTA	nickel-nitrilotriacetic acid
N. tabacum	*Nicotiana tabacum*
OAA	oxaloacetate
OD	optical density
OG	octyl-β-D-glucopyranoside
OsO_4	osmium tetroxide
P	pellet
PAE	predicted aligned error
PAR	photosynthetically active radiation
PBS	Phosphate Buffered Saline
Pc	plastocyanin
PCD	pterin-4α-carbinolamine dehydratase
PCR	polymerase chain reaction
PDB	Protein Data Bank
PDBP	2,3-pentodiulose-1,5-bisphosphate

PDE	predicted distance error
2-PG	2-phosphoglycolate
3-PG	glycerate-3-phosphate or 3-phosphoglycerate
P granules	Processing granules
P_i	inorganic phosphate
pLDDT	predicted local distance difference test
pmf	proton motive force
polylysine	Poly L-lysine hydrobromide
PQ	plastoquinone
PQH_2	plastoquinol
PSI	Photosystem I
PSII	Photosystem II
PTAC	phosphotransacylase
pTM	predicted template modeling
Raf1 and Raf2	Rubisco accumulation factors
RbcL/CbbL	Rubisco large subunit
$RbcL_2$	RbcL dimers
$RbcL_8S_8$	hexadecameric Rubisco (holoenzyme)
RbcS/CbbS	Rubisco small subunit
RBS	ribosome binding site
Rca	Rubisco activase
RLPs	Rubisco-like proteins
R-P	phosphorylated product
Rubisco	ribulose-1,5-bisphosphate carboxylase/oxygenase
RBMs	Rubisco binding motifs
RT	room temperature
RuBP	ribulose-1,5-bisphosphate
s	seconds
S	supernatant
SII	StrepII-tagged
SAGA1 and SAGA2	StArch Granules Abnormal proteins
SDS	sodium dodecyl sulfate

SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
*Se*7942	*Synechococcus elongatus* PCC 7942
SEC	size-exclusion chromatography
SEC-MALS	size-exclusion chromatography coupled to multiangle static light scattering
SSN	sequence similarity network
SSUL	small subunit-like
STA	subtomogram averaging
*Syn*6803	*Synechocystis* sp. PCC 6803
T	total
TAE	Tris-acetate-EDTA
TCEP-HCl	tris(2-carboxyethyl)phosphine-HCl
TEM	transmission electron microscopy
TEMED	N,N,N’,N’-tetramethylethylenediamine
TIM	triosephosphate isomerase
Tris base	2-amino-2-(hydroxymethyl)-1,3-propanediol
UV	ultraviolet
XuBP	xylulose-1,5-bisphosphate

1. Introduction

1.1. Photosynthesis

Photosynthesis is an essential biological process on Earth, converting solar energy into chemical energy and biomass, which forms the foundation of life (Bryant & Frigaard, 2006). It exists in two principal types, (1) oxygenic photosynthesis, employed by most plants, algae, and cyanobacteria, which utilizes chlorophyll *a* (Chl *a*) to split water (H_2O), producing O_2 and fixing CO_2 into organic compounds; and (2) anoxygenic photosynthesis, carried out by certain bacteria that use alternative electron donors such as hydrogen sulfide (H_2S), organic molecules or ferrous iron (Fe^{2+}) instead of H_2O. For example, purple sulfur bacteria use bacteriochlorophyll to split hydrogen sulfide to produce sulfur compounds. Unlike oxygenic photosynthesis, anoxygenic photosynthesis may or may not involve carbon fixation.

1.1.1. Two parts of oxygenic photosynthesis

Oxygenic photosynthesis is the most common type of photosynthesis. In photoautotrophs, such as plants, algae, and cyanobacteria, oxygenic photosynthesis occurs in two main phases, light-dependent and light-independent (Figure **1.1**). The light-dependent reactions take place in the thylakoid membranes of chloroplasts (Figure **1.1A**) in plants and algae, and in plasma membrane invaginations (chromatophores) of bacteria (Hsin et al., 2010), where the specialized photosynthetic pigments – primarily chlorophyll *a* and *b* (Chl *a*, with absorption maxima at ~430 nm/662 nm and Chl *b*, ~453 nm/642 nm, respectively) and accessory pigments such as carotenoids (400-500 nm) – capture photons and extend the range of usable light beyond the spectrum of photosynthetically active radiation (PAR, 400-700 nm) (Croce & van Amerongen, 2014; Falcioni et al., 2023). These pigments are embedded in the photosystem II (PSII, P680) and photosystem I (PSI, P700) complexes, which absorb light optimally at 680 nm and 700 nm, respectively (Duysens et al., 1961). Some cyanobacteria, however, have evolved the ability to synthesize two additional types of chlorophyll, called Chl *d* and Chl *f*, which can absorb far-red light (FR, 700-750 nm) (Gan et al., 2014; Ho et al., 2017).

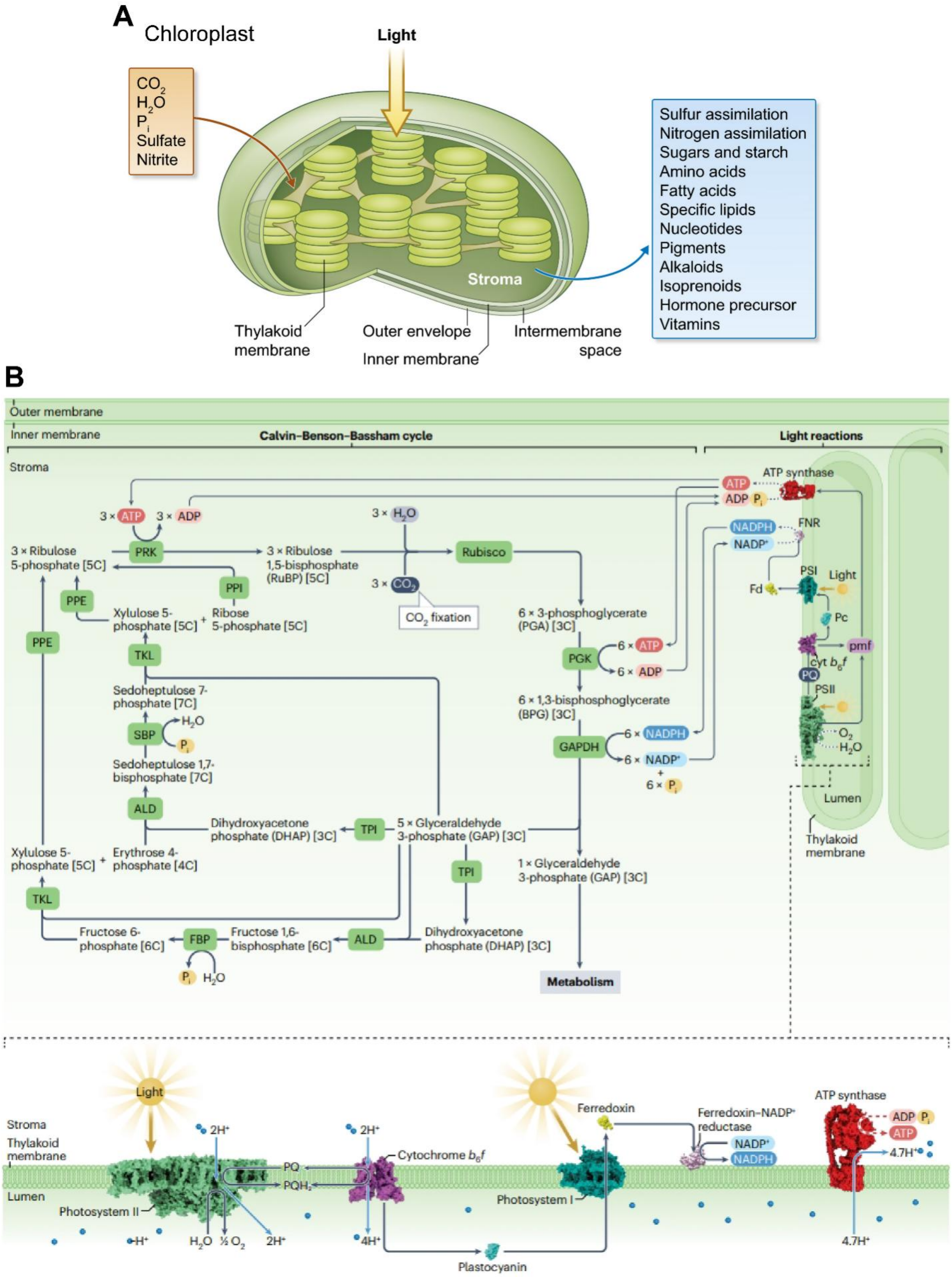


Figure 1.1 Two phases of oxygenic photosynthesis.
(A) Overall structure and metabolic functions of higher plant chloroplasts. Chloroplasts harness sunlight energy to convert low-energy precursors (orange box) into high-energy metabolites

(blue box). **(B)** Overview of the oxygenic photosynthetic linear electron transfer chain and the Calvin-Benson-Bassham (CBB) cycle. In the electron transfer chain (light reactions in thylakoid membrane, right part of this panel and magnified view below), photosystem II (PSII) uses light energy to oxidize water, releasing oxygen, protons (H^+) into the thylakoid lumen, and electrons (e^-) that are transferred to plastoquinone (PQ), which combines with two stromal protons to form plastoquinol (PQH_2). Cytochrome b_6f (Cyt b_6f) then oxidizes PQH_2, transferring electrons to plastocyanin (Pc) and releasing protons into the lumen. Photosystem I (PSI), powered by light, then oxidizes Pc and reduces ferredoxin (Fd), which facilitates $NADP^+$ reduction to NADPH via ferredoxin-$NADP^+$ reductase (FNR). The resulting transmembrane proton motive force (pmf) drives ATP synthase (F-ATPase) to produce ATP from ADP and inorganic phosphate (P_i). In the CBB cycle, which takes place in the stroma of chloroplasts, the key enzyme Rubisco combines CO_2 with RuBP to generate an unstable six-carbon intermediate. This intermediate is then hydrolyzed to create two molecules of 3-phosphoglycerate (3-PG). The 3-PG is then phosphorylated and subsequently reduced to glyceraldehyde-3-phosphate (G3P). G3P can be used in the downstream metabolism for sucrose or starch synthesis, or to regenerate RuBP. In this way, nine ATP molecules and six NADPH molecules are consumed for every output molecule of G3P generated. The structure for the electron transfer chain complexes are as follows: PSII (Protein Data Bank (PDB) ID: 5XNM), Cyt b_6f (PDB ID: 6RQF), PSI (PDB ID: 5L8R), FNR (PDB ID: 1GAQ) and ATP synthase (PDB ID: 6FKF). Figure adapted from (Johnson, 2025; Kirchhoff, 2019).

1.1.1.1. Non-cyclic and cyclic light-dependent reactions

The light-dependent reactions occur in two forms: non-cyclic and cyclic. In the non-cyclic reaction, also known as non-cyclic photophosphorylation, when photons excite electrons in the pigments, the energy is transferred to the reaction center (Chl *a*) via resonance energy transfer (Blankenship & Wiley, 2002), initiating electron flow through an electron transport chain (ETC), consisting of a series of multi-subunit membrane-protein complexes, including PSII, the cytochrome b_6f (Cyt b_6f) complex, and PSI (Johnson, 2025; Nelson & Ben-Shem, 2004). Electrons stripped from water molecules (releasing O_2 as a by-product) travel through the ETC, fueling the pumping of protons (H^+) into the thylakoid lumen. This creates a transmembrane proton motive force (pmf) (Malone et al., 2021; Sarewicz et al., 2021) that powers ATP synthase (F-ATPase) to phosphorylate adenosine diphosphate (ADP) to produce adenosine triphosphate (ATP), a universal energy carrier (Allen, 2002; Johnson, 2025; Mitchell, 1961, 1972). Simultaneously, ferredoxin-$NADP^+$ reductase (FNR) reduces nicotinamide adenine dinucleotide phosphate ($NADP^+$) to nicotinamide adenine dinucleotide phosphate (NADPH), a high-energy electron carrier (Emrich-Mills et al., 2025; Kimata-Ariga et al., 2019). The overall equation for non-cyclic electron flow in green plants and cyanobacteria is:

$$2\ H_2O + 2\ NADP^+ + 3\ ADP + 3\ P_i + \text{light} \rightarrow 2\ NADPH + 2\ H^+ + 3\ ATP + O_2$$

The cyclic reaction, also known as cyclic photophosphorylation, is similar to the non-cyclic reaction, but it occurs only within PSI, and produces only ATP without NADPH production (Allen, 2003). It can meet the additional cellular demand of ATP (Allen, 2003). When light excites electrons in PSI, they are transferred through an ETC involving the Cyt b_6f complex (Yamori et al., 2011). These electrons power the synthesis of ATP and recycle back to PSI to complete the cycle.

It is important to note that photosynthetic organisms only absorb certain wavelengths of light to support photosynthesis, depending on their pigment composition. The non-absorbed parts of the light spectrum give photosynthetic organisms their color (e.g., green plants, blue-green cyanobacteria and purple bacteria), and represent the least efficient light for their photosynthesis (Leister, 2023). Together, these light-dependent reactions produce ATP and NADPH, which fuel the next phase of photosynthesis.

1.1.1.2. Light-independent reactions

The light-independent reactions, also known as the Calvin-Benson-Bassham (CBB) cycle (Bassham et al., 1950), occur in the stroma of chloroplasts of plants and algae, and in the cytosol of photosynthetic prokaryotes, where atmospheric CO_2 is fixed into organic molecules (Figure **1.1B**). The cycle begins when the enzyme Rubisco binds to the five-carbon sugar, ribulose-1,5-bisphosphate (RuBP) and catalyzes its carboxylation, producing an unstable six-carbon phosphorylated intermediate, known as 3-keto-2-carboxyarabinitol-1,5-bisphosphate (3-KCABP or KCABP), which rapidly splits into two molecules of a three-carbon compound, glycerate 3-phosphate, also known as 3-phosphoglycerate (3-PG) (Bracher et al., 2017; Raines, 2003). Using ATP and NADPH produced from the light-dependent reactions, 3-PG is first phosphorylated to 1,3-bisphosphoglycerate (1,3-BPG) by the enzyme phosphoglycerate kinase and then reduced to glyceraldehyde-3-phosphate (G3P) by the enzyme glyceraldehyde 3-phosphate dehydrogenase. G3P is the precursor of carbohydrate products such as glucose, sucrose, and starch, and the starting point for other biosynthetic pathways, including those for amino acids, fatty acids, pigments and nucleotides (Kirchhoff, 2019). While some G3P leaves the cycle to form these biomolecules, most is recycled through a series of complex enzymatic reactions to regenerate RuBP to ensure the continuity of the cycle (Miziorko & Lorimer, 1983; Raines, 2011). The overall equation for the CBB cycle is as follows:

$$3\ CO_2 + 9\ ATP + 6\ NADPH + 6\ H^+ \rightarrow G3P + 9\ ADP + 8\ P_i + 6\ NADP^+ + 3\ H_2O$$

Despite its critical role in carbon fixation, Rubisco suffers from inefficiencies such as poor selectivity between CO_2 and the highly abundant O_2. Photorespiration is a metabolic process that initiates when Rubisco oxygenates its substrate RuBP rather than carboxylating it (Portis & Parry, 2007) (Figure **1.2**). This reaction produces one molecule each of 3-PG and 2-phosphoglycolate (2-PG), a toxic compound that must be rapidly metabolized and recycled to 3-PG in chloroplasts (Bauwe, 2023; Peterhansel et al., 2010; Segura Broncano et al., 2023). However, photorespiration is a complex and energetically costly enzymatic pathway that reduces the efficiency of photosynthesis and results in the loss of fixed carbon (Bracher et al., 2017). This process decreased the photosynthetic output of US soybean and wheat, both of which are C_3 plants – the most common type of plant on Earth, by 36% and 20%, respectively (Walker et al., 2016).

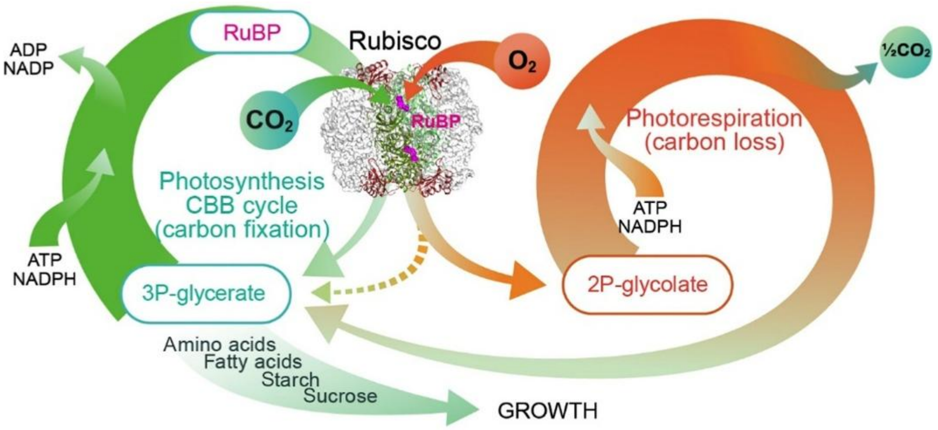


Figure 1.2 Carbon fixation and photorespiration.
In addition to binding CO_2 and participating in the CBB cycle, Rubisco reacts with O_2 to produce one molecule each of 3-PG and 2-PG. The latter is toxic and must be metabolized in an energy-dependent reaction involving peroxisomes and mitochondria (the photorespiratory cycle). Figure adapted from (Hayer-Hartl & Hartl, 2020).

1.2. Rubisco

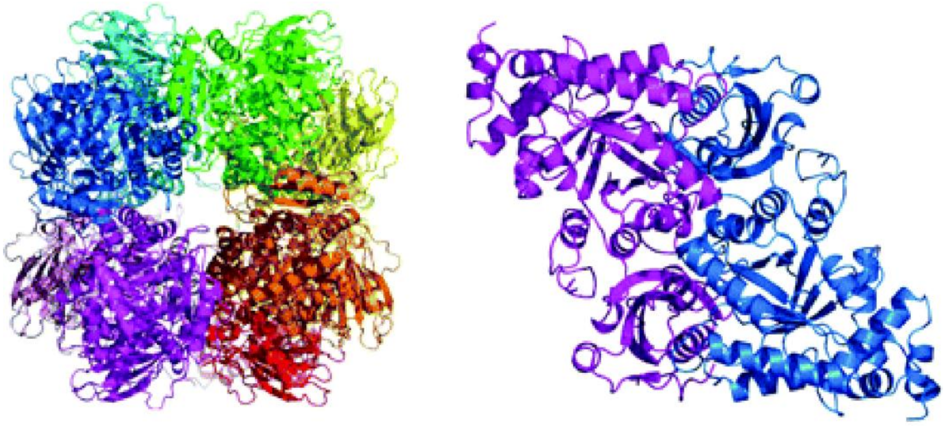
Rubisco is an ancient enzyme that evolved more than 3.5 billion years ago when the Earth's atmosphere was rich in CO_2 and free of O_2 (Whitney, Houtz, et al., 2011). As a key enzyme in carbon fixation, Rubisco catalyzes the carboxylation of RuBP in the CBB cycle of photosynthesis, contributing to the formation of carbohydrates, fatty acids, and amino acids.

However, Rubisco is a comparatively inefficient enzyme, with a catalytic rate of ~3-10 CO_2 molecules per second (Ellis, 2010; Hartman & Harpel, 1994). Moreover, the similar size and electrostatic potential of CO_2 and O_2 make it difficult for Rubisco to discriminate between them. Additionally, the multi-step catalytic process can lead to the formation of "misfire" products, such as xylulose-1,5-bisphosphate (XuBP), and 2,3-pentodiulose-1,5-bisphosphate (PDBP). These phosphorylated sugar derivatives bind tightly to the active sites of Rubisco, rendering the enzyme in an inactive state and inhibiting its catalytic function (Parry et al., 2008; Pearce, 2006). Rubisco activase (Rca), an ATP-dependent catalytic chaperone that has co-evolved with Rubisco, facilitates the reactivation of Rubisco by inducing conformational changes and promoting the release of these sugar phosphate inhibitors from its active sites (Andersson, 2008; Flecken et al., 2020; Portis, 2003; Stotz et al., 2011; Waheeda et al., 2023).

1.2.1. The different forms of Rubisco

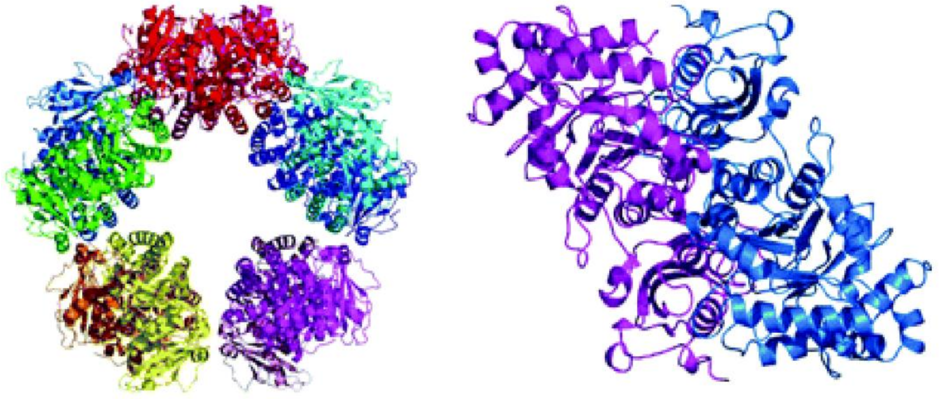
There are four forms of Rubisco found in nature, referred to as forms I-IV (Figure **1.3**) (Tabita et al., 2008). Form I Rubiscos are the most abundant, found in plants, algae, cyanobacteria and proteobacteria. They typically form a ~530 kDa hexadecameric ($RbcL_8S_8$) complex composed of eight large (RbcL, ~50-55 kDa) and eight small (RbcS, ~12-18 kDa) subunits (Andersson & Backlund, 2008; Bracher et al., 2017). They can be further subdivided into form IA (found in some cyanobacteria and some proteobacteria), IB (found in higher plants, green algae, and most cyanobacteria), IC (found in some proteobacteria) and ID (found in certain diatoms and red algae) (Tabita et al., 2008). Furthermore, form IA and IB Rubiscos are classified as green-type because they evolved in cyanobacteria and green algae/plants, while form IC and ID Rubiscos are classified as red-type because they originated in red algae and later spread to diatoms and photosynthetic bacteria (Badger & Bek, 2008; Tabita, 1999; Tabita et al., 2008). The core of the complex adopts a cube-shaped structure formed by four antiparallel RbcL dimers ($RbcL_2$), capped by four RbcS subunits at the top and four at the bottom to form the functional holoenzyme (Hayer-Hartl & Hartl, 2020). Form II Rubiscos are found in some bacteria and dinoflagellates, consisting only of large subunits, forming one or more $RbcL_2$ (Tabita et al., 2008; Tsai et al., 2015). Both form I and form II Rubiscos catalyze the carboxylation reaction in the CBB cycle.

In contrast, form III Rubiscos, found in archaea, contain only $RbcL_2$ units that generally assemble into higher oligomers, such as the Rubisco from *Thermococcus kodakarensis*, which forms ring-shaped complexes of five RbcL$_2$ units (Huang & Szebenyi, 2023). Although these enzymes retain carboxylase/oxygenase activities, the CBB cycle has not been found in archaea (Kono et al., 2017). Instead, form III Rubiscos are involved in nucleotide and nucleoside metabolism (Sato et al., 2007). In addition, some bacteria and archaea contain Rubisco-like proteins (RLPs), which are classified as form IV Rubisco. These proteins are not involved in carbon fixation but in other metabolic pathways (Hanson & Tabita, 2001).



Form I
Spinacia oleracea

Form II
Rhodospirillum rubrum



Form III
Thermococcus kodakaraensis

Form IV
Chlorobium tepidum

Figure 1.3 Representative structures of different forms of Rubisco.
All forms of Rubisco consist of dimers of large subunits (RbcL). Form I contains four dimers (RbcL$_2$) with small subunits attached to the top and bottom. Only Form I has these small subunits. Form II consists of one or more RbcL$_2$. Form III has one or more RbcL$_2$, showing five dimers as above. Form IV (the Rubisco-like proteins or RLPs) only has an RbcL$_2$ structure. Form I Rubisco (PDB ID: 1RCX), form II Rubisco (PDB ID: 5RUB), form III (PDB ID: 1GEH), form IV Rubisco (PDB ID: 1YKW). Figure adapted from (Tabita et al., 2008).

The various forms of Rubisco share a similar RbcL structure, comprising two distinct domains: an N-terminal α+β domain (~150 amino acids) and a C-terminal region (~325 amino acids) containing a $\beta_8\alpha_8$ triosephosphate isomerase (TIM)-barrel domain (~320 amino acids) followed by a short flexible tail (~15 amino acids) (Bracher et al., 2017). Each antiparallel $RbcL_2$ has two active sites located at the interface between the N-terminal domain of one subunit and the C-terminal TIM-barrel domain of the adjacent subunit. While not directly contributing to the formation of active sites, the RbcS subunits are considered an evolutionary adaption to the increasing O_2 levels in the atmosphere, serving to stabilize the RbcL active sites and enhancing CO_2/O_2 specificity of form I Rubisco (Andersson & Backlund, 2008; Spreitzer, 2003) – thereby optimizing catalytic activity.

1.2.2. Auxiliary chaperone factors for Rubisco biogenesis

Due to its catalytic inefficiency and slow turnover rate, Rubisco has been one of the most important targets for improving photosynthetic efficiency for decades with the goal to increase crop yields (Gionfriddo et al., 2023; Parry et al., 2013; Whitney, Houtz, et al., 2011). The assembly of form IB Rubisco of cyanobacteria *in vitro* depends on two key factors – RbcX and Rubisco accumulation factor Raf1. The *RbcX* gene is located between the *RbcL* and *RbcS* genes within the Rubisco operon of several cyanobacteria species, and is evolutionarily conserved in plants and green algae (Bracher et al., 2017). Even when *RbcX* is outside the Rubisco operon, its co-expression with Rubisco *in vitro* enhances the biogenesis of the holoenzyme $RbcL_8S_8$ (Emlyn-Jones et al., 2006). However, *in vitro* reconstitution reactions with GroEL/GroES and RbcX yielded only ~40% of active $RbcL_8S_8$ (Liu et al., 2010). Raf1 is conserved in all eukaryotic and cyanobacterial organisms that also contain RbcX (Hayer-Hartl & Hartl, 2020). Raf1 in combination with GroEL/GroES proved highly efficient in mediating Rubisco assembly for some cyanobacterial species, both *in vitro* and upon co-expression in *E. coli* (Hauser et al., 2015; Kolesinski et al., 2014).

The folding and assembly of plant Rubisco is a complex process involving several limiting steps. The RbcL subunit is encoded by the chloroplast genome, while the RbcS subunit is encoded in the nucleus in plants (Tabita, 1999). RbcL is prone to misfolding and aggregation; therefore, non-native RbcL requires the chloroplast chaperonin system (Cpn60αβ with its cofactors Cpn10/20) for proper folding. Subsequently, a number of auxiliary factors, including RbcX, Rubisco accumulation factors (Raf1 and Raf2), and bundle sheath defective-2 (BSD2) are required to prevent RbcL aggregation and facilitate the assembly to $RbcL_2$ units and to the

$RbcL_8$ core of Rubisco. Finally, binding of RbcS to the $RbcL_8$ results in the formation of the functional $RbcL_8S_8$ holoenzyme (Aigner et al., 2017; Hayer-Hartl & Hartl, 2020) (Figure **1.4**).

The RbcX proteins, which are nuclear-encoded in plants and green algae, form dimers composed of ~15 kDa subunits. The RbcX dimer contains a central hydrophobic groove that binds the RbcL C-terminal sequence motif EIKFE(F/Y)X (Saschenbrecker et al., 2007). RbcX acts as a “molecular staple”, stabilizing $RbcL_2$ and promoting the assembly of $RbcL_8$ with four RbcX dimers bound at the top and four at the bottom (Hayer-Hartl & Hartl, 2020). Similarly, Raf1 forms a dimer consisting of ~40 kDa subunits, and contains an N-terminal α-helical domain and a C-terminal β-sheet dimerization domain connected by a flexible linker (Hauser et al., 2015). Raf1 also functions to stabilize $RbcL_2$, with its α-helical domains contacting the top and bottom edges of the $RbcL_2$, while the β domains being positioned at the equator of the $RbcL_8$ complex. The end-state intermediate of Raf1-mediated Rubisco assembly is an $RbcL_8$ complex bound to four Raf1 dimers. Both RbcX and Raf1 dynamically interact with $RbcL_2$, allowing their displacement by RbcS to form the $RbcL_8S_8$ holoenzyme. BSD2, first identified in maize, is a chloroplast-specific monomeric protein with a molecular weight of 8-10 kDa. Like RbcX and Raf1, BSD2 acts as a Rubisco assembly chaperone, with eight BSD2 monomers binding to $RbcL_8$ to stabilize the intermediate structure before RbcS subunits dock to form the $RbcL_8S_8$ complex (Aigner et al., 2017). In addition to its role in Rubisco assembly, BSD2 has been implicated in post-transcriptional control of RbcL and maintaining Rubisco activity (Busch et al., 2020; Doron et al., 2014; Salesse et al., 2017). Together, RbcX, Raf1 and BSD2 have similar functions in assembly of $RbcL_8S_8$ holoenzyme by stabilizing $RbcL_2$ through distinct interactions with RbcL, suggesting they may function in parallel or cooperate *in vivo* to achieve efficient assembly at a biologically relevant timescale (Bracher et al., 2017). Raf2, found in chloroplasts and cyanobacteria containing form IA Rubisco, forms ~38 kDa dimers. Although Raf2 shares sequence homology to pterin-4α-carbinolamine dehydratase (PCD) enzymes, it lacks catalytic activity (Naponelli et al., 2008; Wheatley et al., 2014). Raf2 has also been proposed to function in Rubisco assembly by interacting with RbcS and facilitating $RbcL_8$ formation. To date the exact mechanism underlying the role of Raf2 in Rubisco biogenesis remains unknown. Notably, RbcS synthesized in the cytosol, may also require chaperone assistance (by stromal Hsp70) for folding after being imported into chloroplasts (Trosch et al., 2015).

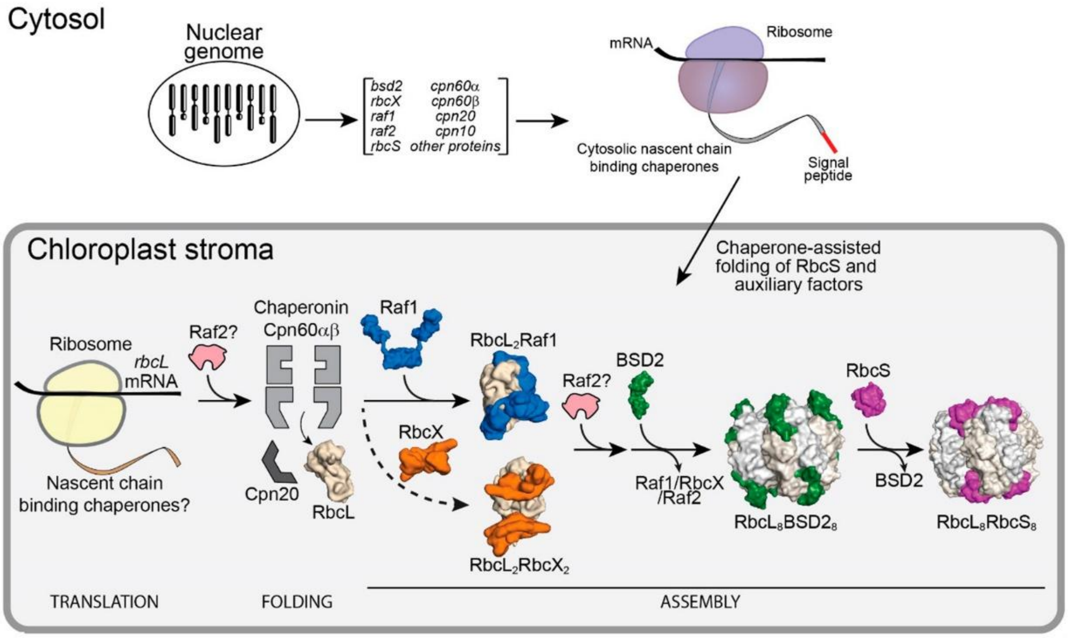


Figure 1.4 Molecular chaperones involved in the biogenesis pathway of plant Rubisco. The chloroplast-encoded RbcL subunit is translated on stromal ribosomes and folded by the Cpn60αβ/Cpn10/20 chaperonin system. Meanwhile, the nuclear-encoded RbcS subunit and assembly factors are synthesized in the cytosol, imported into the chloroplast, and folded with the help of Hsp70/Hsp40. The folded RbcL subunits assemble into $RbcL_2$ dimers and $RbcL_8$ cores assisted by Raf1 and RbcX. Then, BSD2 stabilizes the $RbcL_8$ complex. The final holoenzyme forms when RbcS binds and displaces BSD2. Raf2 also plays an essential but incompletely defined role in $RbcL_8$ assembly, potentially acting either upstream or downstream of chaperonin-mediated folding. This tightly regulated process ensures the proper assembly of functional Rubisco. Abbreviations: RbcL, Rubisco large subunit; RbcS, Rubisco small subunit; Cpn, chaperonin; Raf, Rubisco accumulation factor; BSD2, bundle sheath defective-2. Figure adapted from (Wilson & Hayer-Hartl, 2018).

1.3. Bacterial microcompartments (BMCs)

Bacterial microcompartments (BMCs) are proteinaceous, self-assembling organelles that encapsulate specialized metabolic enzymes within a selectively permeable protein shell, analogous to the lipid-based membrane of eukaryotic organelles (Kennedy et al., 2021; Kerfeld et al., 2018). These icosahedral organelles, typically 40 to 600 nm in diameter (Erbilgin et al., 2014; Liberton et al., 2011) and composed of several thousand protein subunits of 10 to 20 different types (Bobik et al., 2015), improve metabolic efficiency by maintaining enzyme stability, preventing side reactions and sequestering toxic or volatile metabolic intermediates (Kerfeld et al., 2010). Their spatial organization also creates local high concentrations of enzyme substrates that further optimize enzymatic activity (Turmo et al., 2017). BMCs are classified according to their functional roles and associated metabolic pathways, with the two well-known classes being anabolic carboxysomes and catabolic metabolosomes (Axen et al., 2014; Sutter et al., 2021) (Figure **1.5**). Carboxysomes serve as a core component of the CO_2-concentrating mechanism (CCM), encapsulating the primary carboxylating enzyme Rubisco – the most abundant enzyme on Earth – as well as carbonic anhydrase (CA) that converts cytoplasmic bicarbonate to CO_2 (Espie & Kimber, 2011; Kerfeld & Melnicki, 2016; Rae et al., 2013). Metabolosomes encapsulate and optimize metabolic pathways for the degradation of organic substrates, such as 1,2-propanediol (PDU and GRM3), ethanolamine (EUT), ethanol (ETU), choline (Cut and GRM2), small saccharides (GRM5 and PVM), xanthine (XAU), amino alcohols (RMM/AAUM), taurine, and aromatic compounds (ARO) (Doron & Kerfeld, 2024; Kerfeld et al., 2010). These microcompartments play critical roles in ecological nutrient cycling, host-microbe symbiosis, bacterial pathogenesis, and even human cardiovascular disease (Kerfeld & Erbilgin, 2015). Emerging as elegant solutions to atmospheric and metabolic challenges, BMCs also offer a promising blueprint for engineered nanomaterials and metabolic engineering.

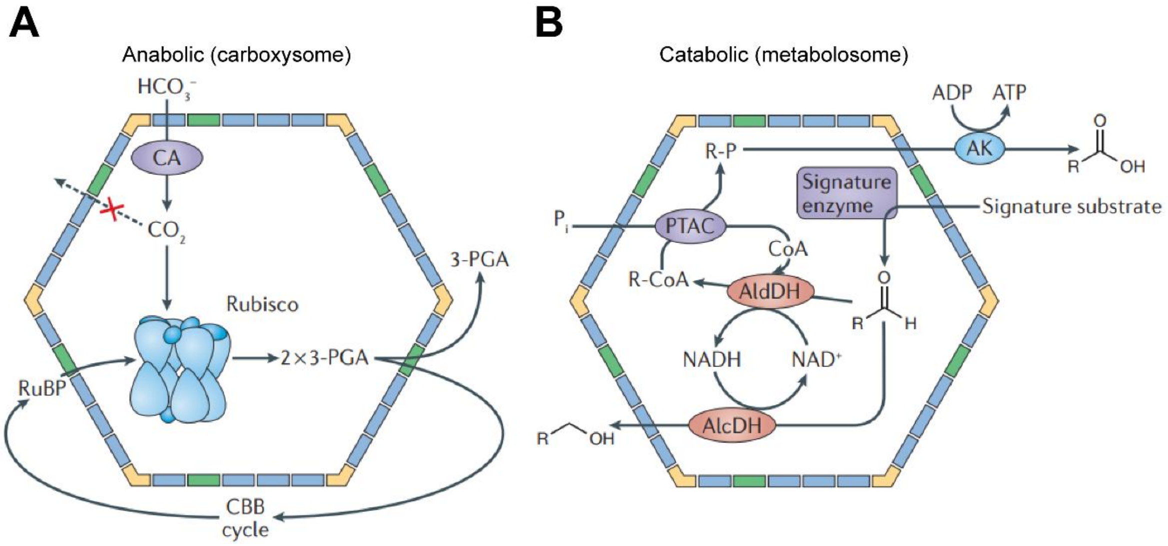


Figure 1.5 Schematic of (A) carboxysome and (B) metabolosome function.
(A) The essential enzymes Rubisco and carbonic anhydrase (CA) are encapsulated within the carboxysome, a proteinaceous compartment with a selectively permeable shell that regulates the flux of specific metabolites, enabling integration with cellular metabolism. **(B)** Metabolosomes have more diverse initial substrates than carboxysomes, but they typically share a common core biochemistry that is based on a substrate-defining signature enzyme, an alcohol dehydrogenase (AlcDH), a Coenzyme A (CoA)-dependent aldehyde dehydrogenase (AldDH) and a phosphotransacylase (PTAC). The signature enzyme metabolizes the substrate to a toxic aldehyde intermediate, which is contained and sequentially detoxified by AlcDH and AldDH to yield an alcohol and a CoA derivate of a carboxylic acid; furthermore, PTAC regenerates CoA, producing a phosphorylated product (R-P) that fuels ATP generation via an

acetyl kinase. This process prevents both cellular damage and carbon loss. Abbreviations: P_i, inorganic phosphate. Figure adapted from (Kerfeld et al., 2018).

1.3.1. Assembly of bacterial microcompartments

Despite their chemically diverse functions, BMC shells assemble through conserved interactions among different classes of homologous shell proteins. These include the self-assembling hexameric (BMC-H) and trimeric (BMC-T) oligomers (containing one or two Pfam00936 domains, respectively), which form the facets of the shell, and the pentameric (BMC-P) proteins, which contain the Pfam03319 domain and cap the vertices to ensure geometric closure of the icosahedral structure (Sutter et al., 2017; Trettel et al., 2024) (Figure **1.6**). Notably, BMC-H proteins assemble into a cyclic hexamer with a convex and a concave side, whereas BMC-T proteins form trimers that resemble BMC-H hexamers in size and shape, referred to as pseudohexamers (Cai et al., 2013). BMC-T proteins are further divided into two subtypes: single-trimer BMC-T^s and double-trimer BMC-T^{ds}, the latter resulting from the dimerization of two trimers via their concave faces.

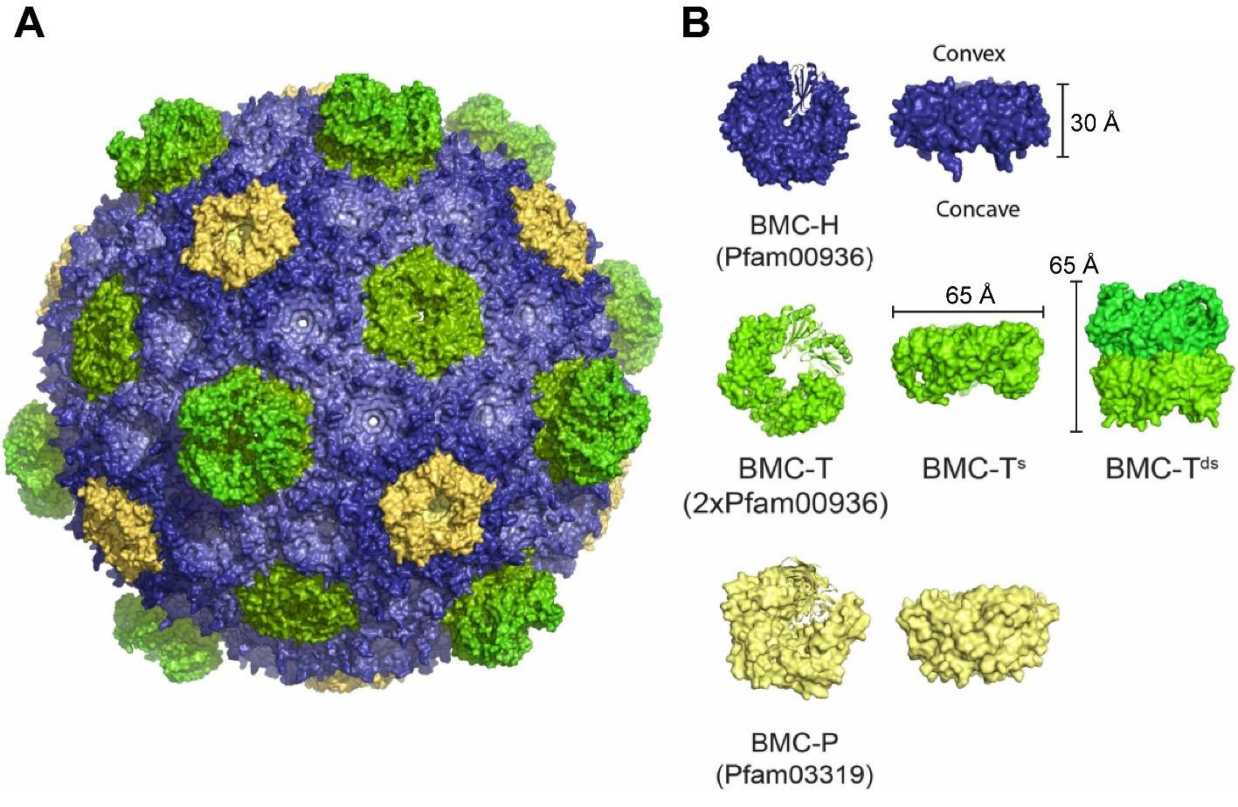


Figure 1.6 Structural representation of a bacterial microcompartment (BMC) shell and its shell protein building blocks.

(A) The overall structure of a polyhedral BMC shell assembled from three types of protein building blocks. **(B)** The structural components of the shell: BMC-H (Pfam00936), BMC-T (2xPfam00936) either single (BMC-T^s) and/or stacked dimers of trimers (BMC-T^{ds}), and BMC-P (Pfam03319). Complete shell (PDB ID: 5V74), BMC-H (PDB ID: 5DJB), BMC-T^s (PDB ID: 5DIH), BMC-T^{ds} (PDB ID: 5V75). BMC-P structure was extracted from the whole shell structure. Figure modified from (Doron & Kerfeld, 2024).

The most abundant BMC-H hexamers interact with all other types of shell proteins, including other BMC-H hexamers, whereas the BMC-T^s trimer, the BMC-T^{ds} dimer of trimers, and the BMC-P pentamer interact exclusively with BMC-H hexamers. There are four types of interfaces between shell proteins in the synthetic BMC shell (~40 nm) of the myxobacterium *Haliangium ochraceum* (*H. ochraceum*): two distinct hexamer-hexamer interfaces, one hexamer-pentamer interface, and one hexamer-pseudohexamer interface (Figure **1.7**) (Sutter et al., 2017). The KAA and PRPH motifs are broadly conserved across BMC-H hexamers and mediate hexamer-hexamer interactions (Sutter et al., 2017). At the hexamer-pentamer interface, the conserved pentamer residues S13, the GAGxGE motif, and the I(V/I)D motif interact with hexamers (Sutter et al., 2017). Additionally, a specific hydrogen bonding network, supported by the KAA and PRPH motifs of the hexamers along the outer edge of each subunit, plays a central role in forming the interface with the pentamer (Sutter et al., 2017).

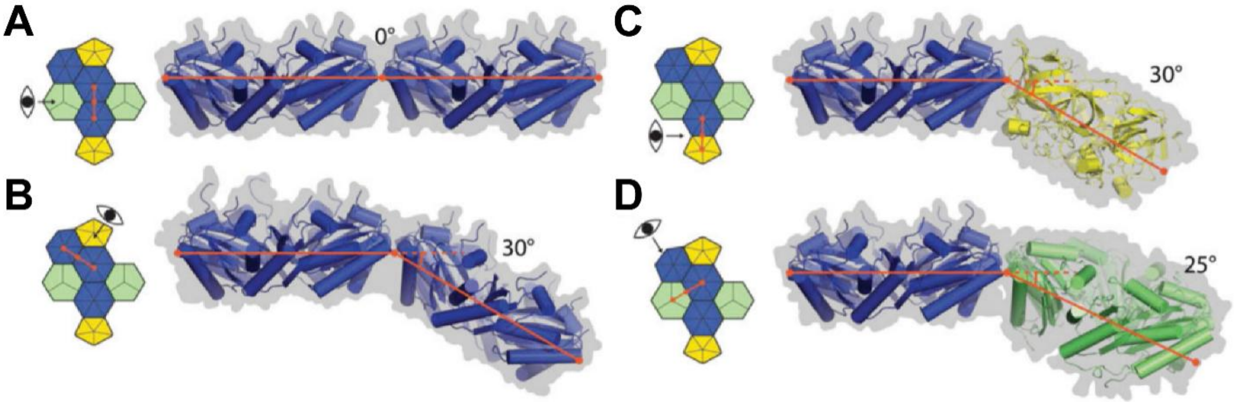


Figure 1.7 Overview of the four distinct interfaces between the hexamers, pentamers, and pseudohexamers.
Structures are displayed in cartoon representation (surface view as grey background) with a pictogram showing their location on the shell. **(A)** Coplanar hexamer-hexamer interface connecting two pentamers. **(B)** Hexamer-hexamer interface that surrounds a pentamer. **(C)** Hexamer-pentamer interface. **(D)** Hexamer-pseudohexamer interface. The tilt angles of the interfaces are labeled. Color coding: hexamer (blue), pentamer (yellow), and pseudohexamer (lime green). Figure adapted from (Sutter et al., 2017).

Recently, six distinct interfaces were identified in purified intact α-carboxysomes (~90 nm) from the marine cyanobacterium *Prochlorococcus* MED4. These interfaces include four hexamer-hexamer interfaces with tilt angles of 0°, 10°, 12°, and 17°, one hexamer-pentamer interface, and one hexamer-pseudohexamer interface (Figure **1.8**) (Zhou et al., 2024). These interfaces allow for more variability in tilt angles between hexamers than the BMC shell of *H. ochraceum*, suggesting that size variations in the BMC shell are achieved by adjusting these angles (Zhou et al., 2024). These interaction patterns appear to be universal among BMCs, as observed in the *H. ochraceum* BMC shell structure. Interestingly, recent studies have identified some simple shells in nature, including ARO, PVM, the *Acidobacterium* microcompartment (ACI) and the Sugar Phosphate Utilization microcompartment, which consist primarily of BMC-H and BMC-P proteins (Sutter et al., 2021).

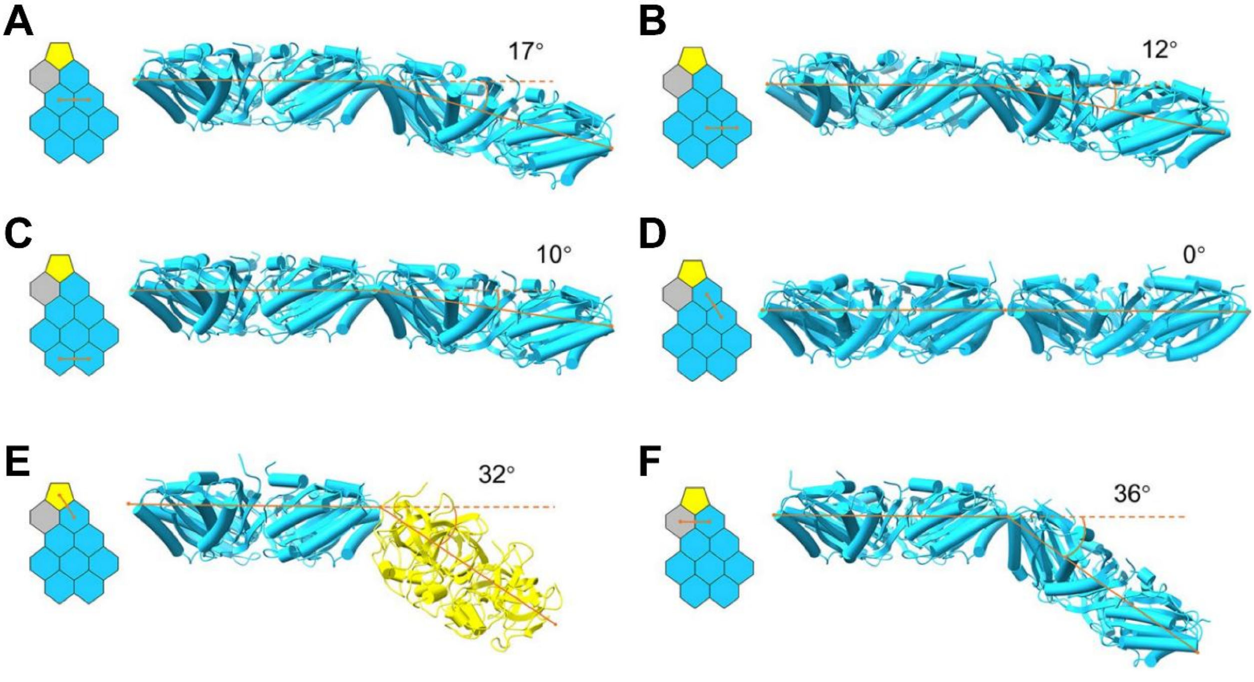


Figure 1.8 Overview of the distinct interfaces among the hexamers, pentamers, and pseudohexamers.
Structures are shown as cartoons with a pictogram showing their location on the shell. The tilt angles of interfaces are labeled. **(A-D)** Hexamer-hexamer interfaces with tilt angles of 17° **(A)**, 12° **(B)**, 10° **(C)**, 0° **(D)**. **(E)** Hexamer-pentamer interface. **(F)** Hexamer-pseudohexamer interface. The tilt angles of the interfaces are labeled. Color coding: hexamer (blue), pentamer (yellow), and pseudohexamer (gray). Figure adapted from (Zhou et al., 2024).

In most BMCs, the enzymatic core is targeted into the interior of the shell by a conserved encapsulation peptide (EP). This short peptide, typically 15-20 residues long, is located at either the N- or C-terminus of cargo proteins (Aussignargues et al., 2015). It was first identified as an N-terminal sequence extension on PduP (AldDH) in metabolosome homologs (Fan et al., 2010) and was later found at the C-terminus of the carboxysomal protein ApN (Kinney et al., 2012). EP is predicted and experimentally verified to form an amphipathic α-helix that mediates specific binding to the inner surface of shell proteins (Kerfeld et al., 2018; Kinney et al., 2012). The N-terminal targeting sequence of PduP has been shown to bind the C-terminal sequences of the shell proteins PduA and PduJ, mediating enzyme encapsulation (Fan et al., 2012). Engineered organelles have been generated by fusing foreign proteins, such as green fluorescent protein (GFP), glutathione S-transferase (GST), or maltose-binding protein (MBP), to EPs and incorporating them into heterologously expressed BMC shells, suggesting that EPs are important for the interactions with shell proteins, but not for organizing the luminal contents of BMCs (Chowdhury et al., 2014; Kerfeld & Erbilgin, 2015). It is also of note that some BMC luminal proteins lack targeting sequences, suggesting that alternative mechanisms may facilitate their assembly (Chowdhury et al., 2014).

1.3.2. Metabolite transport across the shell

All shell proteins possess a central pore along their axis of symmetry, with variations in size and electrostatic properties on the concave and convex sides allowing selective metabolites transport across BMC shells. BMC-H proteins contain smaller pores with a diameter of ~4-7 Å that allow the passage of small metabolites, such as bicarbonate (HCO_3^-) in carboxysomes or alcohols and carboxylic acids in metabolosomes (Kerfeld & Erbilgin, 2015). In contrast, stacked BMC-T^{ds} pores are typically larger, ranging from ~12-14 Å and have been proposed to function in an airlock-like mode (Cai et al., 2013; Klein et al., 2009; Larsson et al., 2017), with their opening and closing regulated by conformational changes of the surrounding side chains (Kerfeld et al., 2018) (Figure **1.9**), enabling the transport of bulkier metabolites. The carboxysome shell is selectively permeable to protons and reducing agents, maintaining a pH equilibrium between the carboxysome interior and the cytoplasm (Huang et al., 2022; Menon et al., 2010), while preserving an oxidizing internal environment within the carboxysomes (Chen et al., 2013), essential for Rubisco condensation, dehydration of HCO_3^-, CA activity, and carboxysome function (Liu, 2022; Long et al., 2021; Pena et al., 2010). The distinct permeability properties conferred by this diverse array of shell proteins distinguish BMCs from other single-component proteinaceous shell systems, such as bacterial encapsulin nanocompartments (Sutter et al., 2008).

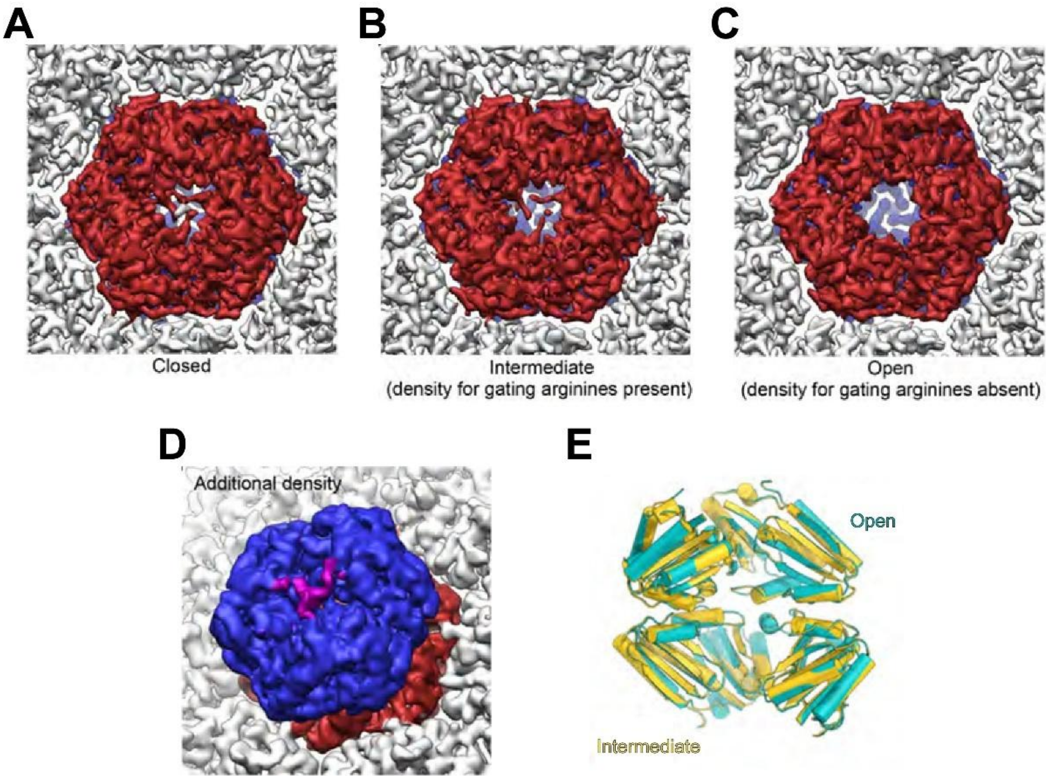


Figure 1.9 Structure and conformational differences of shell-embedded BMC-T^{ds} in the *Haliangium ochraceum* BMC shell.
(A) Closed conformation of the shell-embedded BMC-T^{ds} trimer. The shell-embedded BMC-T^{ds} trimer (dark red) and the protruding trimer (blue) are shown as viewed from the inside of the shell. The inner pore is closed, with three gating arginine residues forming salt bridges with glutamate residues of the neighboring subunit. **(B)** Intermediate conformation of the inner BMC-T^{ds} ring. The gating arginines remain visible, while the overall structure of the inner BMC-T^{ds} subunits has already undergone the conformational changes. **(C)** Open conformation of the pore. The density of the gating arginine residues is absent in the shell-embedded trimer, but is clearly present in the protruding trimer. **(D)** Airlock-like transport mechanism. The inner pore is open while the outer pore is closed, which is consistent with a selective transport mechanism for small molecules across the BMC shell. Weak density (purple) above the closed pore in the protruding BMC-T^{ds} trimer may correspond to the C terminus of BMC-T^{ds}. **(E)** Comparison of overall conformation of BMC-T^{ds} trimers (intermediate and open states, respectively). BMC-T^{ds} closed state (PDB ID: 6MZU; Electron Microscopy Data Bank (EMD) ID: 9307), BMC-T^{ds} widened inner ring (PDB ID: 6MZV; EMD ID: 9308), BMC-T^{ds} open inner pore (PDB ID: 6N07; EMD ID: 9312). Figure modified from (Greber et al., 2019).

1.3.3. Structural flexibility of shell proteins

A critical requirement for the formation of a tightly packed intact shell is the precise shape complementarity among the shell proteins, which must adopt polymorphic conformations to accommodate different edge interactions and other physical stresses. For instance, the BMC-H protein EutS from *Salmonella enterica* has been shown to crystallize in a bent conformation

(Tanaka et al., 2010), whereas CutR from *Streptococcus intermedius* has various screw conformations (Ochoa et al., 2020). Similarly, the BMC-P protein EutN has been shown to crystallize as a homohexamer (Trettel et al., 2024; Wunderlich et al., 2004). Although the physiological relevance of these observations remains uncertain, they suggest that the shell proteins have the structural flexibility needed to adapt to their local environment.

1.4. CO_2-concentrating mechanisms (CCMs)

Photosynthetic organisms, such as cyanobacteria, algae and certain plants (C_4 and crassulacean acid metabolism, CAM plants), utilize CO_2-concentrating mechanisms (CCMs) to enhance carbon fixation by increasing the concentration of CO_2 around Rubisco. Cyanobacteria achieve this through specialized microcompartments called carboxysomes, while algae use pyrenoids. In contrast, C_4 and CAM plants separate the steps of CO_2 uptake and storage for carbon fixation spatially or temporally, respectively. By concentrating CO_2, CCMs suppress Rubisco's oxygenase activity, thereby enhancing photosynthetic efficiency in low-CO_2 aquatic habitats and hot, arid environments.

1.4.1. Carboxysomes

Cyanobacteria (also known as blue-green algae) have existed for about 3.5 billion years (Schopf & Packer, 1987) and have endured dramatic environmental changes, including the steep decline in atmospheric CO_2 levels (Badger & Price, 2003). A remarkable adaptation to these changes was the evolution of carboxysomes: a unique microcompartment surrounded by a proteinaceous outer shell, found in many chemoautotrophic proteobacteria and all cyanobacteria (Kerfeld & Melnicki, 2016) (Figure **1.10**). The cyanobacterial CCM consists of two interconnected systems: (i) bicarbonate pumps in the plasma membrane, which use energy-driven processes to import inorganic carbon (Ci), including CO_2 and HCO_3^-, into the cytosol, and membrane-associated CO_2 hydration enzymes that elevate cytosolic bicarbonate levels; and (ii) the carboxysome, where CA rapidly converts cytosolic HCO_3^- into CO_2, the substrate for Rubisco (Cannon et al., 2010; Price, Shelden, et al., 2011) (Figure **1.10E**). The proteinaceous shell of the carboxysome acts as a selectively permeable barrier, permitting the influx of HCO_3^- while restricting the escape of CO_2 generated by CA within the microcompartment (Figure **1.10E**). This unique architecture creates a CO_2-rich microenvironment that directly addresses the two fundamental limitations of Rubisco: its slow catalytic rate and its competing oxygenase activity with O_2, leading to photorespiration and the production of toxic phosphoglycolate. By increasing local CO_2 concentration, carboxysomes significantly enhance the efficiency of carbon fixation while minimizing energy-wasting photorespiration (Blikstad et al., 2019). It is estimated that up to 25% of all carbon fixation on Earth occurs within carboxysomes (Rae et al., 2013). Disruption of the carboxysome shell abolishes the ability of cyanobacteria to grow under atmospheric CO_2 (0.04%), forcing reliance on high CO_2 levels for autotrophic growth (English et al., 1995; Rae et al., 2013).

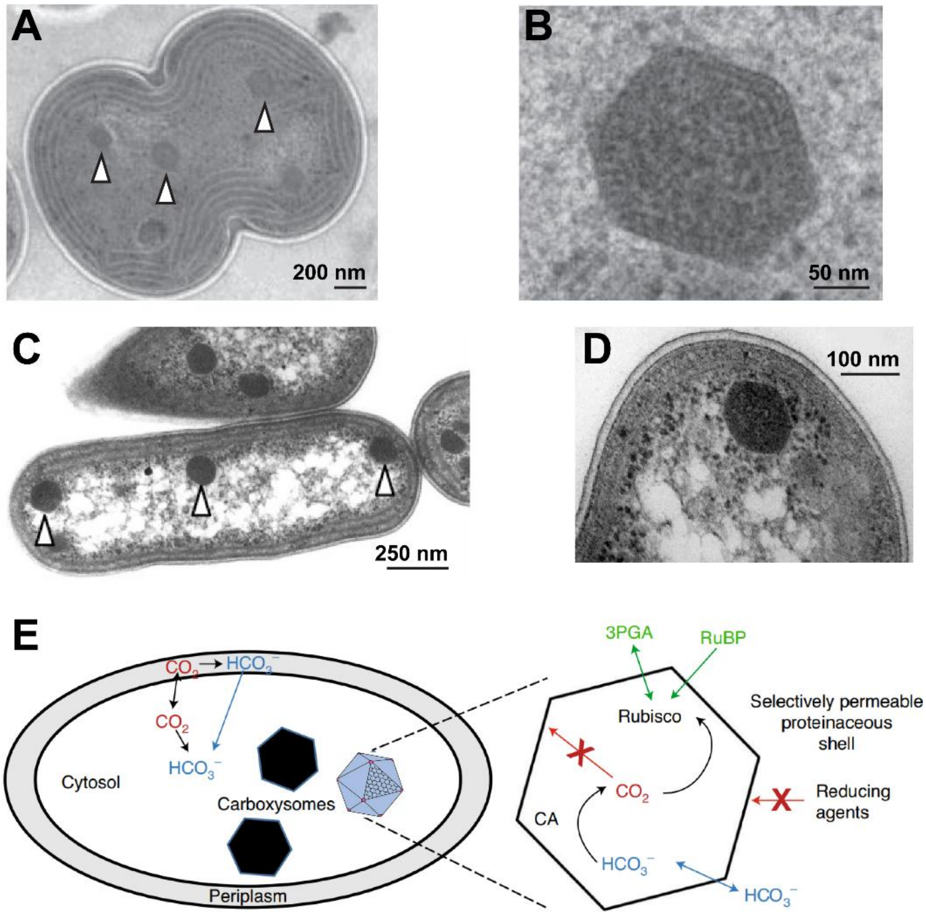


Figure 1.10 Electron micrographs of various carboxysomes and a model for their function. (A) Transmission electron micrograph of β-carboxysomes (arrowheads) in a dividing *Synechocystis* sp. PCC 6803 (*Syn*6803) cell. **(B)** An enlargement of a single carboxysome from *Syn*6803. **(C)** β-carboxysomes (arrowheads) in *Synechococcus elongatus* PCC 7942 (*Se*7942). **(D)** Close-up of a β-carboxysome from *Se*7942. Scale bars are indicated. **(E)** CO_2-concentrating mechanism (CCM) of β-cyanobacteria. The proteinaceous shell allows the entry of HCO_3^-, but restricts the escape of CO_2 generated by CA. It also restricts the entry of reducing agents from the cytosol, creating an internal oxidizing environment within the carboxysome. 3PGA, 3-phosphoglycerate; RuBP, ribulose-1,5-bisphosphate; CA, carbonic anhydrase. Figure modified from (Rae et al., 2013; Yeates et al., 2010; Zang et al., 2021).

1.4.1.1. Two types of carboxysomes

Two types of carboxysomes have evolved through convergent evolution, probably after the divergence of α- and β-cyanobacteria approximately one billion years ago (Correa et al., 2025). The α-carboxysomes contain form 1A Rubisco and are predominantly found in oceanic cyanobacteria, like *Halothiobacillus neapolitanus* (*H. neapolitanus*), whereas the β-carboxysomes contain form 1B Rubisco (the type also found in plants) and are mainly present in freshwater and estuarine cyanobacteria, such as *Synechococcus elongatus* PCC 7942 (*Se*7942) and *Synechocystis* sp. PCC 6803 (*Syn*6803) as well as some filamentous and bloom-forming genera (Badger et al., 2002; Liu, 2022; Rae et al., 2013). Both types share a similar polyhedral architecture (Iancu et al., 2007; Iancu et al., 2010), but differ in size: α-carboxysomes are typically around 150 nm in diameter, whereas β-carboxysomes are larger (200-400 nm) (Rae et al., 2013). In addition to morphology, they differ fundamentally in their operon organization (*cso* operon for α-carboxysome and *ccm* operon for β-carboxysome) (Figure **1.11**), protein composition, and assembly mechanisms (Cai et al., 2015; Cameron et al., 2013; Iancu et al., 2010; Kinney et al., 2011; Rae et al., 2013; Zang et al., 2021).

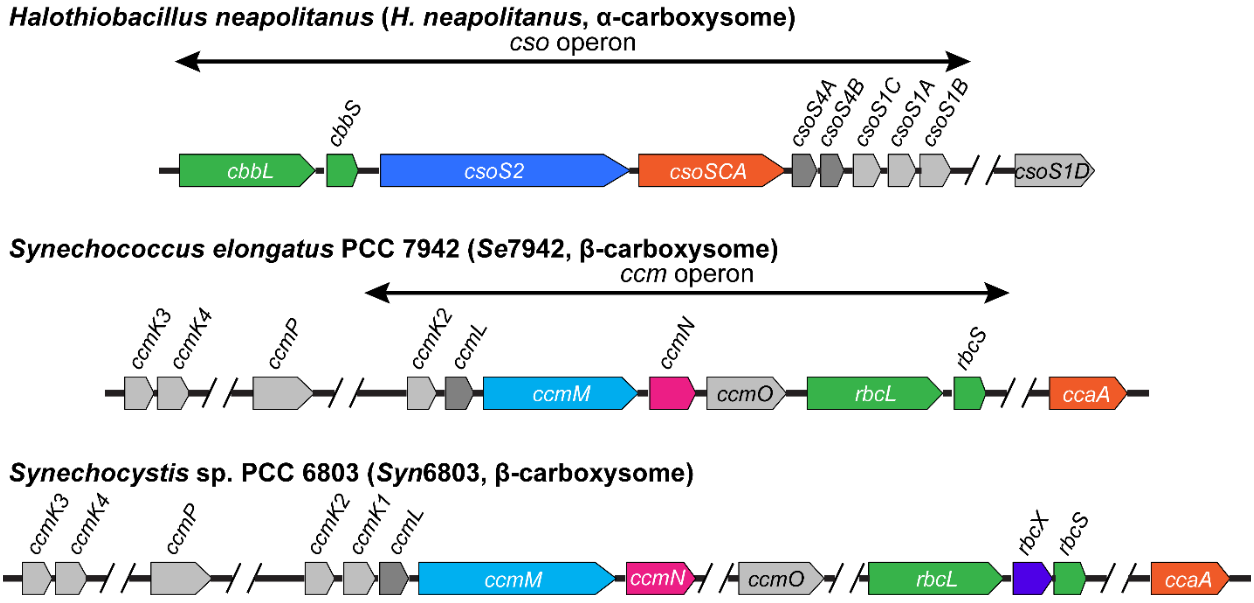


Figure 1.11 Genomic organization of representative α- and β-carboxysomal operons. Key genes of *cso* operon from *Halothiobacillus neapolitanus* (*H. neapolitanus*, α-carboxysome) are shown at the top. Next are the *ccm* operons from *Synechococcus elongatus* PCC 7942 (*Se*7942, β-carboxysome) and *Synechocystis* sp. PCC 6803 (*Syn*6803, β-carboxysome). Genes encoding structurally and/or functionally similar proteins are the same color. Figure modified from (Rae et al., 2013).

1.4.1.1.1. Components of α-carboxysomes

α-Carboxysomes typically contain 8-11 polypeptides encoded mainly by genes clustered in a single *cso* operon (Figure **1.11**). The principal proteins found inside α-carboxysomes include Rubisco (genes *cbbL* and *cbbS*), the scaffolding protein CsoS2 (CsoS2A and CsoS2B) and the carbonic anhydrase CsoSCA. These internal components are encapsulated by a selectively permeable shell formed by shell proteins CsoS1A to -E, along with CsoS4A and -B (Rae et al., 2013).

The CsoS2 proteins are essential in organizing the α-carboxysome architecture by bridging the cargo and the shell. The N-terminus of CsoS2 binds Rubisco, while the C-terminus interacts with shell proteins through multiple interfaces (Li et al., 2020; Oltrogge et al., 2020; Wang et al., 2024). Remarkably, the ribosomal frameshifting of *cis* elements in CsoS2 mRNA results in the expression of two CsoS2 isoforms: the full-length CsoS2B (~900 residues), a large intrinsically disordered polypeptide lacking structured domains, and the C-terminus-truncated CsoS2A. *In vivo* analysis of α-carboxysome biogenesis showed that only CsoS2B is capable of forming intact α-carboxysome (Chaijarasphong et al., 2016).

The CsoSCA (also known as CsoS3) in α-carboxysome is a highly divergent β-CA, which has been characterized in the chemoautotrophic bacterium *H. neapolitanus*. CsoSCA was initially identified as a novel ε-class CA, but is structurally analogous to β-class CA enzymes and contains only one active site that catalyzes the conversion of HCO_3^- to CO_2 (Heinhorst et al., 2006; Sawaya et al., 2006; So et al., 2004). Interestingly, CsoSCA is redox inactivated under reducing conditions (So et al., 2004). Structural studies show that CsoSCA interacts directly with Rubisco via an intrinsically disordered N-terminal domain, mediated by a hydrogen bonding network (Blikstad et al., 2023). Furthermore, CsoSCA and CsoS2 share an overlapping binding site on Rubisco, but utilize substantially different motifs and binding modes, highlighting the plasticity of the Rubisco binding site (Blikstad et al., 2023). These findings suggest that Rubisco functions not only as an enzyme, but also as a central hub to mediate carboxysome assembly through multiple protein-protein interactions. Interestingly, in the photoautotrophic α-cyanobacterium *Cyanobium* sp. PCC 7001 (*Cy*7001), an isoform of CsoSCA (*Cy*CsoSCA) is allosterically activated by the Rubisco substrate RuBP, constituting a feedback loop that couples CA activity to carbon fixation (Pulsford et al., 2024).

The shell proteins CsoS1A to -C share the Pfam00936 BMC domain and form hexameric structures, classifying them as BMC-H proteins (Tan et al., 2021; Tsai et al., 2007; Tsai et al., 2009). CsoS1D, encoded by a gene close to the canonical *cso* operon, contains tandem Pfam00936 domains and is classified as a BMC-T protein (Klein et al., 2009). It may facilitate the gated transport of large metabolites such as RuBP and 3-PG across the shell. CsoS1E, which has been found in most α-cyanobacteria, also has a single Pfam00936 domain (Roberts et al., 2012). Alternatively, CsoS4A and CsoS4B contain the Pfam03319 domain and belong to the family of BMC-P proteins that cap the vertices of the polyhedral shell (Wheatley et al., 2013).

Mutation of the core α-carboxysome genes generally resulted in a high CO_2-requiring (HCR) phenotype. For example, modifications in the *cbbL* gene resulted in mutants that required a CO_2 supplement for growth (Baker et al., 1998). Genetic disruption of the scaffolding protein gene *csoS2* by inserting a kanamycin resistance cassette (*csoS2*::Kan^R) resulted in loss of carboxysomes in *H. neapolitanus* cells and a HCR phenotype (Cai et al., 2015). Similarly, the deletion of the CA gene (*csoSCA*::Kan^R) resulted in carboxysomes with a size and appearance indistinguishable from those of wild-type cells, but the knockout cells grew significantly more slowly in air than wild-type cells (Dou et al., 2008). Gene knockout of either of the shell protein genes *CsoS1A* or *CsoS1C* resulted in HCR phenotypes, and furthermore, knockout of *CsoS1B* has a less pronounced effect on the growth of the strain than the knockout of either *CsoS1A* or *CsoS1C* (Desmarais et al., 2019). Knockout of *csoS1D* had too weak a phenotype to be considered HCR (Desmarais et al., 2019). In addition, loss of CsoS4A and CsoS4B also caused an HCR phenotype, probably due to CO_2 leakage (Cai et al., 2009).

1.4.1.1.2. Components of β-carboxysomes

The structural and catalytic components of β-carboxysomes in *Se*7942 are encoded by the *ccm* operon (Figure **1.11**). Besides the shell proteins (CcmK2, CcmK3, CcmK4, CcmP, CcmL, CcmO), three additional proteins are involved in the formation of β-carboxysomes: the scaffolding protein CM (CcmM/M58), the adaptor protein ApN (CcmN) and Rubisco. The genes for CM, ApN and Rubisco and the shell proteins CcmK2, CcmL and CcmO are located in the *ccm* operon, while the genes for the shell proteins CcmK3, CcmK4 and CcmP, and CA are located in distant loci (Figure **1.11**).

The proteinaceous shell of the β-carboxysome is mainly composed of three types of BMC proteins: BMC-H (CcmK1, CcmK2, CcmK3 and CcmK4), BMC-T (CcmP and CcmO) and BMC-P (CcmL). Among these, the hexameric BMC-H proteins CcmK1, CcmK2 and CcmK4 have central pores lined with positively charged lysine or arginine residues that are thought to allow the selective uptake of negatively charged molecules such as HCO_3^- (Kerfeld et al., 2005; Tsai et al., 2007). Meanwhile, CcmP forms a dimer of pseudohexamers, positioning a gated pore on either side of a central container in an airlock-like arrangement, with the open conformation of these pores larger than those of BMC-H proteins to accommodate RuBP or 3-PG. CcmK3 and CcmK4 have also been proposed to be involved in carboxysome localization and partitioning through interactions with the bacterial cytoskeleton (Rae et al., 2012).

In *Se*7942, the scaffolding protein CM exists in two isoforms. The full-length CM contains an N-terminal γ-carbonic anhydrase-like (γCAL) domain with a seven-turn β-helix, followed by three so-called SSUL (Rubisco small subunit-like) modules, and are connected by flexible linkers (Wang et al., 2019). CM trimerizes via oligomerization of its γCAL domain, and thus has a total of nine SSUL modules (Zang et al., 2021). A shorter monomeric isoform, CM^{Ct}, is generated by an internal ribosome binding site and comprises only the three SSUL modules (Long et al., 2007). Both $(CM)_3$ and CM^{Ct} are essential for functional β-carboxysome biogenesis (Long et al., 2010). $(CM)_3$ and CM^{Ct} link Rubisco complexes by interactions of binding their SSUL modules with a surface cleft between two antiparallel RbcL dimers (Wang et al., 2019), while $(CM)_3$ has a much higher affinity for Rubisco due to the increased avidity of the SSUL modules (Zang et al., 2021). $(CM)_3$ crosslinks $(CA)_4$ via γCAL domains of $(CM)_3$ interactions with the C-terminal tails of $(CA)_4$ (Zang et al., 2021). Notably, CM shows species-dependent variability. In *Syn*6803, alternative forms (e.g., CM73, CM62, CM52, and CM36) arise from variation in the number of SSUL modules.

ApN has been reported to contain a five-turn N-terminal β-helical domain consisting of ~120 amino acids, which includes six bacterial hexapeptide repeat motifs (Pfam00132) and binds the N-terminal γCAL domain of $(CM)_3$ (Cot et al., 2008; Kinney et al., 2012). This domain is followed by a flexible linker region and a highly conserved C-terminal segment of 15~20 amino acids that folds into an amphipathic α-helix known as EP (Aussignargues et al., 2015; Kinney et al., 2012). The EP interacts with shell proteins CcmK2, CcmL and CcmO to initiate the shell formation (Cameron et al., 2013; Kinney et al., 2012).

There are two types of carbonic anhydrases (CA) in β-carboxysomes. The canonical β-carboxysome CA enzyme also belongs to the β-CA class and, like CsoSCA of α-carboxysomes, is inactivated under reducing conditions, indicating that the carboxysome interior is inaccessible to reduction by thioredoxin (Rae et al., 2013). In some species lacking a dedicated CA, the γCAL domain of CM may provide CA activity (de Araujo et al., 2014; Pena et al., 2010). In species that possess both CA and CM, the two proteins may function individually or together as CAs, presumably playing complementary roles (Kimber, 2014; McGurn et al., 2016; Pena et al., 2010). It is unclear why CA activity of CA or CM is preferred over the other, but it may be that CM acts as a structural component, whereas CA is more flexible in adjusting the CA/Rubisco ratio, which is important in regulating CO_2 levels for fixation rate (Rae et al., 2013).

Not all of the β-carboxysome components are essential for carboxysome formation, but absence of these proteins generally resulted in HCR phenotypes (Price & Badger, 1989c; Rae et al., 2012; So et al., 2002; Suzuki et al., 1991). Modifications of *rbcLS* produced mutants that required either high or low levels of CO_2 for growth (Kaplan et al., 1991; Ronen-Tarazi et al., 1995). Knockout of any one of *ccmK2, ccmN* or *ccmO* results in strains that possess a single large polar carboxysome-like body, lacking the characteristic faceted genometry of carboxysomes (Figure **1.12A-D**) (Cameron et al., 2013; Rae et al., 2012). In contrast, deletion of *ccmL* leads to the formation of elongated rod-like structures alongside some apparently normal looking carboxysomes (Figure **1.12E**) (Cameron et al., 2013; Price & Badger, 1989c), while loss of *ccmM* completely abolishes carboxysome formation (Figure **1.12F**) (Long et al., 2007; Long et al., 2010; Rae et al., 2012). Additionally, strains lacking each of *ccmK2, ccmL, ccmM, ccmN*, or *ccmO* exhibit HCR phenotypes (Cameron et al., 2013; Kinney et al., 2012; Marco et al., 1994; Price et al., 1993; Rae et al., 2012). Deleting *ccmK3* or *ccmK4* alone does not affect carboxysome ultrastructure or physiology (Figure **1.12G-H**). However, a tandem deletion of *ccmK3-4* results in cells with a regular-shaped icosahedral carboxysome structure (Figure **1.12I**), but physiologically deficient at low CO_2 conditions (Rae et al., 2012). Similarly, inactivation mutants of *ccaA* produce carboxysomes with similar size, shape and abundance in cells compared to the wild-type strain (Figure **1.12J**), but they are unable to grow autotrophically at air levels of CO_2 (Price & Badger, 1989c; So et al., 2002).

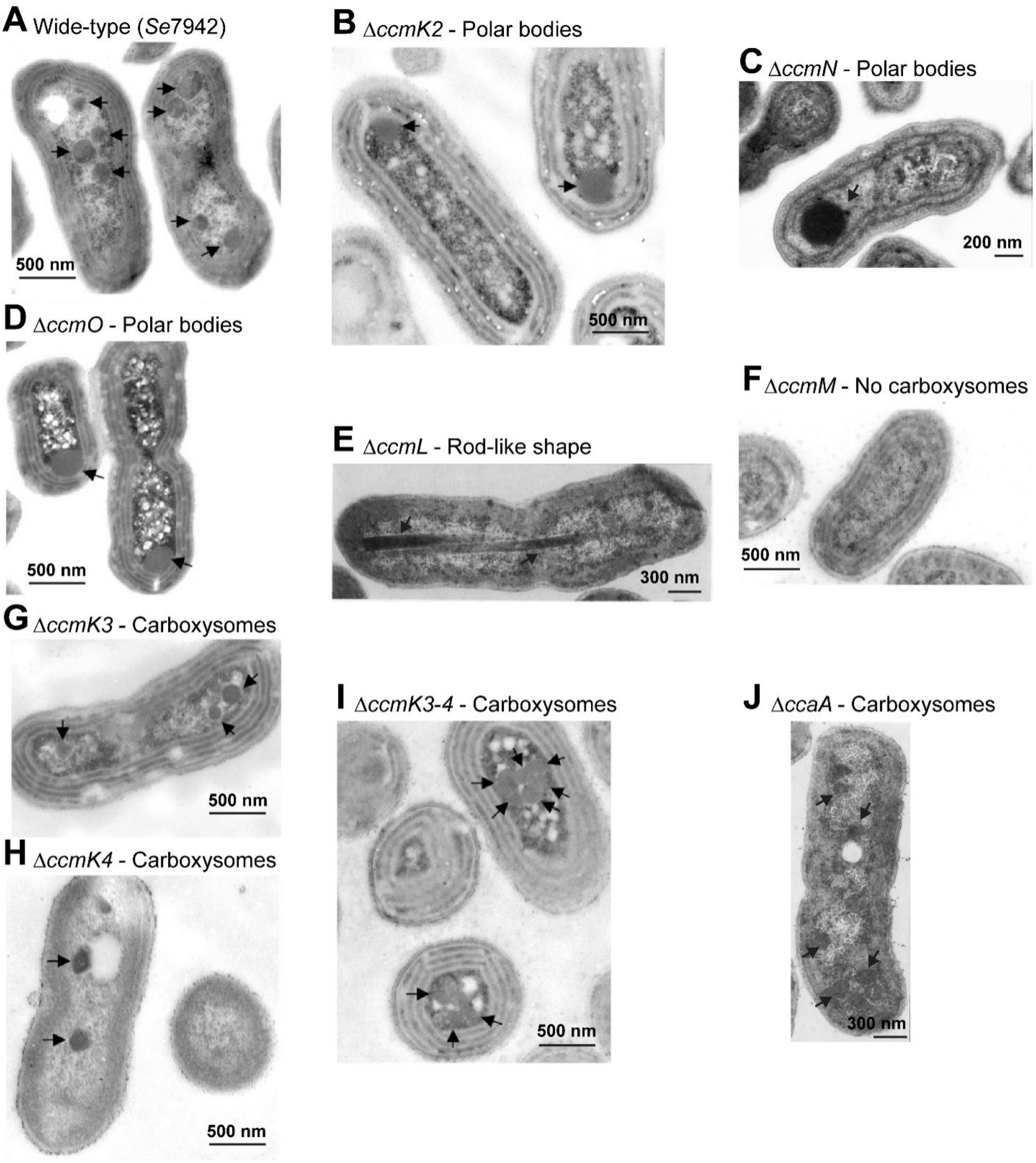


Figure 1.12 Ultrastructural carboxysome phenotypes in wild-type and BMC mutant cells of *Se*7942.

Representative transmission electron micrographs of wild-type and mutant strains are shown (black arrows indicate the structures). **(A)** Wild-type *Se*7942. **(B-D)** *Se*7942 mutants lacking carboxysomes but containing a single polar protein aggregate: **(B)** Δ*ccmK2*. **(C)** Δ*ccmN*. **(D)** Δ*ccmO*. **(E)** *Se*7942 Δ*ccmL*, showing two long, rod-like structures in the center of the cell. **(F)** *Se*7942 Δ*ccmM*, with no detectable carboxysomes. **(G)** *Se*7942 Δ*ccmK3*. **(H)** *Se*7942 Δ*ccmK4*. **(I)** *Se*7942 Δ*ccmK3-4*. **(J)** *Se*7942 Δ*ccaA*. Scale bars are indicated. Figure modified from (Kinney et al., 2012; Price & Badger, 1989c; Rae et al., 2012).

1.4.1.2. Structural characterization of carboxysomes

The BMC-H (CcmK1 and CcmK2), BMC-T (CcmO) and BMC-P (CcmL) shell proteins from *Halothece* sp. PCC 7418 (*Ha*7418) upon recombinant expression in *Escherichia coli* (*E. coli*) resulted in the self-assembly of carboxysomal shells. Cryo-electron microscopy (cryo-EM) single-particle analysis revealed three different structural classes with a resolution of 2.6 Å (a pseudo-icosahedral T = 4 shell), 3 Å (a pseudo-icosahedral T = 3 shell), and 4.1 Å (an elongated shell), respectively (Sutter et al., 2019). The majority (75%) adopted a pseudo-icosahedral T = 4 structure with a diameter of 245 Å, a smaller population (3.4%) formed a pseudo-icosahedral T = 3 structure (210 Å), and 4.5% assembled into elongated shells with dimensions of 245 × 310 Å. Interestingly, only BMC-H and BMC-P proteins were present in the assembled shells, suggesting that internal proteins may also contribute to the formation of an intact β-carboxysome shell (Sutter et al., 2019). Similarly, co-expression of *H. neapolitanus* α-carboxysome BMC-H protein (CsoS1A), BMC-P protein (CsoS4A) with or without BMC-T protein (CsoS1D) also produced the T = 3 form of shell-like particles that closely resembled β-carboxysome shells, with CsoS1D also absent from the final structure. When CsoS1A and CsoS4A were co-expressed with an N-terminal GFP-tagged 30-residue C-terminal peptide of CsoS2 (GFP-S2-CTP), two shell sizes were observed: T = 3 (~217 Å) and T = 4 (~247 Å), with proportions of 14% and 86%, respectively. These results suggested that encapsulated scaffold molecules may increase shell size (Tan et al., 2021). In addition, in presence of only CsoS1A and CsoS4A shell proteins from *H. neapolitanus* produced predominantly small T = 3 shells (~21 nm), with a low abundance of medium-sized T = 4 shells (~23 nm) (Ni et al., 2023). Upon inclusion of the internalized scaffolding proteins CsoS2A/B, most shells adopted a larger T = 9 configuration (~35 nm) containing CsoS2B, alongside some medium-sized T = 4 shells (~23 nm) that failed to incorporate CsoS2A/B. Although only three C-terminal fragments of CsoS2B (Arg712-Arg731, Leu773-Gly823, Glu829-Gly869) were resolved, these findings reveal that CsoS2B binds to the shell proteins through multivalent interactions, with a highly conserved and repetitive Ile(Val)-Thr-Gly ([IV]TG) motif of CsoS2B identified at the CsoS1A-CsoS2B interfaces, suggesting that CsoS2B acts as a molecular thread to facilitate α-carboxysome shell assembly (Ni et al., 2023).

The low-resolution cryo-EM maps of an intact α-carboxysome from *Cy*7001 enabled the construction of a structural model for both the carboxysome shell and its internal enzyme organization, demonstrating that Rubisco is arranged in concentric layers with two main forms of Rubisco-Rubisco interface: side-by-side and top-to-bottom (Evans et al., 2023). Recently, the structure of intact α-carboxysomes from the marine cyanobacterium *Prochlorococcus* MED4 was solved using block-based reconstruction of the cryo-EM single-particles, revealing the architecture and assembly pattern of the shell, the three-layered internal organization of Rubisco, and the multivalent interactions between the middle and C-terminal regions of CsoS2 with shell capsomers and the N-terminal regions of CsoS2 with Rubisco (Zhou et al., 2024).

Cryo-electron tomography (cryo-ET) combined with subtomogram averaging (STA) of Rubisco within intact, native α-carboxysomes revealed two different organizational arrangements of the enzyme. In *H. neapolitanus*, luminal Rubisco assembles into higher-order fibrils with six-fold pseudosymmetry, which is mediated by its small subunits, CbbS (Metskas et al., 2022). However, it is disputable whether this arrangement requires CsoS2, since extra density of CsoS2 was only observed near Rubisco in proximity to the carboxysome shell (Ni et al., 2022). In contrast, CsoS2 scaffolds Rubisco throughout the lumen and shell of *Cy*7001 carboxysome, in which Rubiscos are aligned in a radial direction with their 4-fold axes roughly pointing to the center of the carboxysome (Ni et al., 2022). Unlike α-carboxysomes, cryo-ET and STA analyses showed that *Se*7942 β-carboxysomes are rather heterogeneous in shape and size, with interiors densely packed with Rubiscos that are generally arranged in a lattice-like pattern, but with their 4-fold axes pointing to various directions (Kong et al., 2024). Three prominent configurations were observed between adjacent Rubiscos: head-to-head, head-to-side, and side-by-side (Kong et al., 2024). Rubiscos in the outermost layer are regularly aligned along the shell facets, with ~30% probably forming direct interactions with the shell (Kong et al., 2024). In contrast, in *Se*7942 grown under low-light conditions, Rubiscos are mostly oriented with their 4-fold axes along the radial direction across all layers, with no significant difference in the orientation of Rubisco within the outermost layer compared to the inner layers, indicating structural plasticity of β-carboxysomes under different environmental conditions (Sun et al., 2024).

1.4.1.3. Biogenesis pathways of α- and β-carboxysomes

Cyanobacterial carboxysomes are a potential target for genetic engineering into plant chloroplasts to improve photosynthetic CO_2 assimilation efficiency with the goal to increase crop yields (Nguyen et al., 2024; Occhialini et al., 2016). Understanding the mechanism of carboxysome biogenesis would aid in bioengineering efforts. Current evidence suggests that

Rubisco condensation in α-carboxysomes is driven by CsoS2B, which forms multivalent interactions with Rubisco via its N-terminus while binding shell proteins through its C-terminus (Figure **1.13**). These structures assemble via a ‘partial shell first’ or ‘concomitant shell-core’ assembly pathway (Dai et al., 2018; Kerfeld et al., 2018; Liu, 2022), whereby the shell forms independently and/or simultaneously with Rubisco encapsulation (Iancu et al., 2010; Liu, 2022). In contrast, the assembly of β-carboxysomes begins with the condensation of Rubisco, facilitated by the SSUL modules of the scaffolding protein CM and its isoform CM^{Ct} (Wang et al., 2019; Zang et al., 2021), and proceeds via an ‘inside-out’ mechanism in which the adaptor ApN initiates shell formation around the pro-carboxysome condensate containing Rubisco and carbonic anhydrase (Cameron et al., 2013; Gonzalez-Esquer et al., 2015; Zang et al., 2021). This process is driven by liquid-liquid phase separation and multiple protein-protein interactions of the core components. Nevertheless, how carboxysomes assemble and mature during biogenesis is still not fully understood. In this project, we will explore the stepwise maturation process of pro-β-carboxysomes.

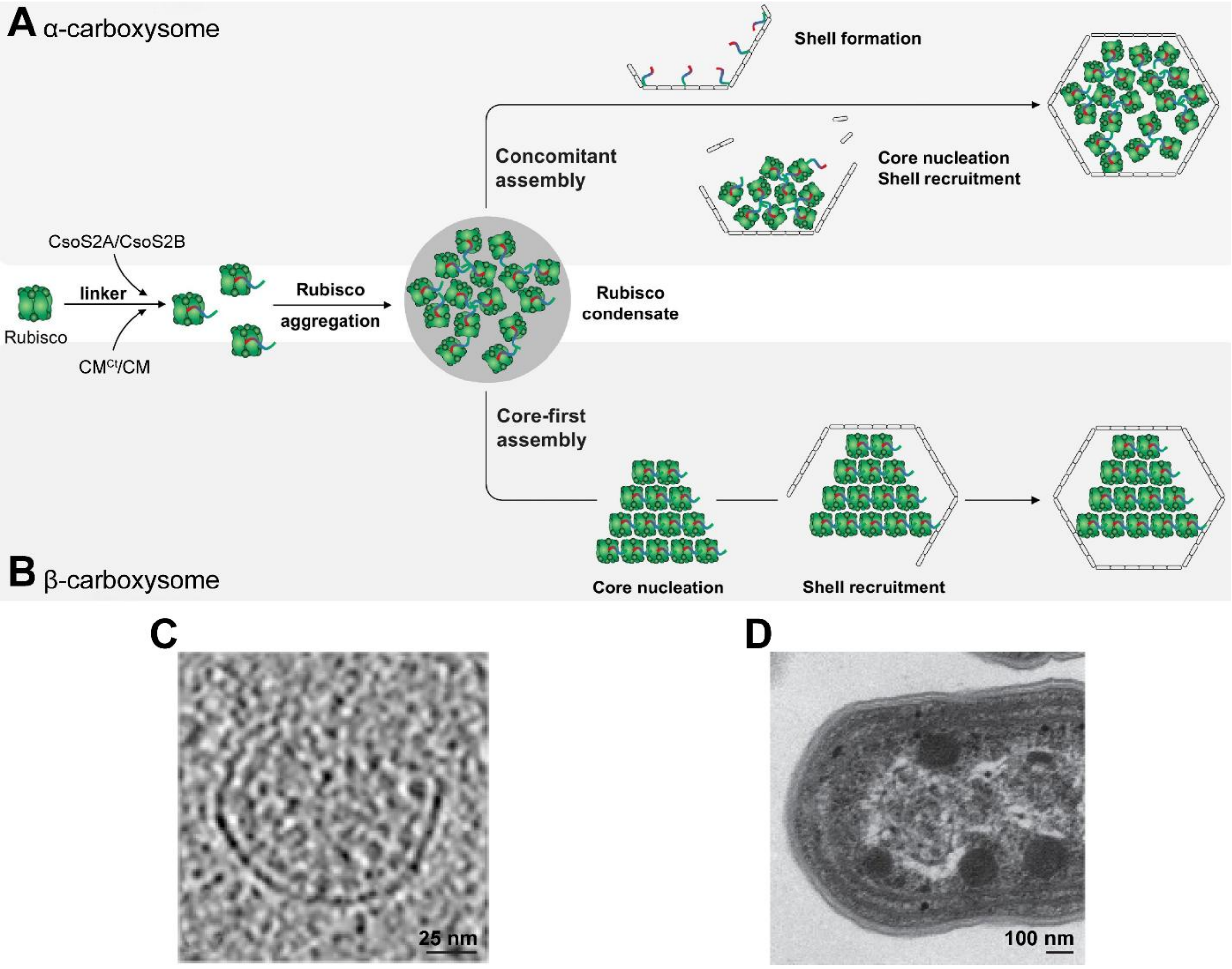


Figure 1.13 Proposed assembly pathways of carboxysomes.
(A) Concomitant shell and core assembly: the carboxysome shell assembles simultaneously with the aggregation of Rubisco. **(B)** Core-first assembly: Rubisco condensation forms the densely packed enzyme core first, followed by the shell recruitment and encapsulation. **(C-D)** Representative micrographs of **(C)** a partial α-carboxysome of *H. neapolitanus*, and **(D)** intact β-carboxysomes in cyanobacteria. Scale bars are indicated. Figure adapted from (Iancu et al., 2010; Kerfeld & Erbilgin, 2015; Liu, 2022).

1.4.1.4. Redox regulation of carboxysome function

Upon shell formation, the internal environment of the β-carboxysome becomes oxidizing (Chen et al., 2013). The semipermeable shell selectively restricts the entry of small proteins (~40 Å in diameter), such as thioredoxin and glutaredoxin, as well as other endogenous reducing agents, through the formidable barrier of small pores (~6 Å in diameter) of shell proteins, or up to 14 Å in the case of gated pores (Pena et al., 2010). The thioredoxins in the cytoplasm could reduce redox damage to carboxysome components during carboxysome biogenesis and maturation (Lindahl & Florencio, 2003; Mata-Cabana et al., 2007; Rae et al., 2013). Additionally, independent redox modulation within the carboxysome lumen is thought to be vital for the activity of its encapsulated enzymes (Sun et al., 2016). Notably, the CA enzyme is inactive in the reducing environment of the cytosol and is activated only in the oxidizing environment within the carboxysome, generating CO_2 where it is needed (Pena et al., 2010; Price et al., 1992). Furthermore, in the absence of redox regulation within the carboxysome, disulfide bond formation in the SSUL modules of CM^{Ct} and $(CM)_3$ decreases the affinity of the scaffolding proteins for Rubisco (Wang et al., 2019; Zang et al., 2021), in contrast to the affinity in the reducing environment of the cytosol during pro-carboxysome condensate formation (Wang et al., 2019; Zang et al., 2021). Biochemical evidence suggests that α-carboxysomes may mature through a similar oxidation process (Carpenter et al., 2022; Heinhorst et al., 2006), but the relationship between affinity and the chemical environment remains unknown (Metskas et al., 2022).

1.4.1.5. Key advances in carboxysomes construction in plant chloroplasts

When a simplified set of the α-carboxysome proteins of *Cy*7001, including Rubisco, the linker protein CsoS2 and the shell protein CsoS1A, were co-expressed in *Nicotiana tabacum* (*N. tabacum*), they were capable of self-organizing to form structures similar to α-carboxysomes, while some long rod-shaped carboxysomes were also observed (Long et al., 2018). In another study, when almost all α-carboxysome components of *H. neapolitanus* (including CbbL, CbbS, CsoS2, CsoSCA, CsoS1A/B/C and CsoS4A/B) were introduced into *N. tabacum*, formation of somewhat elongated functional α-carboxysomes was observed (Chen, Hojka, et al., 2023). However, these transplastomic plants exhibited an HCR phenotype. These results imply that engineering fully functional CO_2-fixing modules and entire CCMs into plants to improve crop photosynthesis and productivity requires better understanding of the protein stoichiometry for BMCs assembly and functionality, the introduction of bicarbonate transporters in the chloroplast inner membrane, and the elimination of endogenous chloroplastic CAs to prevent the dehydration of HCO_3^- and diffusional loss of CO_2 from the stroma prior to entry into the carboxysome (Nguyen et al., 2024; Price & Badger, 1989b).

1.4.2. Pyrenoids

Nearly all eukaryotic algae and some nonvascular plants (including most hornwort species) concentrate Rubisco within chloroplast-associated microcompartments called pyrenoids (Wang et al., 2015), which are estimated to account for approximately one third of global photosynthetic CO_2 capture (Catherall et al., 2025). While recent research has extended to diatoms, hornworts, and other algae, the green algal species *Chlamydomonas reinhardtii* (*C. reinhardtii*; *Cr*) remains the most extensively studied model organism. The *C. reinhardtii* pyrenoid has three characteristic structural features: a spheroidal liquid-like matrix of Rubisco and a protein named Essential Pyrenoid Component 1 (EPYC1) formed through phase separation, which is driven by the multivalent interactions between RbcS subunits and five Rubisco binding motifs (RBMs) in the intrinsically disordered repeat linker protein EPYC1 (Freeman Rosenzweig et al., 2017; He et al., 2020; Wunder et al., 2018); pyrenoid tubules, which are specialized tubular thylakoids that traverse the matrix (Engel et al., 2015); and a starch sheath that surrounds the matrix (Itakura et al., 2019; Toyokawa et al., 2020). HCO_3^- is transported from the stroma into the pyrenoid tubules, where the carbonic anhydrase 3 (CAH3) converts it to CO_2 (Blanco-Rivero et al., 2012; Sinetova et al., 2012). The starch sheath potentially limits the diffusion of CO_2 away from the matrix, thereby concentrating it in the surrounding pyrenoid matrix. Together, these features are thought to increase the CO_2 fixation rate by up to threefold (Fei et al., 2022). Recent advances have been made in assembling a pyrenoid-like structure in the model plant *Arabidopsis thaliana* (*A. thaliana*; *At*). The expression of *Cr*EPYC1 and a plant-algal hybrid Rubisco (comprising *At*RbcL and *Cr*RbcS) has generated an algal-like ‘proto-pyrenoid’ matrix in *A. thaliana* chloroplasts via liquid-liquid phase separation (Atkinson et al., 2020). Further progress was made by introducing the *C. reinhardtii* StArch Granules Abnormal proteins, SAGA1 and SAGA2, into *A. thaliana*, which resulted in atypical spherical starch granules being recruited to surround and partially cover the proto-pyrenoid condensate (Atkinson et al., 2024). Furthermore, expressing SAGA1 and MIssing THylakoids 1 (MITH1) together in *A. thaliana* has been shown to result in the formation of matrix-traversing thylakoid membranes (Hennacy et al., 2024). In order to improve CO_2 fixation by synthetic pyrenoids in plants, efforts will focus on combining the starch sheath with the matrix-traversing membrane, as well as engineering bicarbonate transporters and carbonic anhydrases to facilitate the delivery of CO_2 through the chloroplast membranes (Fei et al., 2022).

1.4.3. C_4 photosynthesis and crassulacean acid metabolism (CAM) photosynthesis in plants

C_4 photosynthesis and crassulacean acid metabolism (CAM) are two specialized forms of photosynthesis that have evolved in plants to optimize photosynthetic efficiency and improve water use efficiency under challenging conditions, particularly in hot and dry environments (Cui et al., 2025). C_4 plants spatially separate initial CO_2 fixation and CBB cycle in different cell types. They capture CO_2 in their mesophyll cells, where the enzyme phosphoenolpyruvate carboxylase catalyzes the combination of CO_2 and phosphoenolpyruvate into a four-carbon compound, oxaloacetate (OAA) (Kellogg, 2013). OAA is then converted to malate or aspartate and transported to bundle-sheath cells, where CO_2 is enzymatically regenerated from malate or aspartate, creating a high concentration of CO_2 around Rubisco (Kellogg, 2013). This process can increase the CO_2 concentration near Rubisco by up to 10-fold above ambient levels (von Caemmerer & Furbank, 2016). As a result, C_4 plants, such as maize, sorghum and sugarcane, have higher photosynthetic efficiency, faster growth rates and higher water use efficiency than C_3 plants (e.g., rice, wheat, soybean). Unlike C_4 plants, CAM photosynthesis, found in succulents (e.g., cacti, agave, pineapple) and some epiphytes (e.g., orchids), separates CO_2 fixation temporally rather than spatially. CAM plants close their stomata during the day to minimize water loss via evaporation and open their stomata at night to take in CO_2. The CO_2 is fixed by phosphoenolpyruvate carboxylase into carboxylic acids, such as malate, citrate, or isocitrate

(Topfer et al., 2020). These organic acids are stored in vacuoles and released to the photosynthesizing cells during the day while the stomata remain closed (Topfer et al., 2020). This adaptation minimizes water loss and prevents photorespiration by concentrating CO_2 internally, allowing CAM plants to thrive in extremely dry conditions.

1.5. Liquid-liquid phase separation (LLPS)

Cells confine and organize complex biochemical reactions spatiotemporally within specialized membrane-bound organelles, or membraneless compartments. Membrane-bound compartments are impermeable to most biological molecules and thus physically separate their interior from the exterior, such as endoplasmic reticulum, Golgi apparatus, mitochondria and lysosomes (Banani et al., 2017; Friedman & Nunnari, 2014; Luzio et al., 2007). Membraneless compartments, on the other hand, which include Processing granules (P granules), nucleoli, centrosomes, Cajal bodies, and stress granules, lack a physical barrier to separate their internal components from the surrounding medium, allowing their components to diffuse freely in and out (Brangwynne et al., 2009; Hyman et al., 2014). However, unlike membrane-bound compartments, whose homeostasis is maintained by the out-of-equilibrium action of a variety of receptors, channels, pumps, and motors embedded in organellar membranes, the concentration homeostasis in membraneless compartments is established due to attractive interactions among molecules and the solution entropy of the phase-separating components (Deviri & Safran, 2021). Liquid-liquid phase separation (LLPS) is a universal mechanism that organizes biomolecules into dynamic, membraneless compartments, also known as biomolecular condensates. These condensates regulate critical cellular processes, such as genome organization, immune response, neuronal function, and disease pathology in mammalian and other higher eukaryotic cells, as well as cellular dynamics and organization, stress response, and pathogenic mechanisms in prokaryotes (Alberti, 2017; Alberti et al., 2019; Che et al., 2023; Yusuf et al., 2025).

1.5.1. Multivalent interactions involved in LLPS

Condensate components are broadly divided into two classes: (1) ‘scaffolds’, which are resident molecules that are essential for condensate formation, such as PML for PML nuclear bodies, Spd5 for *Caenorhabditis elegans* (*C. elegans*) centrosomes and TIA1 for stress granules; (2) ‘clients’, which consist of molecules that are dispensable for condensate assembly, but are major components localized to condensates by direct binding to scaffolds (Banani et al., 2017), or that are selectively recruited into condensates after their formation, making the condensates compositionally malleable (Holehouse & Pappu, 2018). These biomolecules (proteins/nucleic acids) undergo LLPS driven by thermodynamic forces, behaving like polymers in solution and separating into a single condensed phase (condensate that often resembles liquid-like droplets) and a more dilute surrounding phase (Brangwynne et al., 2009; Deviri & Safran, 2021; Guillen-Boixet et al., 2020). LLPS is mediated by two types of multivalent interactions: (i) conventional multivalent interactions between proteins and/or RNAs (e.g., protein-protein, protein-RNA, or RNA-RNA interactions), and (ii) non-specific weak interactions (charge-charge interactions, π-π stacking, cation-π interactions, hydrophobic contacts, and hydrogen bonds) between amino acids in intrinsically disordered proteins (IDPs) and other biomolecules, or multivalent binding of folded domains that are tethered to intrinsically disordered regions (IDRs) (Dignon et al., 2020; Wang et al., 2021). The phase behavior of condensate-forming molecules is influenced by the overall net charge and the spatial distribution of charged residues. Aromatic residues, tryptophan (Trp), tyrosine (Tyr), and phenylalanine (Phe), serve as primary stickers that form reversible non-covalent crosslinks (Bremer et al., 2022; Ginell & Holehouse, 2023). Among these residues, Trp is the strongest aromatic sticker, followed by Tyr and Phe (Dzuricky et al., 2020; Ginell & Holehouse, 2023; Wang et al., 2018). Similarly, arginine (Arg) has been shown to be a substantially stronger driver of cation-π interactions than lysine (Lys) (Fisher & Elbaum-Garfinkle, 2020; Ginell & Holehouse, 2023; Wang et al., 2018).

The stickers-and-spacers model has been widely used to understand various biomolecular assembly behavior, offering a useful framework in which every residue, subregion, or domain can be designated as a sticker or a spacer (Figure **1.14**) (Ginell & Holehouse, 2023). Stickers and spacers could be individual residues or local subregions within a flexible disordered region (Pak et al., 2016; Wang et al., 2018; Yang et al., 2019) (Figure **1.14A)**, or stickers could be folded domains, while spacers could be flexible disordered linkers (Harmon et al., 2017; Li et al., 2012) (Figure **1.14B**). For a folded domain, stickers could be adhesive patches, while spacers would be non-adhesive regions, akin to patchy-colloids (Ginell & Holehouse, 2023; Sanders et al., 2020). Stickers represent regions that mediate attractive intermolecular interactions and govern the concentration threshold for assembly by forming reversible, non-covalent physical crosslinks that control the formation of stoichiometric or nonstoichiometric complexes (Ginell & Holehouse, 2023). In contrast, spacers are interspersed between stickers and act as flexible connectors that scaffold the stickers, engendering flexibility and conformational heterogeneity, while influencing polymer solubility, determining intra-condensate topology, and tuning the cooperativity of sticker-mediated interactions (Ginell & Holehouse, 2023).

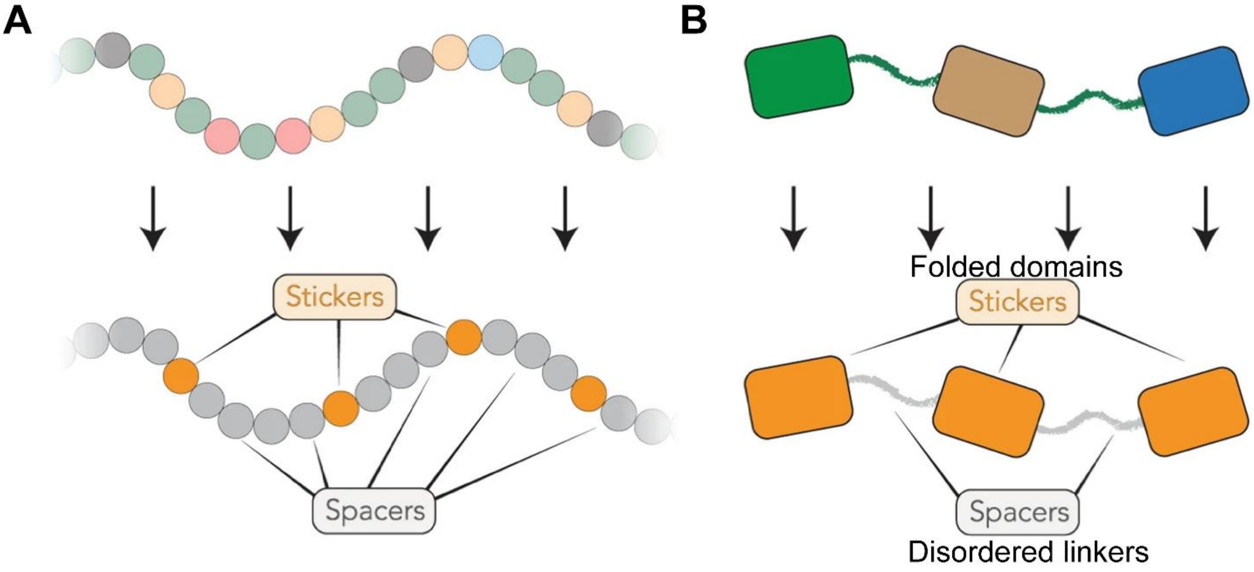


Figure 1.14 Schematic illustration of sticker-spacer mapping across different protein architectures.
(A) An intrinsically disordered region (IDR) can be mapped onto a stickers-and spacers model. In this model, adhesive residues (stickers) mediate intermolecular interactions and are interspersed with flexible non-adhesive spacers. Different colored beads represent amino acids with distinct physico-chemical properties. The relative proportion of stickers and spacers, the strength of sticker-mediated interactions, and the intrinsic dynamics of the spacers contribute to the material properties of a condensate. **(B)** Sticker-spacer mapping for a multidomain protein, in which folded domains function as stickers and are connected by flexible intrinsically disordered linkers that serve as spacers. Figure adapted from (Ginell & Holehouse, 2023).

Phase separation can occur through homotypic interactions, in which a single type of biomolecule drives segregative LLPS, excluding most of the other solutes from the biomolecular condensates, or through heterotypic interactions, in which two or more different biomolecules undergo associative LLPS and cooperatively form biomolecular condensates, as observed in the complex coacervation within carboxysomes and pyrenoids (Figure **1.15**) (Deviri & Safran, 2021). Alternatively, combinations of homotypic and heterotypic interactions can give rise to diverse types of LLPS, leading to a rich variety of biomolecular condensates that may themselves consist of multiple phases, such as the layered nucleolus (Deviri & Safran, 2021; Feric et al., 2016).

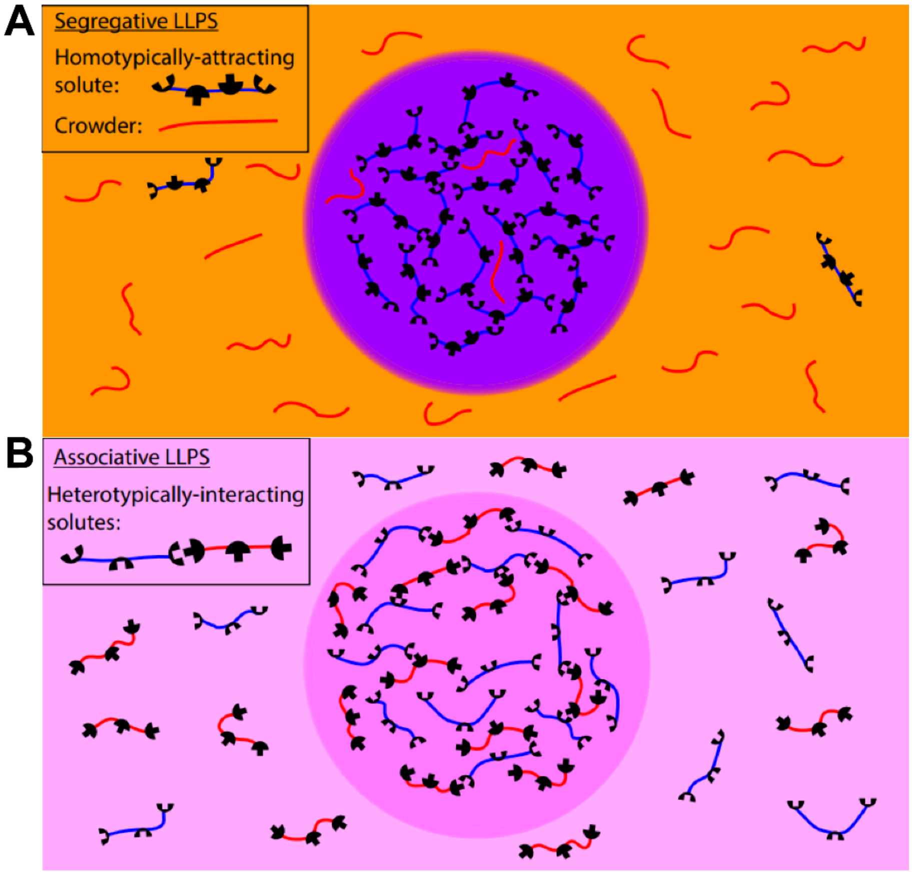


Figure 1.15 Illustration of segregative and associative LLPS.
(A) A homotypically attractive solute (blue) and an inert "crowder" (red) undergo segregative LLPS, forming one phase enriched in the attractive solute but poor in the crowder (purple), and another phase enriched in the crowder but poor in the attractive solute (light orange). **(B)** Two types of solute molecules (red and blue) engage in heterotypical interactions, resulting in associative LLPS and the formation of a dense phase (light purple) enriched in both solutes that coexist with a dilute phase (light pink) with low concentrations of both solutes. Figure adapted from (Deviri & Safran, 2021).

Beyond molecular interactions, LLPS dynamics are sensitive to environmental and chemical cues (Chen et al., 2020; Iserman et al., 2020; Krainer et al., 2021). In addition to the concentration and identities of biomolecules, LLPS can be tuned (either promoted or disrupted) by different stimuli, including temperature, salt type and concentration, co-solutes, pH and the volume excluded by other biomolecules (Alberti et al., 2019) (Figure **1.16**). In special cases, additional factors such as ATP, poly ADP-ribose, small molecules such as 1,6-hexanediol and chemical chaperones, and large molecular crowders such as polyethylene glycol, and oxidation can also influence LLPS (Dignon et al., 2020).

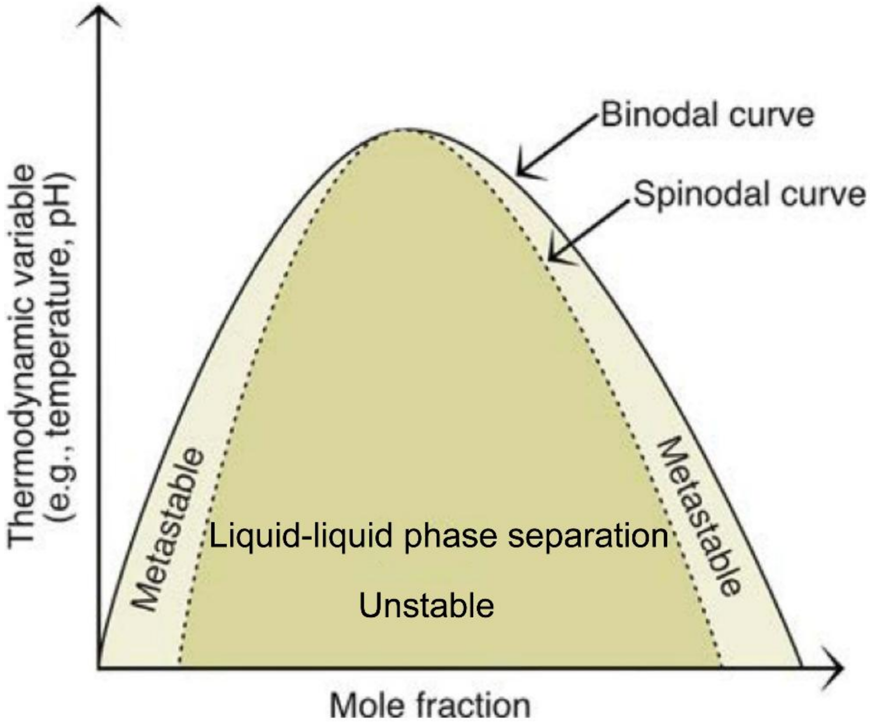


Figure 1.16 Schematic diagrams illustrating the thermodynamic conditions for LLPS. Binodal and spinodal curves define distinct phase behaviors in a mixture and are typically represented in phase diagrams that plot protein concentration against a variable, such as pH, temperature, or salt concentration. The binodal curve marks the boundary of thermodynamic equilibrium between the single-phase state (outside the curve) and two-phase state (inside the curve), where distinct liquid phases coexist. Inside this boundary, compositions separate via nucleation and growth, particularly in the metastable state between the binodal and spinodal curves. Inside the spinodal curve, the system becomes thermodynamically unstable, and phase separation proceeds spontaneously via spinodal decomposition. Figure modified from (Catherall et al., 2025).

1.5.2. Implications of LLPS in cellular compartmentalization

LLPS has several advantages compared with membrane-bound compartments. The phase boundary can provide an energy barrier that selectively partitions certain biomolecules, concentrating specific components while excluding others, without requiring physical membranes. This selective concentration increases the local effective concentration of enzymes and associated substrates, enhancing the efficiency and specificity of biochemical reactions, such as the catalysis of CO_2 fixation of Rubisco in carboxysomes and pyrenoids. Additionally, the phase-separated state may facilitate quality control by enabling the selective recruitment of functional Rubisco and accessory proteins while preventing premature or disorganized encapsulation. The liquid-like properties of biomolecular condensates promote biochemical specificity, facilitate efficient substrate diffusion, buffer intracellular concentrations of molecules, and enable rapid response to environmental fluctuations. Moreover, the reversible assembly and disassembly of membraneless organelles could provide a convenient mechanism for partitioning their components during cell division (Holehouse & Pappu, 2018). Some condensates, such as *C. elegans* P granules, undergo asymmetric inheritance by preferentially condensing at one end of the cell before cell division (Brangwynne et al., 2009; Saha et al., 2016). In contrast, some condensates, like pyrenoids in *C. reinhardtii*, disassemble prior to mitosis, enabling symmetric inheritance by evenly distributing constitutive components to daughter cells (Freeman Rosenzweig et al., 2017). However, the absence of a physical membrane compromises the long-term stability and storage within condensates, and limits the retention of very small molecules, such as ions, thereby hindering the maintenance of a stable pH gradient (Banani et al., 2017).

1.5.3. Methods to study biomolecular condensates

Several tools have been developed to study the physico-chemical properties and underlying molecular interactions of biomolecular condensates. Microscopy is an essential tool to visualize the structure and composition of biomolecular condensates, particularly with the advanced confocal microscopy and super-resolution imaging, but these methods usually require fluorescent labeling or antibody staining. Electron microscopy, especially cryo-ET, allows for the precise visualization of biomolecular condensates without labels. A common approach to dissect the molecular mechanisms underlying LLPS is to combine microscopy with *in vitro* reconstitution assays using purified proteins (Wang et al., 2021). Imaging-based techniques, such as fluorescence recovery after photobleaching (FRAP), fluorescence loss in photobleaching (FLIP), and fluorescence correlation spectroscopy (FCS), are used to evaluate the material properties and dynamics of biomolecular condensates (Ishikawa-Ankerhold et al., 2012). Additionally, some optogenetic tools have been developed to achieve temporal control over the formation and dissolution of biomolecular condensates *in vivo* by activating chimeric proteins with light or small molecules (Reinkemeier & Lemke, 2021). One such system fuses the *A. thaliana* protein domain Cry2, which oligomerizes upon blue-light activation, to the IDRs of target proteins in order to regulate multivalency using blue light (Shin et al., 2017). Optogenetic systems can also be modified to gain spatial control over the formation of condensates by fusing core proteins to proteins, such as telomeric repeat-binding factor TRF1 or catalytically dead Cas9, which binds specific genomic loci and, upon light-triggered oligomerization, localize to these sites to induce phase separation at defined genomic regions (Shin et al., 2018).

1.6. Aim of the study

Introducing a cyanobacterial carboxysome-like CCM into crop plant chloroplasts is predicted to significantly improve photosynthetic efficiency (Hines et al., 2021; McGrath & Long, 2014; Price, Badger, et al., 2011; Weerasooriya et al., 2022; Wu et al., 2023). This advancement could ultimately increase crop yields and help to address the challenges to global food security posed by the ever-growing global population and the impact of climate change (Hayer-Hartl & Hartl, 2020). However, understanding carboxysome biogenesis is essential for successfully transplanting these structures into plants. β-Carboxysomes, which contain form IB Rubisco, a homolog of plant form IB Rubisco, are of particular interest in this context. Here, we focus on investigating the biogenesis pathway of β-carboxysomes. Recent biochemical and structural studies suggest that β-carboxysomes assemble via an ‘inside-out’ mechanism, in which Rubisco forms condensates via interaction with the SSUL modules of the scaffolding protein $(CM)_3$ and its shorter isoform CM^{Ct} before formation of the proteinaceous shell (Cameron et al., 2013; Chen et al., 2013). However, the distinct steps of this stepwise assembly process remain incompletely understood.

In this project, we focus on the well-studied model organism *Se*7942, using the minimum set of genes to reconstitute *in vitro* the pro-β-carboxysome condensate that forms prior to shell formation. Our goal is to elucidate the stepwise mechanism of pro-β-carboxysome maturation and determine when and how the shell adaptor protein ApN participates in the process. We obtained the soluble ApN protein by overexpressing ApN in *E. coli*, and confirmed its oligomeric state in solution using SEC-MALS analysis and cryo-EM. Pull-down and sedimentation assays, as well as fluorescence microscopy-based condensate formation analysis combined with *in vivo* mutational analysis revealed that ApN alone is not the functional unit. Recombinant expression of CM and ApN resulted in a hetero-tetrameric complex consisting of one protomer of CM and three protomers of ApN, as confirmed by SEC-MALS. We also used cryo-EM to obtain structural insight into this hetero-tetrameric $(ApN)_3$:CM complex, although structure determination proved difficult due to its structural lability at the air-water interface during cryo-EM sample preparation. Moreover, the role of $(ApN)_3$:CM in carboxysome maturation was confirmed by *in vivo* complementation assay (in collaboration with Benedict M. Long, New South Wales and G. Dean Price, Australian Capital Territory). Importantly, using *in vitro* reconstitution we demonstrated that this hetero-complex is recruited to the pro-carboxysome condensate through multiple protein-protein interactions, including the interaction of C-terminal SSUL modules of CM protomer with N-terminal γCAL domain of adjacent CM protomers, the interaction of SSUL modules with Rubisco, and the interaction of the γCAL domain with the C-terminal tail of $(CA)_4$. We found that pro-β-carboxysome condensate formation proceeds according to the following sequence of steps: First, Rubisco forms a densely packed enzyme core mediated by the abundant scaffolding proteins CM^{Ct} and $(CM)_3$. Then, CA is incorporated before or simultaneously with $(ApN)_3$:CM to the rim of the initial Rubisco/CM^{Ct}/$(CM)_3$ condensate. Presence of $(ApN)_3$:CM will then mediate shell protein assembly.

2. Materials and methods

2.1. Materials

2.1.1. Chemicals and reagents

Table 2.1 Chemicals and reagents

Chemicals or reagents	**Source**	**Identifier**
Acrylamide/Bis Solution, 37.5:1 (30% w/v)	SERVA Electrophoresis GmbH	Cat. No. 10688.01
Alexa Fluor 405 NHS Ester (Succinimidyl Ester)	Thermo Fisher Scientific	Cat. No. A30000
Alexa Fluor 488 NHS Ester (Succinimidyl Ester)	Thermo Fisher Scientific	Cat. No. A20000
Alexa Fluor 568 NHS Ester (Succinimidyl Ester)	Thermo Fisher Scientific	Cat. No. A20003
Alexa Fluor 647 NHS Ester (Succinimidyl Ester)	Thermo Fisher Scientific	Cat. No. A20006
2-Amino-2-(hydroxymethyl)-1,3-propanediol (Tris base)	Sigma-Aldrich	Cat. No. T1503
Ammonium persulfate (APS)	Sigma-Aldrich	Cat. No. A3678
Ampicillin sodium salt	Carl ROTH	Art. No. K029.2
Bacto agar	BD Difco	Cat. No. 214030
Bacto tryptone	Gibco	Cat. No. 211699
Bacto yeast extract	Gibco	Cat. No. 212720
Biotin	Sigma-Aldrich	Cat. No. B4501
3-(3-cholamidopropyl-dimethylammonio)-2-hydroxy-1-propanesulfonate (CHAPSO)	Sigma-Aldrich	Cat. No. C3649
cOmplete EDTA-free Protease Inhibitor Cocktail	Sigma-Aldrich	Cat. No. 5056489001

'Der Blaue Jonas' Single-step Coomassie Blue protein gel dye	BIOZOL	Cat. No. GRP1
1,4-Dithiothreitol (DTT)	Thermo Fisher Scientific	Cat. No. R0862
DNA Oligonucleotides	Sigma-Aldrich	N/A
Ethanol	Thermo Fisher Scientific	Cat. No. 10048291
Ethylenediaminetetraacetic acid (EDTA)	Supelco	Cat. No. 1.08452
Formaldehyde (Formalin)	Electron Microscopy Sciences	Cat. No. 50-00-0
2.5% glutaraldehyde	Electron Microscopy Sciences	Cat. No. 111-30-8
Glycerol (86%)	Carl ROTH	Art. No. 4043.2
Glycine	Carl ROTH	Art. No. 3908.3
Hydrochloric acid (HCl, 37%)	Sigma-Aldrich	Cat. No. 258148
Hydrogen peroxide, 30% (H_2O_2, Certified ACS)	Thermo Fisher Scientific	Cat. No. H325500
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Carl ROTH	Art. No. HN77.5
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Carl ROTH	Art. No. CN08.3
Kanamycin sulfate	Sigma-Aldrich	Cat. No. K4000
Lead citrate	Electron Microscopy Sciences	Cat. No. 17810
LR-white resin	ProSciTech	Cat. No. C023
Magnesium acetate ($Mg(OAc)_2$)	Supelco	Cat. No. 1.05819
β-mercaptoethanol (β-ME)	Sigma-Aldrich	Cat. No. M6250
Sodium thiosulfate ($Na_2S_2O_3$)	Sigma-Aldrich	Cat. No. 217263
Nickel-nitrilotriacetic acid (Ni-NTA) agarose	QIAGEN	Cat. No. 30250
2-(N-morpholino)ethanesulfonic acid (MES)	SERVA Electrophoresis GmbH	Cat. No. 29834.03
N,N,N',N'-Tetramethylethylenediamine (TEMED)	Sigma-Aldrich	Cat. No. T7024
Octyl-β-D-glucopyranoside (OG)	Sigma-Aldrich	Cat. No. O8001

Osmium tetroxide (OsO_4)	Sigma-Aldrich	Cat. No. 201030
PageRuler prestained protein ladder	Thermo Fisher Scientific	Cat. No. 26616
Phosphate Buffered Saline (PBS)	Thermo Fisher Scientific	Cat. No. 20012068
Poly L-lysine hydrobromide (polylysine)	Polysciences	Cat. No. 09730
Potassium chloride (KCl)	Carl ROTH	Art. No. 6781.1
Potassium hydroxide (KOH)	Carl ROTH	Art. No. P747.2
Sodium chloride (NaCl)	Carl ROTH	Art. No. 9265.2
Sodium dodecyl sulfate (SDS)	SERVA Electrophoresis GmbH	Cat. No. 20765.03
Strep-Tactin XT 4Flow resin	IBA Lifesciences	Cat. No. 2-5010-010
SYBR Safe DNA Gel Stain	Thermo Fisher Scientific	Cat. No. S33102
Tris(2-carboxyethyl)phosphine (TCEP)-HCl	Thermo Fisher Scientific	Cat. No. 20491
Uranyl acetate	Polysciences	Cat. No. 21447

2.1.2. Proteins/enzymes, consumables and commercial kits

Table 2.2 Proteins/enzymes, consumables and commercial kits

Reagent or resource	Source	Identifier
Proteins/enzymes		
Bovine serum albumin (BSA)	Thermo Fisher Scientific	Cat. No. 23209
DpnI	NEW ENGLAND Biolabs	Cat. No. R0176L
Herculase II Fusion DNA Polymerase	Agilent	Cat. No. 600679
Lysozyme	Sigma-Aldrich	Cat. No. L6876
His6-Usp2	(Baker et al., 2005)	N/A
Consumables or kits		
Consumables		
Millex-GP filter (0.22 µm)	Merck Millipore	Cat. No. SLGP033RB

Novex Value 4-12% Tris-Glycine gels	Thermo Fisher Scientific	Cat. No. XV04125PK20
Novex WedgeWell 14% Tris-Glycine gels	Thermo Fisher Scientific	Cat. No. XP00145BOX
NuPAGE 4-12% Bis-Tris protein gels	Thermo Fisher Scientific	Cat. No. NP0323BOX
Protein LoBind Tubes (0.5 mL)	Eppendorf	Cat. No. EP0030108094
Protein LoBind Tubes (1.5 mL)	Eppendorf	Cat. No. EP0030108116
Quantifoil R1.2/1.3 300 mesh copper grids	Plano	Cat. No. S143-2
Quantifoil R1.2/1.3 300 mesh gold grids	Plano	Cat. No. S143-8
Quantifoil R1.2/1.3 400 mesh gold grids (graphene oxide)	Electron Microscopy Sciences	Cat. No. GOQ400R1213Au10
Quantifoil R2/1 300 mesh copper grids	Plano	Cat. No. S174-2
Quantifoil R2/4 300 mesh copper grids (graphene oxide)	Electron Microscopy Sciences	Cat. No. GOQ300R24Cu10
Slide-A-Lyzer mini Dialysis Devices (MWCO 3.5 kDa)	Thermo Fisher Scientific	Cat. No. 69550
Vivaspin 20 MWCO 10 kDa centrifugal concentrator	Cytiva	Cat. No. 28932360
µ-Slide angiogenesis 15 well	Ibidi	Cat. No. 81506
Kits		
Wizard SV Gel and PCR Clean-Up system	Promega	Cat. No. A9282
Gibson Assembly Cloning Kit	NEW ENGLAND Biolabs	Cat. No. E2611L
QIAprep Spin Miniprep Kit	QIAGEN	Cat. No. 27106

2.1.3. Media and buffer

Table 2.3 Instruments and software

Media or buffers	**Composition**
Media	
Lysogeny broth (LB) medium	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl
LB plates	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 15 g/L agar
BG-11 medium	(Price & Badger, 1989a)
BG-11 plates	(Woodger et al., 2003)
Buffers	
Buffer A	50 mM Tris-HCl pH 8.0/300 mM NaCl/20 mM β-mercaptoethanol (β-ME)/5% (v/v) glycerol
Buffer B	50 mM Tris-HCl pH 8.0/300 mM KCl/5 mM DTT/10% (v/v) glycerol
Buffer C	50 mM Tris-HCl pH 8.0/500 mM NaCl/5 mM DTT/10% (v/v) glycerol
Buffer D	50 mM Tris-HCl pH 8.0/150 mM KCl
Buffer E	50 mM Tris-HCl pH 8.0/500 mM KCl
Buffer F	50 mM Tris-HCl pH 8.0/300 mM KCl
Buffer G	50 mM Tris-HCl pH 8.0/100 mM KCl/10 mM $Mg(OAc)_2$
Buffer H	50 mM Tris-HCl pH 8.0/50 mM KCl/10 mM $Mg(OAc)_2$
Buffer I	50 mM Tris-HCl pH 8.0/100 mM KCl/5 mM DTT
Native electrophoresis buffer	25 mM Tris base/192 mM glycine
2 × Native loading buffer	100 mM Tris-HCl pH 6.8/0.2% (w/v) bromophenol blue/20% (v/v) glycerol/50 mM DTT
2-(N-morpholino)ethanesulfonic acid (MES) running buffer	50 mM MES/50 mM Tris base/1 mM EDTA/0.1% (w/v) SDS
2 × SDS sample buffer	100 mM Tris-HCl pH 6.8/4% (w/v) SDS/0.2% (w/v) bromophenol blue/20% (v/v) glycerol/50 mM DTT
Tris-glycine running buffer	25 mM Tris base/192 mM glycine/0.1% (w/v) SDS

2.1.4. Instruments and software

Table 2.4 Instruments and software

Instruments or software	**Supplier**
Instruments	
ÄKTA pure chromatography system and columns (HisTrap HP, HiPrep 26/10, HiLoad 26/600 Superdex 75, HiLoad 26/600 Superdex 200, and Superdex 200 Increase 10/300 GL)	Cytiva
Allegra X-15R refrigerated centrifuge with SX4750A ARIES swinging-bucket rotor assembly	Beckman Coulter
Avanti J6-MI centrifuge equipped with a JS 4.2 swinging-bucket rotor assembly	Beckman Coulter
Function Line B12 incubator	Heraeus
BioPhotometer plus	Eppendorf
Centrifuge 5424 equipped with a FA-45-24-11 rotor	Eppendorf
Centrifuge 5427R equipped with a FA-45-30-11 rotor	Eppendorf
Circulator heating water bath	Lauda E100
CryoCube F570h	Eppendorf
DAWN EOS MALS detector	Wyatt Technology
Electrophoresis power supply	Bio-Rad
EmulsiFlex-C5 high pressure homogenizer	Avestin, Inc.
GIF Quantum Energy Filters	Gatan
Glacios transmission electron microscopy (TEM)	Thermo Fisher Scientific
Innova 40 benchtop incubator shaker	New Brunswick Scientific
Innova S44i incubator shaker	Eppendorf
ITC200 calorimeter	MicroCal/GE
JEM F200 TEM	JEOL
K2 summit direct electron detector	Gatan
K3 direct electron detector	Gatan

Mini Gel Tank	Thermo Fisher Scientific
Mini-PROTEAN Tetra System	Bio-Rad
Nanodrop ND 1000 Spectrophotometer	Peqlab
Optilab rEX refractive index detector	Wyatt Technology
PDC-32G-2	Harrick Plasma
PD MiniTrap G-10 column	Cytiva
Perfection V850 Pro flatbed scanner	EPSON
pH meter	WTW inoLab
Rio 16 camera	Gatan
Talos Arctica TEM	Thermo Fisher Scientific
TCS SP8 AOBS confocal laser scanning microscope (HCX PL APO 63×/1.2 water objective)	Leica Microsystems
Thermomixer Comfort	Eppendorf
Titan Krios	Thermo Fisher Scientific
T3 Thermocycler	Biometra
Ultra-micro quartz cuvette	Hellma
Ultramicrotome	Leica EM UC7
UV absorbance detector (1100 series)	Agilent
Vitrobot Mark IV	Thermo Fisher Scientific
Vortex-Genie 2	Scientific Industries
V-560 UV/VIS Spectrophotometer	Jasco
Software	
AlphaFold 3	https://alphafoldserver.com/
CryoSPARC version 4.2	Structura Biotechnology Inc. https://cryosparc.com/
DigitalMicrograph 3.5	Gatan
EFI-EST (Enzyme Function Initiative-Enzyme Similarity Tool)	https://efi.igb.illinois.edu/efi-est/
EFI-GNT (Enzyme Function Initiative-Genome Neighborhood Tool)	https://efi.igb.illinois.edu/efi-gnt/
ESPript 3.0	https://espript.ibcp.fr/

Expasy server/ProtParam	https://web.expasy.org/cgi-bin/protparam/protparam
Fiji	https://imagej.net/Welcome
ImageJ	National Institutes of Health
OriginPro 2020	OriginLab https://www.originlab.com/
PyMOL version 3.1.3.1, Schrödinger, LLC	DeLano Scientific http://www.pymol.org/
SerialEM version 4.1	https://bio3d.colorado.edu/SerialEM/
SigmaPlot 14	https://www.grafiti.com/

2.2. Sequence alignment and analysis of the genome neighborhood of ApN and CM

The depicted ApN sequence alignments were performed using full-length protein sequences from the following selected cyanobacterial strains: *Synechococcus elongatus* PCC 7942 (FACHB-805, WP_011242446.1), *Synechocystis* sp. PCC 6803 (AGF50461.1), *Chlorogloeopsis fritschii* PCC 6912 (WP_016874098.1), *Gloeobacter violaceus* PCC 7421 (WP_011142090.1), *Nostoc* sp. PCC 7120 (WP_010995038.1), *Thermosynechococcus vestitus* BP-1 (WP_011056787.1), *Tolypothrix tenuis* PCC 7101 (WP_096574049.1), *Synechocystis* sp. PCC 6714 (WP_028947280.1). The alignments were visualized using ESPript 3.0 (Robert & Gouet, 2014).

Alignments of CM proteins were performed using full-length sequences from the same cyanobacterial species selected above for ApN.

The Enzyme Similarity Tool (EFI-EST) was used to generate a sequence similarity network (SSN) to identify functionally related proteins, using the BLAST-derived sequence of *Se*7942 ApN as the input query. In this network, the most related proteins (ApN and its closest homologs in the UniProt, UniRef90, or UniRef50 databases) are grouped together into clusters, which allows visualization of the relationships among protein sequences. The term “bacteria” was used to restrict the retrieved sequences to this taxonomy group, resulting in the selection of 543 genomes for SSN construction. The SSN file was then uploaded to the Genome

Neighborhood Tool (EFI-GNT) to generate genome neighborhood networks (GNNs), in which the neighboring genes are sorted into neighbor Pfam families (Oberg et al., 2023; Zallot et al., 2019).

2.3. Molecular biological methods

2.3.1. Plasmid construction

Plasmid construction was performed by Kun Zang and me, except for pUC18-*ΔKO* and pSE4-*ccmKLMNO* generated by Nghiem D. Nguyen, Loraine M. Rourke and Jiwon Lee.

For the preparation of DNA of *ccmN*, the cyanobacterium *Synechococcus elongatus* PCC 7942 (*Se*7942) (Institut Pasteur Paris) was cultured in BG-11 medium at 30 °C and shaking at 50 rpm under continuous light until reaching high cell density. Cells were pelleted by centrifugation at 10,000 × g for 10 minutes (min) at 4 °C using an Eppendorf Centrifuge 5427R equipped with a FA-45-30-11 rotor (Eppendorf). The cell pellet was resuspended in 100 mL of buffer 50 mM Tris-HCl pH 8.0, 50 mM NaCl and lysed by five cycles of heating (3 min at 95 °C) and snap freezing in liquid nitrogen. The lysate was centrifuged at 20,000 × g for 10 min at 4 °C, and 1 μL of the supernatant was used directly as template for subsequent polymerase chain reaction (PCR). All plasmids were constructed using PCR amplification with Herculase II Fusion DNA Polymerase (Agilent) on a T3 Thermocycler (Biometra). The PCR products were analyzed via Agarose gel electrophoresis in Tris-acetate-EDTA (TAE) buffer, followed by digestion with DpnI (NEW ENGLAND Biolabs) at 37 °C for 2 hours (h) to remove the template DNA. The resulting DNA fragments were then purified using the Wizard SV Gel and PCR Clean-Up system (Promega), following the manufacturer’s protocol. For plasmid assembly, the Gibson Assembly Cloning Kit (NEW ENGLAND Biolabs) was used. Finally, plasmid purification was performed using the QIAprep Spin Miniprep Kit (QIAGEN) according to the manufacturer’s instructions.

pHUE-*Se*ApN was generated by amplification of the *ccmN* gene from *Se*7942 genomic DNA and subsequent cloning into the pHUE vector with His6-ubiquitin (His6ubi) tag fused at the N-terminus (Baker et al., 2005; Catanzariti et al., 2004). pET28a-StrepII-*Se*ApN was generated by amplification of the *ccmN* gene from pHUE-*Se*ApN with Strep-tag II (GSWSHPQFEKGSG) inserted after the initial methionine and subsequently cloned into the pET28a vector. pET28a-*Se*CM-His10 or pET28a-*Se*CM were generated by amplifying the *ccmM* gene from pHUE-*Se*M58 (Zang et al., 2021) and cloning into the pET28a vector, either with or without a C-terminal 10 × His tag (GSGGSHHHHHHHHHH).

The bicistronic plasmids pET28a-*Se*CM-His6ubi-*Se*ApN and pET28a-*Se*CM-His10-StrepII-*Se*ApN were obtained by amplifying His6ubi-*Se*ApN or StrepII-*Se*ApN with their preceding ribosome binding site (RBS) from pHUE-*Se*ApN or pET28a-StrepII-*Se*ApN and subsequently cloning C-terminally into pET28a-*Se*CM or pET28a-*Se*CM-His10, respectively. The full-length CM fragment in both bicistronic plasmids was replaced by its first 198 residues (1-198) to generate pET28a-*Se*CM$^{\text{Nt}}$-His6ubi-*Se*ApN and pET28a-*Se*CM$^{\text{Nt}}$-His10-StrepII-*Se*ApN.

pET28a-*Se*CM-His10-*Se*ApN or pET28a-*Se*CM-His10-*Se*ApNΔN10 were generated by replacing the StrepII-*Se*ApN fragment in pET28a-*Se*CM-His10-StrepII-*Se*ApN with either the full-length ApN or ApN lacking the first 10 residues (2-11) after the initial methionine.

Site-directed mutagenesis was then used to generate cysteine-to-alanine (C49A/C88A), cysteine-to-serine (C49S/C88S), leucine-to-aspartic acid (L76D), leucine-to-arginine (L76R), and threonine-to-arginine (T94R) mutations across all relevant backbones, resulting in the following mutant constructs: pET28a-*Se*CM-His6ubi-*Se*ApN$_{\text{C49A/C88A}}$, pET28a-*Se*CM-His6ubi-*Se*ApN$_{\text{C49S/C88S}}$, pET28a-*Se*CM-His10-StrepII-*Se*ApN$_{\text{C49A/C88A}}$, pET28a-*Se*CM-His10-StrepII-*Se*ApN$_{\text{C49S/C88S}}$, pET28a-*Se*CM$^{\text{Nt}}$-His10-StrepII-*Se*ApN$_{\text{C49A/C88A}}$, pET28a-*Se*CM$^{\text{Nt}}$-His10-StrepII-*Se*ApN$_{\text{C49S/C88S}}$, pET28a-*Se*CM$^{\text{Nt}}$-His10-StrepII-*Se*ApN$_{\text{L76D}}$, pET28a-*Se*CM$^{\text{Nt}}$-His10-StrepII-*Se*ApN$_{\text{L76R}}$, and pET28a-*Se*CM$^{\text{Nt}}$-His10-StrepII-*Se*ApN$_{\text{T94R}}$.

pUC18-*ΔKO* was generated through PCR amplification of the native upstream (A) and downstream (B) flanking regions (~1,000 bp) of *Se*7942 *ccmKLMNO* genes in the *ccm* operon, including EcoRI and BamHI sites in flank A and BamHI and HindIII sites in flank B (5' to 3'). These fragments were A-tailed and ligated into pGem-T. Flank A was cloned into pUC18-*Amp*^{R} to generate pUC18-FlankA, and flank B was inserted into pUC18-FlankA to generate pUC18-FlankAB. The *Kan*^{R} cassette from pUC4K was isolated using BamHI and cloned directly into pUC18-FlankAB to generate pUC18-*ΔKO*.

pSE4-*ccmKLMNO* was generated from a synthesized *ccmKLMNO* fragment with its native promoter region and flanking NcoI and XbaI sites into the corresponding sites of the pSE4 vector, replicating the coding sequence of the *Se*7942 *ccm* operon. A single additional NheI site (in *ccmM* gene) was eliminated during synthesis to enable successful assembly. The oligonucleotides used for PCR amplification and the plasmids generated in this study are listed in Table **2.5**.

Table 2.5 List of plasmids and primer sequences

Plasmid	**Oligonucleotides sequence (5' to 3')**	**Source**
pHUE-*Se*ApN	TTGCGCCTCCGCGGTGGAATGCATCTACCGCCCCTA	This study
	GGCCGCGTCGACGGTACCTTAGCGATCGGGGAACAT	
	ATGTTCCCCGATCGCTAAGGTACCGTCGACGCGGCC	
	TAGGGGCGGTAGATGCATTCCACCGCGGAGGCGCAA	
pET28a-StrepII-*Se*ApN	CCTCAGTTCGAAAAAGGCAGCGGCCATCTACCGCCCCTAGAGCCCCCG	This study
	AGCCGGATCTAAGCTTTTAGCGATCGGGGAACATGCTCT	
	GTTCCCCGATCGCTAAAAGCTTAGATCCGGCTGCTAACAAAG	
	ATGGCCGCTGCCTTTTTCGAACTGAGGATGACTCCAGCTGCCCATGGTATATCTCCTTCTTA	
pET28a-*Se*CM-His10	TTAAGAAGGAGATATACCATGCCGAGCCCAACAACG	This study
	GATGGTGATGATGGTGGTGATGGTGACTGCCTCCAGATCCCGGCTTTTGAATCAACAGTTCAGCC	
	CACCATCACCACCATCATCACCATCACCATTAAAAGCTTAGATCCGGCTGCTAACAAAG	
	CGTTGTTGGGCTCGGCATGGTATATCTCCTTCTTAA	
pET28a-*Se*CM	TTAAGAAGGAGATATACCATGCCGAGCCCAACAACG	This study
	CAGCCGGATCTAAGCTTTTACGGCTTTTGAATCAACAGTTCAGCC	
	CTGTTGATTCAAAAGCCGTAAAAGCTTAGATCCGGCTGCTAACAA	
	CGTTGTTGGGCTCGGCATGGTATATCTCCTTCTTAA	
pET28a-*Se*CM-His6ubi-*Se*ApN	ATTCAAAAGCCGTAACCTCTAGAAATAATTTTGTTTAACTTTAAG	This study
	TAGAGGCCCCAAGGGGTTATGCTAG	
	TTAAGAAGGAGATATACCATGCCGAGCCCAACAACG	
	AATTATTTCTAGAGGTTACGGCTTTTGAATCAACAGTTCAGCCAC	
pET28a-*Se*CM^{Nt}-His6ubi-*Se*ApN	TTAAGAAGGAGATATACCATGCCGAGCCCAACAACG	This study
	TATTTCTAGAGGTTAGTGGAAATCAGCAGTAGCTGCTGGA	
	ACTGCTGATTTCCACTAACCTCTAGAAATAATTTTGTTTA	
	CGTTGTTGGGCTCGGCATGGTATATCTCCTTCTTAA	
pET28a-*Se*CM-His10-StrepII-*Se*ApN	CCATCATCACCATCACCATTAACCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGAT	This study
	TAGAGGCCCCAAGGGGTTATGCTAG	
	CTAGCATAACCCCTTGGGGCCTCTA	
	CAAAATTATTTCTAGAGGTTAATGGTGATGGTGATGATGGTGGTGATGGTGACT	
pET28a-*Se*CM^{Nt}-His10-StrepII-*Se*ApN	TTAAGAAGGAGATATACCATGCCGAGCCCAACAACG	This study
	TGACTGCCTCCAGATCCGTGGAAATCAGCAGTAGCTGCTGGAGTA	
	CAGCTACTGCTGATTTCCACGGATCTGGAGGCAGTCACCATCACC	

	CGTTGTTGGGCTCGGCATGGTATATCTCCTTCTTAA	
pET28a-*Se*CM-His10-*Se*ApN	GAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGCATCTACCGCCCCTAGAGCC TAGAGGCCCCAAGGGGTTATGCTAG CTAGCATAACCCCTTGGGGCCTCTA GGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGTTAATGGTGATGGTGATGATGG	This study
pET28a-*Se*CM-His10-*Se*ApNΔN10	GAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGATCGCTACTTTGCCAGTGGTGAGG TAGAGGCCCCAAGGGGTTATGCTAG CTAGCATAACCCCTTGGGGCCTCTA GGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGTTAATGGTGATGGTGATGATGG	This study
All C49A/C88A mutants	TTGAGATCGCAAGTGGCGTCGCTATTGGCCTTGGATCGGTGAT ATCACCGATCCAAGGCCAATAGCGACGCCACTTGCGATCTCAA GTATTGTCGGTCGTCAGGCTGCCCTGGGTGCATCCACCACTCT AGAGTGGTGGATGCACCCAGGGCAGCCTGACGACCGACAATAC	This study
All C49S/C88S mutants	TTGAGATCGCAAGTGGCGTCAGTATTGGCCTTGGATCGGTG CACCGATCCAAGGCCAATACTGACGCCACTTGCGATCTCAA AGTATTGTCGGTCGTCAGGCTAGTCTGGGTGCATCCACCACTCTG CAGAGTGGTGGATGCACCCAGACTAGCCTGACGACCGACAATACT	This study
pET28a-*Se*CM^{Nt}-His10-StrepII-*Se*ApN$_{L76D}$	GCCTTGCTTGCTGCCGGAGTGGACATTGTTGGCCAGAGTATTGTC GACAATACTCTGGCCAACAATGTCCACTCCGGCAGCAAGCAAGGC	This study
pET28a-*Se*CM^{Nt}-His10-StrepII-*Se*ApN$_{L76R}$	GCCTTGCTTGCTGCCGGAGTGCGCATTGTTGGCCAGAGTATTGTC GACAATACTCTGGCCAACAATGCGCACTCCGGCAGCAAGCAAGGC	This study
pET28a-*Se*CM^{Nt}-His10-StrepII-*Se*ApN$_{L94R}$	GCTTGCCTGGGTGCATCCACCCGCCTGGTTAACACCTCGATCGA TCGATCGAGGTGTTAACCAGGCGGGTGGATGCACCCAGGCAAGC	This study
pHUE-*Se*M58		(Zang et al., 2021)
pHUE-*Se*γCAL(1-198)		(Zang et al., 2021)
pHUE-*Se*CcaA		(Zang et al., 2021)
pHUE-*Se*CcaAΔC2		(Zang et al., 2021)
pHUE-E$_{GFP}$C2$_{17}$		(Zang et al., 2021)
pHUE_*Syn*6301_*ccmM*_M35 (197-539)		(Wang et al., 2019)
pET11a_*Syn*6301_*rbcL*		(Saschenbrecker et al., 2007)
pET11a_*Syn*6301_*rbcS*		(Saschenbrecker et al., 2007)
pSE4-*ccmKLMNO*		This study
pUC18-*ΔKO*		This study

2.3.2. Transformation of competent cells

For plasmid amplification, 10 μL of DNA ligation mixture was transformed into 50 μL of *E. coli* DH5α competent cells. The cells and the DNA ligation mixture or plasmid were incubated on ice for 30 min, then heat shocked in a circulator heating water bath (Lauda E100) at 42 °C for 1 min, then immediately cooled on ice for 4 min, followed by incubation with 450 µL of LB medium without antibiotics at 37 °C for 1 h with shaking at 1150 rpm (Eppendorf Thermomixer Comfort). Afterwards, the cells were pelleted by centrifugation for 2 min at 3000 rpm and room temperature (RT) using an Eppendorf Centrifuge 5424 equipped with a FA-45-24-11 rotor (Eppendorf), and 250 µL of supernatant was removed. The cell pellet was resuspended with the remaining medium and plated onto pre-warmed LB agar plates containing the appropriate antibiotic (100 µg/mL ampicillin or 50 µg/mL kanamycin, depending on the plasmid resistance marker). Plates were incubated overnight at 37 °C in a Function Line B12 incubator (Heraeus). Positive clones were selected and confirmed by DNA sequencing. For recombinant protein expression, 1 µL of purified plasmid was added to 50 µL of *E. coli* BL21 (DE3) competent cells, followed by incubation on ice, heat shock, and subsequent steps as described above for plasmid amplification.

2.4. Biochemical and biophysical assays

2.4.1. Protein expression and purification

Protein expression and purification were performed by Kun Zang and me. The proteins used in this study are listed in Table **2.6**.

Table 2.6 Proteins purified from the respective plasmid

Plasmid	**Protein**
pHUE-*Se*ApN	ApN
pET28a-StrepII-*Se*ApN	SIIApN
pET28a-*Se*CM-His6ubi-*Se*ApN	CM-ApN
pET28a-*Se*CM-His6ubi-*Se*ApN$_{C49A/C88A}$	CM-(ApN-2A)
pET28a-*Se*CM-His6ubi-*Se*ApN$_{C49S/C88S}$	CM-(ApN-2S)
pET28a-*Se*CM^{Nt}-His6ubi-*Se*ApN	CM^{Nt}-ApN
pET28a-*Se*CM-His10-StrepII-*Se*ApN	(CM-H$_{10}$)-SIIApN; CM-H$_{10}$; CM^{Ct}-H$_{10}$
pET28a-*Se*CM-His10-StrepII-*Se*ApN$_{C49A/C88A}$	(CM-H$_{10}$)-(SIIApN-2A)
pET28a-*Se*CM-His10-StrepII-*Se*ApN$_{C49S/C88S}$	(CM-H$_{10}$)-(SIIApN-2S)

pET28a-*Se*CM^{Nt}-His10-StrepII-*Se*ApN	(CM^{Nt}-H$_{10}$)-SIIApN; CM^{Nt}-H$_{10}$
pET28a-*Se*CM^{Nt}-His10-StrepII-*Se*ApN$_{C49A/C88A}$	(CM^{Nt}-H$_{10}$)-(SIIApN-2A)
pET28a-*Se*CM^{Nt}-His10-StrepII-*Se*ApN$_{C49S/C88S}$	(CM^{Nt}-H$_{10}$)-(SIIApN-2S)
pET28a-*Se*CM^{Nt}-His10-StrepII-*Se*ApN$_{L76D}$	SIIApN-L76D; CM^{Nt}-H$_{10}$
pET28a-*Se*CM^{Nt}-His10-StrepII-*Se*ApN$_{L76R}$	SIIApN-L76R; CM^{Nt}-H$_{10}$
pET28a-*Se*CM^{Nt}-His10-StrepII-*Se*ApN$_{T94R}$	SIIApN-T94R; CM^{Nt}-H$_{10}$
pET28a-*Se*CM-His10-*Se*ApN	(CM-H$_{10}$)-ApN
pET28a-*Se*CM-His10-*Se*ApNΔN10	(CM-H$_{10}$)-ApNΔN10
pHUE-*Se*M58	CM (M58)
pHUE-*Se*γCAL(1-198)	CM^{Nt}
pHUE-*Se*CcaA	CA
pHUE-*Se*CcaAΔC2	CAΔC (CcaAΔC2)
pHUE-EGFPC2$_{17}$	EGFPCA$_{C17}$ (EGFPC2$_{17}$)
pHUE_*Syn*6301_*ccmM*_M35 (197-539)	CM^{Ct} (M35)
pET11a_*Syn*6301_*rbcL*	RbcL$_8$
pET11a_*Syn*6301_*rbcS*	RbcS

The proteins *Se*CM (Zang et al., 2021), *Se*CM^{Nt} (Zang et al., 2021), *Se*CA (Zang et al., 2021), *Se*CAΔC (Zang et al., 2021), EGFPCA$_{C17}$ (Zang et al., 2021), *Syn*CM^{Ct} (Wang et al., 2019) and *Syn*Rubisco (Saschenbrecker et al., 2007) proteins were expressed and purified as previously described.

2.4.1.1. Protein expression

For other proteins used in this study, individual clones were transferred into 5 mL of lysogeny broth (LB) medium supplemented with the appropriate antibiotic (100 µg/mL ampicillin for pHUE-vector proteins and pET11a-vector proteins, or 50 µg/mL kanamycin for pET28a-vector proteins) and grown for 4 h at 37 °C with shaking at 180 rpm in an Innova 40 benchtop incubator shaker (New Brunswick Scientific). This starter culture was then transferred to 1 L of antibiotic-supplemented LB medium and grown at 37 °C for ~3 h in a shaker flask using an Innova S44i incubator shaker (Eppendorf) to reach an optical density (OD$_{600}$) of 0.6-0.8. The temperature was reduced to 18 °C, and after ~1 h, 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce protein expression. The induced culture was incubated overnight for 16~18 h at 18 °C. Cells were harvested by centrifugation at 4,000 rpm and 4 °C for 30 min using an Avanti J6-MI centrifuge equipped with a JS 4.2 swinging-bucket rotor assembly (Beckman Coulter). The cell pellet was washed with medium and subjected to secondary centrifugation at 4,200 rpm and 4 °C for 30 min using an Allegra X-15R refrigerated centrifuge with a SX4750A ARIES swinging-bucket rotor assembly (Beckman Coulter). Cells were either flash-frozen in liquid nitrogen (N_2) for storage or processed immediately as described below.

2.4.1.2. Protein purification

The harvested cells were resuspended and incubated in ice-cold lysis buffer A (50 mM Tris-HCl pH 8.0/300 mM NaCl/20 mM β-mercaptoethanol (β-ME)/5% (v/v) glycerol) with or without 20 mM or 50 mM imidazole and supplemented with 1 mg/mL lysozyme, 2.5 U/mL *Serratia marcescens* DNAse (*Sm*DNAse) and complete EDTA-free protease inhibitor cocktail (Sigma-Aldrich) for 1 h at 4 °C with gentle agitation. Cell lysis was performed using an EmulsiFlex-C5 high pressure homogenizer (Avestin, Inc.) at 10,000-15,000 psi. The cell lysate was clarified by centrifugation at 22,000 rpm and 4 °C for 30 min, after which the supernatant was filtered through a 0.22 μm pore-size Millex-GP filter (Merck Millipore). The subsequent purification was performed as described below.

2.4.1.2.1. Purification of ApN

The filtered supernatant was loaded onto a buffer A/20 mM imidazole pre-equilibrated HisTrap HP column (Cytiva) at 4 °C, washed with 10 column volumes (CV) of buffer A/20 mM imidazole and eluted with a 20 CV linear gradient (20-500 mM imidazole) buffer A. The His6ubi moiety was cleaved by mixing the protein eluate with the 6 × His-tagged deubiquitylating enzyme Usp2 (His6-Usp2) (Baker et al., 2005) overnight at 4 °C. The cleaved protein was buffer exchanged with a HiPrep 26/10 desalting column (Cytiva) into buffer A and then applied to a HisTrap HP column (Cytiva) pre-equilibrated with buffer A to trap His6-Usp2, the cleaved His6ubi moiety and any uncleaved protein. The flow-through was concentrated to ~5 mL and subjected to size-exclusion chromatography (SEC) on HiLoad 26/600 Superdex 200 (Cytiva) in buffer B (50 mM Tris-HCl pH 8.0/300 mM KCl/5 mM DTT/10% (v/v) glycerol). Fractions containing protein were concentrated using a Vivaspin 20 MWCO 10 kDa centrifugal concentrator (Cytiva) and then applied to a SEC column (HiLoad 26/600 Superdex 75, Cytiva) equilibrated with buffer B. Finally, the purified protein was concentrated by ultrafiltration using a Vivaspin 20 MWCO 10 kDa centrifugal concentrator (Cytiva), aliquoted and flash-frozen in liquid N_2 for storage at -80 °C. All the purified proteins described below follow the same concentration and storage procedure.

2.4.1.2.2. Purification of SIIApN

The filtered supernatant was loaded onto a Strep-Tactin XT 4Flow resin (IBA Lifesciences) gravity column, which was pre-equilibrated with buffer A. The column was then washed with 5 CV of buffer A, and the bound protein was eluted using buffer A supplemented with 50 mM biotin (Sigma-Aldrich). The eluted protein was further purified by sequential SEC on HiLoad 26/600 Superdex 200 and HiLoad 26/600 Superdex 75 (Cytiva) that were equilibrated with buffer B. The purified proteins were then concentrated and stored.

2.4.1.2.3. Purification of CM-ApN, CM^{Nt}-ApN and mutants

The CM-ApN or CM^{Nt}-ApN complex was purified using the same protocol as described for ApN. Notably, only the eluted fractions containing the CM-(His6ubi-ApN) complex or CM^{Nt}-(His6ubi-ApN) complex from the HisTrap HP column (Cytiva) were pooled. The His6ubi moiety was then cleaved and all proteins containing a 6 × His tag were removed as described for ApN. The sample was then subjected to SEC on a HiLoad 26/600 Superdex 200 column (Cytiva) pre-equilibrated with buffer B for further separation. The purified proteins were then concentrated and stored. The CM-ApN mutants were purified as described for CM-ApN.

2.4.1.2.4. Purification of (CM-H_{10})-SIIApN, (CM^{Nt}-H_{10})-SIIApN and mutants

The filtered supernatant was first loaded onto a nickel-nitrilotriacetic acid (Ni-NTA) agarose (QIAGEN) gravity column, equilibrated with buffer A/50 mM imidazole. The column was then washed with 5 CV of buffer A/50 mM imidazole. The unbound proteins in the flow-through and wash fractions were collected and loaded onto the Strep-Tactin XT 4Flow resin (IBA Lifesciences) gravity column to isolate the StrepII-tagged proteins (SIIApN). The Ni-NTA-bound protein was eluted with buffer A/500 mM imidazole and then applied to a Strep-Tactin XT 4Flow resin (IBA Lifesciences) column equilibrated with buffer A. The flow-through and subsequent 5 CV wash with buffer A contained only the 10 × His-tagged proteins (CM-H_{10} and CM^{Ct}-H_{10}). The Strep-Tactin-bound (CM-H_{10})-SIIApN complex was eluted with buffer A/50 mM biotin, and further purified by SEC on a HiLoad 26/600 Superdex 200 column (Cytiva) pre-equilibrated with buffer B. The 10 × His-tagged proteins (CM-H_{10} or CM^{Ct}-H_{10}) from the earlier Strep-Tactin flow-through and wash were also subjected to SEC in buffer C (50 mM Tris-HCl pH 8.0/500 mM NaCl/5 mM DTT/10% (v/v) glycerol). The purified proteins were then concentrated and stored. (CM^{Nt}-H_{10})-SIIApN was purified as described for (CM-H_{10})-SIIApN. Besides, the 10 × His-tagged proteins (CM^{Nt}-H_{10}) was finally purified by SEC (HiLoad 26/600 Superdex 200, Cytiva) pre-equilibrated in buffer D (50 mM Tris-HCl pH 8.0/150 mM KCl) containing 10% (v/v) glycerol. The (CM-H_{10})-SIIApN mutants and (CM^{Nt}-H_{10})-SIIApN mutants were purified as described for (CM-H_{10})-SIIApN and (CM^{Nt}-H_{10})-SIIApN, respectively.

2.4.1.2.5. Purification of (CM-H_{10})-ApN and (CM-H_{10})-ApNΔN10

The supernatant was loaded onto a Ni-NTA agarose (QIAGEN) gravity column, equilibrated with buffer A/50 mM imidazole, and washed with 5 CV of the same buffer. The bound protein was eluted with buffer A/500 mM imidazole and (CM-H_{10})-ApN complex or (CM-H_{10})-ApNΔN10 complex enriched fractions were pooled. These fractions were then further purified by SEC using a HiLoad 26/600 Superdex 200 column (Cytiva) pre-equilibrated with buffer B. The purified proteins were then concentrated and stored.

2.4.1.3. Determination of protein concentration

Protein concentrations were determined using the Beer-Lambert equation, with the absorbance at 280 nm obtained using a Nanodrop ND 1000 Spectrophotometer (Peqlab) and extinction coefficients calculated from their protein sequences with Expasy server/ProtParam (Wilkins et al., 1999). For homo-oligomeric proteins (*Se*CM, *Se*CA, *Se*ApN, and *Se*SIIApN), the calculation included one copy of the respective protein sequence. For *Syn*Rubisco, one copy of the sequence of each of the RbcL and RbcS subunits was used. For hetero-complexes, for example, $(ApN)_3$:CM, the calculation included three copies of ApN and one copy of CM sequence.

2.4.2. SDS-PAGE and native-PAGE

2.4.2.1. SDS-PAGE

All the proteins in this study were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The protein samples were denatured in 2 × SDS sample buffer (100 mM Tris-HCl pH 6.8/4% (w/v) SDS/0.2% (w/v) bromophenol blue/20% (v/v) glycerol/50 mM DTT) at 95 ºC for 10 min. Electrophoresis was performed for ApN-related proteins using either 14% Tris-Glycine gels (Novex WedgeWell 14% Tris-Glycine, mini protein gel, Thermo Fisher Scientific) or self-made 15% Tris-Glycine gels (Table **2.7**) in a Mini Gel Tank (Thermo Fisher Scientific) or Mini-PROTEAN Tetra System (Bio-Rad), respectively, in Tris-glycine running buffer (25 mM Tris base/192 mM glycine/0.1% (w/v) SDS) at a constant voltage of 150 V at RT. Other proteins were applied to 4-12% Bis-Tris gels (NuPAGE 4-12% Bis-Tris protein gels, Thermo Fisher Scientific) and run using a Mini Gel Tank (Thermo Fisher Scientific) in 2-(N-morpholino)ethanesulfonic acid (MES) running buffer (50 mM MES/50 mM Tris base/1 mM EDTA/0.1% (w/v) SDS) at a constant voltage of 180 V at RT. To fix and visualize the proteins, the gels were stained with 'Der Blaue Jonas' Single-step Coomassie Blue protein gel dye (BIOZOL) at RT for 1 h, followed by destaining in double-distilled water (ddH_2O) at RT for approximately 12 h. Gel images were digitized using a Perfection V850 Pro flatbed scanner (EPSON).

Table 2.7 Preparation of SDS-PAGE gels

Component (for 5 gels)	**15% resolving gel (20 mL)**	**5% stacking gel (5 mL)**
Acrylamide/Bis Solution, 37.5:1 (30% w/v)	10	0.8375
1.5 M Tris-HCl pH 8.8	5	-
1 M Tris-HCl pH 6.8	-	0.625
10% (w/v) SDS	0.2	0.05
ddH_2O	4.6	3.375
10% (w/v) APS	0.2	0.05
TEMED	0.008	0.005

For the following assays, proteins were retrieved from a -80 °C freezer CryoCube F570h (Eppendorf) and thawed on ice before performing the biochemical assays.

2.4.2.2. Native-PAGE

Samples were prepared in 2 × native loading buffer (100 mM Tris-HCl pH 6.8/0.2% (w/v) bromophenol blue/20% (v/v) glycerol/50 mM DTT). Electrophoresis was performed using 4-12% Tris-Glycine gels (Novex Value 4-12% Tris-Glycine, mini protein gel, Thermo Fisher Scientific) in a Mini Gel Tank (Thermo Fisher Scientific) in native electrophoresis buffer (25 mM Tris base/192 mM glycine) at a constant voltage of 80 V at 4 °C.

2.4.3. Size-exclusion chromatography coupled to multiangle static light scattering (SEC-MALS)

SEC-MALS data were obtained in collaboration with Kun Zang.

Purified proteins at 2 mg/mL were analyzed using static and dynamic light scattering by auto-injection of the sample onto a SEC column (Superdex 200 Increase 10/300 GL, Cytiva) at a flow rate of 0.35 mL/min in buffer D/5 mM DTT for most of the analyzed proteins, unless otherwise indicated, or buffer E (50 mM Tris-HCl pH 8.0/500 mM KCl) containing 5 mM DTT for CM and CM-H_{10} at RT. The column was in line with the following detectors: a variable ultraviolet (UV) absorbance detector set at 280 nm (Agilent 1100 series), a DAWN EOS MALS detector (Wyatt Technology, 690 nm laser) and an Optilab rEX 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 mg/mL. Bovine serum albumin (BSA, Thermo Fisher Scientific) was used as the calibration standard.

For analysis of protein samples (CM^{Nt}-H_{10})-SIIApN and (CM^{Nt}-H_{10})-(SIIApN-2A) upon oxidation, proteins were buffer exchanged into buffer F (50 mM Tris-HCl pH 8.0/300 mM KCl) using a PD MiniTrap G-10 column (Cytiva) to remove DTT, followed by incubation of the protein (2 mg/mL) with 5 mM H_2O_2 on ice for 1 h prior to SEC-MALS analysis in buffer D.

All of the SEC-MALS data are provided in Appendix **A.3**.

2.4.4. Strep-Tactin pull-down assay

The interaction between $(ApN)_4$ and the pro-carboxysome core proteins, $(CM)_3$, Rubisco, $(CA)_4$ or CM^{Ct} were detected by Strep-Tactin pull-down assay. All proteins were prepared as 10-fold concentrated stock solutions and diluted to the following final concentrations in the reaction system (500 µL): $(SIIApN)_4$ (1.5 µM), $(CM)_3$ (1.0 µM), Rubisco ($RbcL_8S_8$ holoenzyme, 0.25 µM), $(CA)_4$ (1.5 µM) or CM^{Ct} (3 µM). $(SIIApN)_4$ was incubated with $(CM)_3$, Rubisco, $(CA)_4$ or CM^{Ct} in buffer D/5 mM DTT in 1.5 mL Protein LoBind Tubes (Eppendorf) at 25 ºC for 30 min using a thermomixer (Eppendorf Thermomixer Comfort). Carry-over of salt was accounted for in the assay buffer. The mixture was then loaded onto a Strep-Tactin XT 4Flow resin (IBA Lifesciences) gravity column (500 µL) pre-equilibrated with buffer D/5 mM DTT, washed extensively with this buffer, and eluted with the same buffer supplemented with 50 mM biotin. Fractions were analyzed by SDS-PAGE and Coomassie staining.

The pull-down assay of $(SIIApN)_4$ and $(CM)_3$ was performed under reducing or oxidizing conditions. Under reducing conditions in buffer D/5 mM DTT, proteins were mixed and incubated at 25 ºC or 37 ºC for 30 min. The pull-down assay was then carried out as above. Under oxidizing conditions, $(SIIApN)_4$ was initially buffer exchanged using a PD MiniTrap G-10 column (Cytiva) into buffer F. First, DTT-free $(SIIApN)_4$ and $(CM)_3$ were oxidized separately with 2 mM H_2O_2 on ice for 30 min. Then, buffer exchange was performed using a PD MiniTrap G-10 column (Cytiva) into buffer F and buffer E, respectively, to remove the unreacted H_2O_2. The two oxidized proteins were then mixed and incubated in buffer D in 1.5 mL Protein LoBind Tubes (Eppendorf) at 25 ºC or 37 ºC for 30 min using a thermomixer (Eppendorf Thermomixer Comfort). Carry-over of salt was accounted for in the assay buffer. The reaction was then loaded onto a pre-equilibrated gravity-flow Strep-Tactin XT column (IBA Lifesciences) and washed extensively with buffer D. Finally, it was eluted with buffer D/50 mM biotin.

2.4.5. Sedimentation assay

Sedimentation assay was used to test if $(ApN)_4$ can be recruited into biomolecular condensates formed by $(CM)_3$, CM^{Ct}, Rubisco and $(CA)_4$. All proteins were prepared as 10-fold concentrated stock solutions and diluted to the following final concentrations in the reaction system (20 µL): Rubisco ($RbcL_8S_8$ holoenzyme, 0.5 µM), CM^{Ct} (2 µM), $(CM)_3$ (0.25 µM), $(CA)_4$ (0.125 µM), and $(ApN)_4$ (0.25 or 2.5 µM). Proteins were incubated in buffer G (50 mM Tris-HCl pH 8.0/100 mM KCl/10 mM $Mg(OAc)_2$) containing 5 mM DTT in 1.5 mL Protein LoBind Tubes (Eppendorf) at 25 ºC for 10 min using a thermomixer (Eppendorf Thermomixer Comfort). Carry-over of salt was accounted for in the assay buffer. Samples were immediately fractionated by centrifugation at 20,817 × g for 20 min at 25 ºC using an Eppendorf Centrifuge 5427R equipped with a FA-45-30-11 rotor (Eppendorf) into total (T), supernatant (S) and pellet (P) fractions, and analyzed by SDS-PAGE and Coomassie staining.

2.4.6. Isothermal titration calorimetry (ITC)

ITC data were collected in collaboration with Claire Basquin of the Department of Structural Cell Biology (MPIB).

We used ITC to analyze the stoichiometry of the binding of $(CA)_4$ and $(ApN)_3$:CM and to confirm the oligomeric state of CA. ITC measurements were carried out on a ITC200 calorimeter (MicroCal/GE) at 20 ºC. Prior to the measurements, the proteins were dialyzed into buffer G/1 mM TCEP at 4 °C using Slide-A-Lyzer mini Dialysis Devices (MWCO 3.5 kDa, Thermo Fisher Scientific), with the buffer solution changed three times over the course of two days. During the ITC experiments, $(CA)_4$ (264.6 µM) or $E_{GFP}CA_{C17}$ (407.85 µM) was loaded into the syringe. The sample cell was filled with 300 µl of $(ApN)_3$:CM^{Nt} (35.3 µM) and the reference cell contained buffer. For each measurement point, 10 µl of $(CA)_4$ or $E_{GFP}CA_{C17}$ was injected at time intervals of 3 min. Data were analyzed using the software Origin 2021, and fitted to a one-site binding model.

2.4.7. Turbidity assay

Homo-demixing of $(ApN)_3$:CM, $(ApN)_3$:CM^{Nt}, $(CM)_3$, and condensate formation of $(ApN)_3$:CM, $(ApN)_3$:CM^{Nt}, $(ApN)_4$, $(CM)_3$ or CM^{Ct} with Rubisco, and $(ApN)_3$:CM, $(ApN)_3$:CM^{Nt}, $(ApN)_4$ or $(CM)_3$ with $(CA)_4$, and $(ApN)_3$:CM with $(CA\Delta C)_4$ were monitored by light scattering. Measurements were performed at 25 °C in buffer H (50 mM Tris-HCl pH 8.0/50 mM KCl/10 mM $Mg(OAc)_2$) containing 5 mM DTT or buffer G/5 mM DTT as indicated in figure legends. All proteins were prepared as 10-fold concentrated stock solutions and diluted to the concentrations indicated in the figures or figure legends in the reaction system

(100 µL). Carry-over of salt was accounted for in the assay buffer. For the formation of condensates with Rubisco or $(CA)_4$, Rubisco or $(CA)_4$ was first added to the assay buffer pre-blanked in an ultra-micro quartz cuvette (chamber volume 100 µL, pathlength 10 mm; Hellma), followed by the addition of other proteins and rapidly mixing by vortexing. Absorbance at 340 nm was then monitored over time on a V-560 UV/VIS Spectrophotometer (Jasco) set to 25 °C. Data plotted using Origin 2021.

2.4.8. Fluorescence microscopy assay

Fluorescence microscopy assay was performed by Kun Zang.

2.4.8.1. Condensate formation analyzed by fluorescence microscopy

The proteins intended for phase separation analysis by microscopy were labeled with a fluorescent dye at the N-terminus. Rubisco holoenzyme was labeled with Alexa Fluor 568 NHS ester (Thermo Fisher Scientific) ($RbcLS_{AF568}$) following the manufacturer's instructions to yield approximately 6.5 dye molecules per Rubisco holoenzyme. Prior to use, the purified CM^{Ct}, $(CM)_3$, and $(CA)_4$ proteins were reduced by treatment with 1 M DTT (final concentration of 5 mM) and then incubated on ice for 30 min. $(ApN)_4$, $(ApN)_3$:CM, $(ApN)_3$:CM^{Nt} and CM^{Ct} were labeled with the fluorophore Alexa Fluor 405 NHS ester (Thermo Fisher Scientific) at ~1.3, ~1.6, ~1.2 and ~0.7 dye molecules per protein, respectively. $(CM)_3$ and CM^{Ct} were labeled with the fluorophore Alexa Fluor 488 NHS ester (Thermo Fisher Scientific) at ~3.7 and ~0.9 dye molecules per protein, respectively, while $(CA)_4$ and CM^{Ct} were labeled with the fluorophore Alexa Fluor 647 NHS ester (Thermo Fisher Scientific) at ~2.7 and ~1.8 dye molecules per protein, respectively. Each labeled protein was mixed with the unlabeled protein at a ratio of 1:10. Proteins were prepared as 10-fold concentrated stock solutions and added to the reaction system (20 µL) in buffer H/5 mM DTT or buffer G/5 mM DTT, and the final concentrations of proteins were as indicated in figures. Carry-over of salt was accounted for in the assay buffer. Proteins were combined as indicated in the figure or figure legends in 0.5 mL Protein LoBind Tubes (Eppendorf) at 25 °C using a thermomixer (Eppendorf Thermomixer Comfort), reactions were then transferred to an uncoated chambered coverslip (µ-Slide angiogenesis 15 well; Ibidi) and incubated for an additional 5 min before analysis. Samples were illuminated with the 405 nm diode laser, 488 nm argon laser and white light lasers (NKT superK EXTREME) (at 570 nm and 647 nm) for fluorescence imaging. Images were acquired by focusing on the bottom of the plate using a Leica TCS SP8 AOBS confocal laser scanning microscope (HCX PL APO 63×/1.2 water objective, PMT detector).

The software Fiji (Schindelin et al., 2012) was used to generate the images of condensates. The image contrast was adjusted to optimize the visibility of phase-separated condensates and a Gaussian blur filter (sigma = 2 pixels) was applied to reduce pixel noise and smoothen the images.

2.4.8.2. Fluorescence recovery after photobleaching (FRAP)

FRAP experiments were carried out with a Leica TCS SP8 AOBS confocal laser scanning microscope equipped with an HCX PL APO 63×/1.2 water objective and a photomultiplier tube (PMT) detector. The samples were prepared as described above. Proteins were diluted to the following final concentrations in the reaction system (20 µL): CM^{Ct}_{AF405} (2 µM) or $(CM)_{3AF488}$ (0.25 µM) was mixed with unlabeled Rubisco ($RbcL_8S_8$ holoenzyme, 1 µM), respectively, and incubated in buffer G/5 mM DTT in 0.5 mL Protein LoBind Tubes (Eppendorf), accounting for salt carry-over in the assay buffer. FRAP analysis of Rubisco was performed by mixing $Rubisco_{AF405}$ ($RbcL_8S_8$ holoenzyme, 1 µM) with unlabeled CM^{Ct} (2 µM) and unlabeled $(CM)_3$ (0.25 µM). After incubation for 5 min at 25 °C using a thermomixer (Eppendorf Thermomixer Comfort), reactions were transferred to an uncoated chambered coverslip (µ-Slide angiogenesis 15 well; Ibidi) and incubated for an additional 15 min before analysis. Images were acquired in a single focal plane before and 5 min after photobleaching, with a 5-second (s) time interval. Bleaching was performed with a bleach point model with either a 405 nm diode laser (3% intensity) or a 488 nm argon laser (100% intensity) and repeated three times with a dwell time of 100 milliseconds (ms). Image analysis was conducted using Fiji software (Schindelin et al., 2012).

2.4.9. Cryo-EM sample preparation and data processing

2.4.9.1. Sample preparation and data collection

Unless otherwise specified, holey carbon supported copper or gold grids (Quantifoil R1.2/1.3 300 mesh, Quantifoil Micro Tools) were plasma-cleaned for 30 s in a PDC-32G-2 (Harrick Plasma) before using (within 30 min) (see Table **2.8**). Samples were diluted to working concentrations immediately before being applied to the glow-discharged grids. Then, a 3.5 µl aliquot of each sample was applied to a glow-discharged grid, followed by semiautomatically blotting for 3 s at a blot force of 4 and plunge-freezing into liquid ethane cooled by liquid N_2 using a Vitrobot Mark IV (Thermo Fisher Scientific) at 25 °C and 90% humidity.

The cryo-EM data collection on the Titan Krios (Thermo Fisher Scientific) transmission electron microscope (TEM) was performed by Daniel Bollschweiler and Tillman Schäfer.

Table 2.8 Preparation of cryo-EM grids

Sample	**Concentrations (µM)**	**Buffer**	**Grids**	**Movies**
$(ApN)_4$	6.2	Buffer F/5 mM DTT	Quantifoil R1.2/1.3 300 mesh, copper	628
$(SIIApN)_4$	6.2	Buffer F/5 mM DTT	Quantifoil R1.2/1.3 300 mesh, copper	720
$(ApN)_4$ polylysine treated	4 or 2.5 or 2.5	Buffer F/5 mM DTT	Quantifoil R1.2/1.3 300 mesh, copper	1,158 or 7,510 or 8,591
$(ApN)_4$	6.2	Buffer F/5 mM DTT	Quantifoil R1.2/1.3 400 mesh, gold (graphene oxide)	428
$(ApN)_3$:CM	4	Buffer I	Quantifoil R1.2/1.3 300 mesh, gold	1,953
$(SIIApN)_3$:CM-H_{10}	4	Buffer I	Quantifoil R1.2/1.3 300 mesh, gold	1,221
$(SIIApN)_3$:CM-H_{10} OG treated	8.2	Buffer I	Quantifoil R1.2/1.3 300 mesh, gold	736
$(SIIApN)_3$:CM-H_{10} CHAPSO treated	8.2	Buffer I	Quantifoil R1.2/1.3 300 mesh, gold	1,715

$(SIIApN)_3$:CM-H_{10}	8.2	Buffer I	Quantifoil R2/4 300 mesh, copper (graphene oxide)	978
$(SIIApN)_3$:CM-H_{10} oxidized	4	Buffer I	Quantifoil R2/1 300 mesh, copper	2,881
$(ApN\text{-}2A)_3$:CM	4	Buffer I	Quantifoil R1.2/1.3 300 mesh, gold	1,956
$(SIIApN\text{-}2A)_3$:CM-H_{10}	4	Buffer I	Quantifoil R1.2/1.3 300 mesh, gold	1,926
$(ApN\text{-}2S)_2$:CM	4	Buffer I	Quantifoil R1.2/1.3 300 mesh, gold	2,129
$(SIIApN\text{-}2S)_2$:CM-H_{10}	4	Buffer I	Quantifoil R1.2/1.3 300 mesh, gold	1,841

2.4.9.1.1. Sample preparation of $(ApN)_4$ and $(SIIApN)_4$

Purified $(ApN)_4$ was diluted to around 6.2 μM in buffer F/5 mM DTT and cryo-grids were prepared as described above. Cryo-grids were screened and the selected grid on stage was used for data collection directly on a Glacios TEM (Thermo Fisher Scientific), which was equipped with a K2 summit direct electron detector (Gatan) and operated at an accelerating voltage of 200 kiloelectron volt (keV). A total of 628 movies were automatically collected with the SerialEM software package (Mastronarde, 2005) at a nominal magnification of 36,000× (yielding a calibrated physical pixel size of 1.181 Å). A total dose of 63 electrons per Å^2 was accumulated over 40 frames within a total exposure time of 18.4 s, with defocus values ranging from –1.0 to –2.8 μm.

Purified $(SIIApN)_4$ was diluted to around 6.2 μM in buffer F/5 mM DTT and cryo-grids were prepared as described above. Cryo-grids were screened and the selected grid on stage was used for data collection directly on a Talos Arctica TEM (Thermo Fisher Scientific), which was equipped with a K3 direct electron detector (Gatan) and operated at an accelerating voltage of 200 keV. A total of 720 movies were automatically collected with the SerialEM software package (Mastronarde, 2005) at a nominal magnification of 36,000× (yielding a calibrated physical pixel size of 1.171 Å). A total dose of 63 electrons per Å^2 was accumulated over 40 frames within a total exposure time of 6.4 s, with defocus values ranging from –1.0 to –2.8 μm.

To overcome the preferred orientation bias of $(ApN)_4$ induced by the air-water interface, three approaches were tested: (1) Poly L-lysine hydrobromide (polylysine, Polysciences) treated grids (Lander et al., 2012): hydrophilized carbon grids were incubated with 5 µL of 0.1% polylysine for 90 s, then washed twice with 5 µL drops of ddH_2O, and dried completely at RT before applying $(ApN)_4$ (4 µM). (2) Support film-coated grids (Fan & Sun, 2022): graphene oxide on gold grids (Quantifoil R1.2/1.3 400 mesh, Electron Microscopy Sciences) were used for grids preparation with $(ApN)_4$ (6.2 µM). (3) Polylysine treatment combined with stage-tilt data collection (Tan et al., 2017): $(ApN)_4$ (2.5 µM) was applied to grids, with polylysine-treated grids used for untilted and 45° tilt data collection. For (1) and (2), 1,158 and 428 movies were collected, respectively, as described for $(SIIApN)_4$. For (3), cryo-grids were initially screened on a Talos Arctica TEM (Thermo Fisher Scientific). Selected grids were transferred to a Titan Krios TEM (Thermo Fisher Scientific), which was equipped with GIF Quantum Energy Filters (Gatan), and a K3 direct detector (Gatan), and operated at an accelerating voltage of 300 keV. A total of 7,510 movies were collected with untilted stage and 8,591 movies were collected for 45° tilted stage with the SerialEM software package (Mastronarde, 2005) at a nominal magnification of 105,000× (yielding a calibrated physical pixel size of 0.8512 Å). A total dose of 63 electrons per $Å^2$ was accumulated over 30 frames within a total exposure time of 2.7 s, with defocus values ranging from –0.9 to –2.2 µm.

2.4.9.1.2. Sample preparation of $(ApN)_3$:CM, $(SIIApN)_3$:CM-H_{10} and $(SIIApN)_3$:CM-H_{10} oxidized

Purified $(ApN)_3$:CM was diluted to around 4 µM in buffer I (50 mM Tris-HCl pH 8.0/100 mM KCl/5 mM DTT). Carry-over of salt was accounted for in the assay buffer. Cryo-grids were prepared and a total of 1,953 movies were collected as described for $(SIIApN)_4$.

Purified $(SIIApN)_3$:CM-H_{10} complex was diluted to 4 µM in buffer I. Carry-over of salt was accounted for in the assay buffer. Cryo-grids were prepared and a total of 1,221 movies were collected as described for $(ApN)_4$.

To prevent the dissociation of $(SIIApN)_3$:CM-H_{10} induced by the air-water interface, three approaches were tested: (1) Treatment with the non-ionic detergent Octyl-β-D-glucopyranoside (OG, Sigma-Aldrich) (Benton et al., 2018): $(SIIApN)_3$:CM-H_{10} was diluted to 8.2 µM in buffer I containing OG (final concentration: 0.008% (w/v)) prior to cryo-grids preparation. (2) Treatment with the zwitterionic detergent 3-(3-cholamidopropyl-dimethylammonio)-2-hydroxy-1-propanesulfonate (CHAPSO, Sigma-Aldrich) (Chen et al., 2019): $(SIIApN)_3$:CM-H_{10} was diluted to 8.2 μM in buffer I supplemented with CHAPSO (final concentration: 4 mM) prior to cryo-grids preparation. (3) Support film-coated grids (Fan & Sun, 2022): graphene oxide on copper grids (Quantifoil R2/4 300 mesh, Electron Microscopy Sciences) were used for grids preparation with $(SIIApN)_3$:CM-H_{10} (8.2 μM). For (1), (2), and (3), 736, 1,715, and 978 movies were collected, respectively, as described for $(SIIApN)_4$.

$(SIIApN)_3$:CM-H_{10} complex purified without DTT in the final SEC step, was pre-incubated with H_2O_2 (final concentration: 2 mM) in buffer F on ice for 30 min to mimic oxidation. Then the sample was diluted to 4 μM in buffer G, incubating for 10 min at RT. Carry-over of salt was accounted for in the assay buffer. Samples were prepared with copper grids (Quantifoil R2/1 300 mesh, Plano) and a total of 2,881 movies were collected as described for $(SIIApN)_4$.

2.4.9.1.3. Sample preparation of $(ApN\text{-}2A)_3$:CM, $(SIIApN\text{-}2A)_3$:CM-H_{10}, $(ApN\text{-}2S)_2$:CM, and $(SIIApN\text{-}2S)_2$:CM-H_{10}

Purified $(ApN\text{-}2A)_3$:CM, $(SIIApN\text{-}2A)_3$:CM-H_{10}, $(ApN\text{-}2S)_2$:CM, or $(SIIApN\text{-}2S)_2$:CM-H_{10} complex was diluted to around 4 μM in buffer I, respectively. Carry-over of salt was accounted for in the assay buffer. Cryo-grids were prepared and movies were collected as described for $(SIIApN)_4$ (see Table **2.8**).

2.4.9.2. Data processing

All of the datasets were processed with cryoSPARC version 4.2 (Punjani et al., 2017). The raw movies were imported into cryoSPARC for all subsequent processing steps (including patch motion correction, patch contrast transfer function estimation, blob picking, template picking, two-dimensional classification). The processing details are provided in Appendix **A.4**. The cryo-EM analysis was validated by converting the asymmetric unit of the crystal structure (PDB: 7D6C) (Sun et al., 2021) to a 10 Å resolution map using molmap (UCSF Chimera tool), followed by the generation of orientation-spanning reference projections in cryoSPARC. These projections are consistent with our experimental data and closely resemble the 2D classes of the dimer of hetero-trimers shown in Figures **3.9A,B** and **3.11C,D** and Appendix **A.4**.

2.5. Protein structure prediction

AlphaFold 3 (Abramson et al., 2024) was used to predict the structure of the $(ApN)_4$ and $(ApN)_3$:CM complexes. The protein sequences were obtained from UniProt and submitted to the AlphaFold 3 server for modeling. Four identical copies of the ApN sequence were input to generate the $(ApN)_4$ structure. For the $(ApN)_3$:CM complex, three ApN sequences and one CM sequence were used. Predicted structural model can be found in Chapters **3.1** and **3.3**, and Appendix **A.5**. The specific number of amino acids used for ApN and CM is detailed in the figure legends.

AlphaFold 3 provides several confidence metrics for assessing model quality. The predicted local distance difference test (pLDDT) provides a per-residue estimate of its confidence on a scale of 0-100, with higher values indicating greater reliability. The predicted aligned error (PAE) matrix reflects the confidence in the relative position and orientation of different domains present in the predicted structure, with higher values indicating higher predicted error and therefore greater uncertainty. The predicted distance error (PDE) assesses the error in the distance matrix of the predicted structure as compared to the true structure. The predicted template modeling (pTM) score evaluates overall folding accuracy. A score above 0.5 suggests a structure comparable to the native fold and the higher is better. The interface predicted template modeling (ipTM) measures the accuracy of subunit positioning within the complex. Scores of 0.8 or higher indicate high-quality predictions.

PyMOL (Schrodinger, 2015) was used to generate all structure visualizations by Kun Zang.

2.6. *In vivo* experiments

The *in vivo* experiments were performed by Nghiem D. Nguyen, Loraine M. Rourke and Jiwon Lee within the research groups of Benedict M. Long and G. Dean Price in Australia.

2.6.1. Cyanobacterial transformation

*Se*7942 transformations were performed by harvesting a seven-day-old lawn of cells grown on modified BG-11 plates containing 1% agar (Woodger et al., 2003). Cells were washed and resuspended in modified BG-11 until the OD_{730} reached ~5.0. Approximately 1 µg pDNA was added to 200 µL aliquots of cells and incubated at 22 °C in the dark for 24 h. Cells were then spread onto modified BG-11 agar plates supplemented with appropriate antibiotics (25 µg/mL kanamycin; 10 µg/mL spectinomycin) and 2 mM sodium thiosulfate. The agar plates were then incubated at 30 °C in 3% CO_2 (v/v in air) until single colonies appeared. To generate deletion cell lines, single colonies were transferred to fresh modified BG-11 plates and sub-cultured three times. Genetic segregation was then confirmed by diagnostic PCR.

2.6.2. Cell growth analysis

*Se*7942 cells and ApN mutants were first grown over 2-3 days in ~3.0% CO_2 and then diluted in fresh modified BG-11 medium to an OD_{730} of ~0.1 in triplicate 35 mL cultures. Cultures were grown at 30 °C with 80 µmol photons $m^{-2}s^{-1}$ and were bubbled with air. Hourly culture measurements (OD_{730}) were conducted over a period of 8 h, and the maximum growth rate was determined by calculating the slope of logarithmic regressions of the data and transformed into doubling times. Data were plotted in SigmaPlot v14.

2.6.3. Cell ultrastructure analysis

*Se*7942 cells were grown in modified BG-11 medium containing 20 mM HEPES-KOH pH 8.0 at 30 °C, 3% CO_2, and an illumination intensity of approximately 80 µmol photons $m^{-2}s^{-1}$ over 2-3 days. Cells in 10 mL culture were fixed for at least 4 h with 4% formaldehyde/2.5% glutaraldehyde (Electron Microscopy Sciences) in PBS (pH 7.4). Cells were then centrifuged at 5,000 × g for 5 min, washed three times with PBS, and then post-fixed in 1% (w/v) osmium tetroxide for 4 h. Fixed cells were dehydrated through a graded ethanol series and embedded in LR-white resin (ProSciTech). Ultrathin sections were obtained using an ultramicrotome (Leica EM UC7) and stained with 2% (w/v) uranyl acetate, followed by lead citrate. The sections were then viewed using a JEM F200 TEM (JEOL) operated at 80 keV. Images were taken with a Rio 16 camera (Gatan). Data were plotted in OriginPro 2020.

3. Results

The goal of this study was to clarify the role of the shell adaptor protein in linking pro-β-carboxysome assembly with carboxysome shell formation. We began with the demonstration that the recombinantly expressed ApN protein, forming a $(ApN)_4$ complex alone is not recruited to a reconstituted pro-carboxysome condensate. Next, we co-expressed CM and ApN using a bicistronic plasmid, yielding a hetero-tetrameric $(ApN)_3$:CM complex capable of incorporation into the mature pro-β-carboxysome via multiple protein-protein interactions. *In vivo* complementation of the ΔK2-O *Se*7942 strain with pSE4*ccmK2-O* carrying ApN mutants demonstrated that the $(ApN)_3$:CM hetero-complex is essential for β-carboxysome biogenesis. Finally, analysis of *in vitro* pro-β-carboxysome reconstitution by fluorescence microscopy analysis revealed a stepwise assembly mechanism, with the $(ApN)_3$:CM hetero-complex positioning at the rim of the pro-β-carboxysome condensate as a prerequisite for the recruitment of shell proteins and initiation of shell formation.

The main findings of this study are presented in Chapter **3**, while the sequence alignment of proteins, protein purifications, SEC-MALS analysis of proteins, additional attempts to obtain the cryo-EM structure, and the predicted structural models are detailed in the Appendix.

3.1. ApN forms a tetramer in solution

The protein ApN is conserved in sequence and function (Kinney et al., 2012). Analysis of the ApN sequence from different species showed that the β-helix domain is followed by a presumably flexible region of ~50 poorly conserved amino acids containing the EP at the C-terminus, and there are some conserved residues in both the N-terminal β-helix domain and the C-terminal EP (Figure **3.1A,B**). ApN serves as an adaptor, linking the carboxysome core and shell via its N-terminal β-helix domain (Kinney et al., 2012), which binds to the γCAL domain of CM (Cot et al., 2008), while the C-terminal EP interacts with shell proteins (Cai et al., 2016; Cameron et al., 2013; Kinney et al., 2012). Although the crystal structures of several shell proteins (Kerfeld et al., 2005; Tanaka et al., 2008), Rubisco (Newman et al., 1993), and of the N-terminal γCAL domain of CM (Figure **3.1A**) (Pena et al., 2010; Zang et al., 2021) have been solved, the structure of ApN was only determined in a hetero-trimeric complex consisting of two SUMO-tagged ApN (SUMO-ApN) protomers and one CM^{209} (residues 1-209) protomer, $(SUMO\text{-}ApN)_2{:}CM^{209}$. Due to the low abundance of ApN and the limited information on the stoichiometry of β-carboxysome components, the precise composition and role of ApN in β-carboxysome assembly remained unclear. Furthermore, ApN was reported to be prone to aggregation (Sun et al., 2021), which may hinder *in vitro* studies of pro-β-carboxysome assembly.

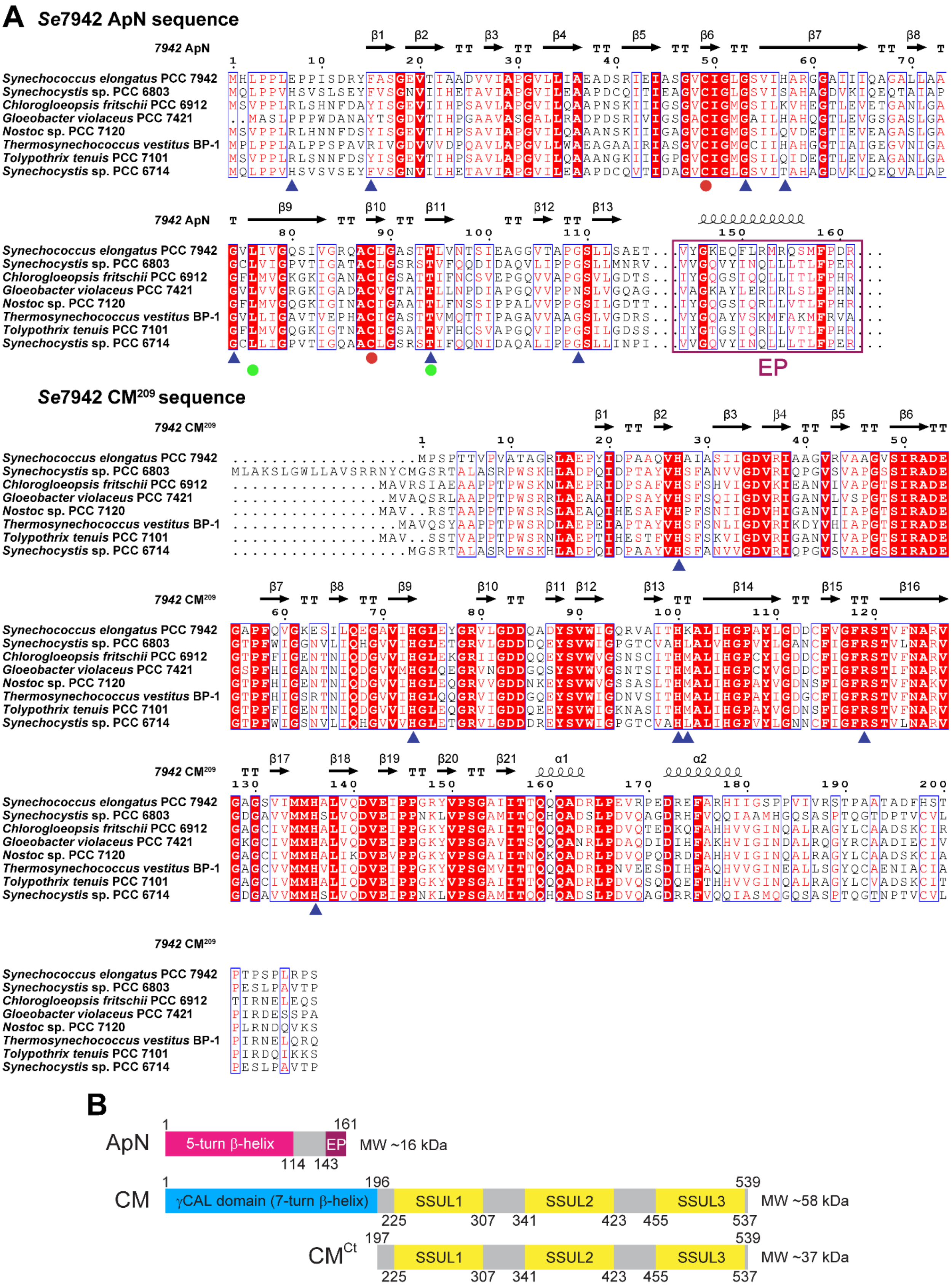


Figure 3.1 The sequence alignment and schematic representation of the shell adaptor protein ApN and the scaffolding protein CM.

(A) Multiple-sequence alignments of ApN and CM^{209}. The selected sequences of ApN and CM^{209} were aligned as described in Chapter **2.2**. For ApN, the presumably flexible region between the N-terminal β-helix domain and the C-terminal encapsulation peptide (EP) was omitted from the alignment. Full-length alignment of both ApN and CM are provided in Appendix **A.1**. The secondary structural elements of *Se*7942 ApN and CM^{209} are shown above the alignment. The contact residues between ApN and CM^{209}, as reported by (Sun et al., 2021), are indicated with blue triangles at the bottom. The conserved residues Cys49 and Cys88 are highlighted by red circles, and Leu76 and Thr94 are indicated by green circles, respectively. EP sequences are marked by a pink box. Identical residues are shown in white text on a red background, and similar residues are shown in red. Homologous regions are outlined in blue. **(B)** Domain structures of the shell adaptor protein (ApN), the full-length scaffolding protein (CM) and its truncated form (CM^{Ct}). Amino acid numbers and approximate molecular weights of each domain are indicated. The SSUL modules of CM and CM^{Ct} are numbered SSUL1 to SSUL3 from N- to C-terminus. EP, encapsulation peptide, SSUL, small subunit-like.

Through recombinant expression of *Se*7942 ApN in *E. coli* with either a cleavable His6ubi tag or a Strep-tag II fused at the N-terminus, we were able to purify both the non-tagged ApN (16.3 kDa protomer) and the Strep-tagged ApN (SIIApN; 17.7 kDa protomer) in the presence of reducing agent. Details of the protein expression and purification procedures are provided in Chapter **2.4.1** and Appendix **A.2**. Size-exclusion chromatography coupled to multiangle static light scattering (SEC-MALS) analysis of ApN indicated a molar mass of ~63.5 kDa (theoretical mass: 65.31 kDa) (Figure **3.2A**), suggesting that ApN is a tetramer in solution. SIIApN also showed a similar oligomeric state (~69.3 kDa; theoretical mass: 70.85 kDa) (Figure **3.2A**), confirming that ApN is tetrameric in solution.

Next to obtain insight into the structure of ApN, we performed cryo-EM single-particle analysis on both untagged ApN and SIIApN under reducing conditions to mimic the cytosolic environment in cyanobacteria. Reference-free two-dimensional (2D) class averages revealed a four-fold symmetric complex composed of four triangular-shaped protomers for both untagged ApN (Figure **3.2B**) and SIIApN (Figure **3.2C**), consistent with the SEC-MALS analysis showing that ApN forms a tetramer. The presence of the N-terminal Strep-tag II does not affect the conformation of ApN, and no density corresponding to the tag was observed (Figure **3.2C**). The top twenty 2D class averages of $(ApN)_4$ and $(SIIApN)_4$ are provided in Appendix **A.4.1**. However, the 2D class averages show an uneven particle distribution, a single preferred orientation, and a high background caused by thick ice. To address these various problems, especially the orientation bias of particles induced by the air-water interface during cryo-EM sample preparation, we tested several commonly used strategies. These strategies included pretreating grids with polylysine (Lander et al., 2012), using support film-coated grids (Fan & Sun, 2022), and collecting stage-tilt data on polylysine-treated grids (Tan et al., 2017). Results of these tests are provided in Appendix **A.4.1**. However, none of these approaches was effective in overcoming the challenges posed by the low molecular weight $(ApN)_4$ complex (~65 kDa; ~40 kDa of the rigid body). Consequently, critical information was missing in Fourier space, making 3D reconstruction of the ApN proteins unattainable.

The triangular shape of each ApN protomer observed in the 2D class averages aligns with the end-on view of ApN in the crystal structure of the hetero-trimeric (SUMO-$ApN)_2$:CM^{209} complex (Sun et al., 2021). The tetrameric assemblies seen in the 2D class averages closely resemble the tetrameric arrangement of ApN (residues 11-114) predicted by AlphaFold 3 (Figure **3.2D**), yielding pTM and ipTM scores of 0.58 and 0.55, respectively (Xu & Zhang, 2010). The full-length ApN structural model predicted by AlphaFold 3 is provided in Appendix **A.5.1**. Aligning the cryo-EM 2D class averages with the AlphaFold 3 predicted structural model of $(ApN)_4$ showed that the end-on view in the 2D class averages places the C-terminus at the top and the N-terminus at the bottom (Figure **3.2B,D**), with the wall A to wall C interfaces between protomers following an anti-clockwise direction. Conversely, with the N-terminus on top, the wall A to wall C interfaces follow a clockwise orientation (Figure **3.2D**). Several particles of $(SIIApN)_4$ show an end-on view with the N-terminus on top (Figure **3.2C**).

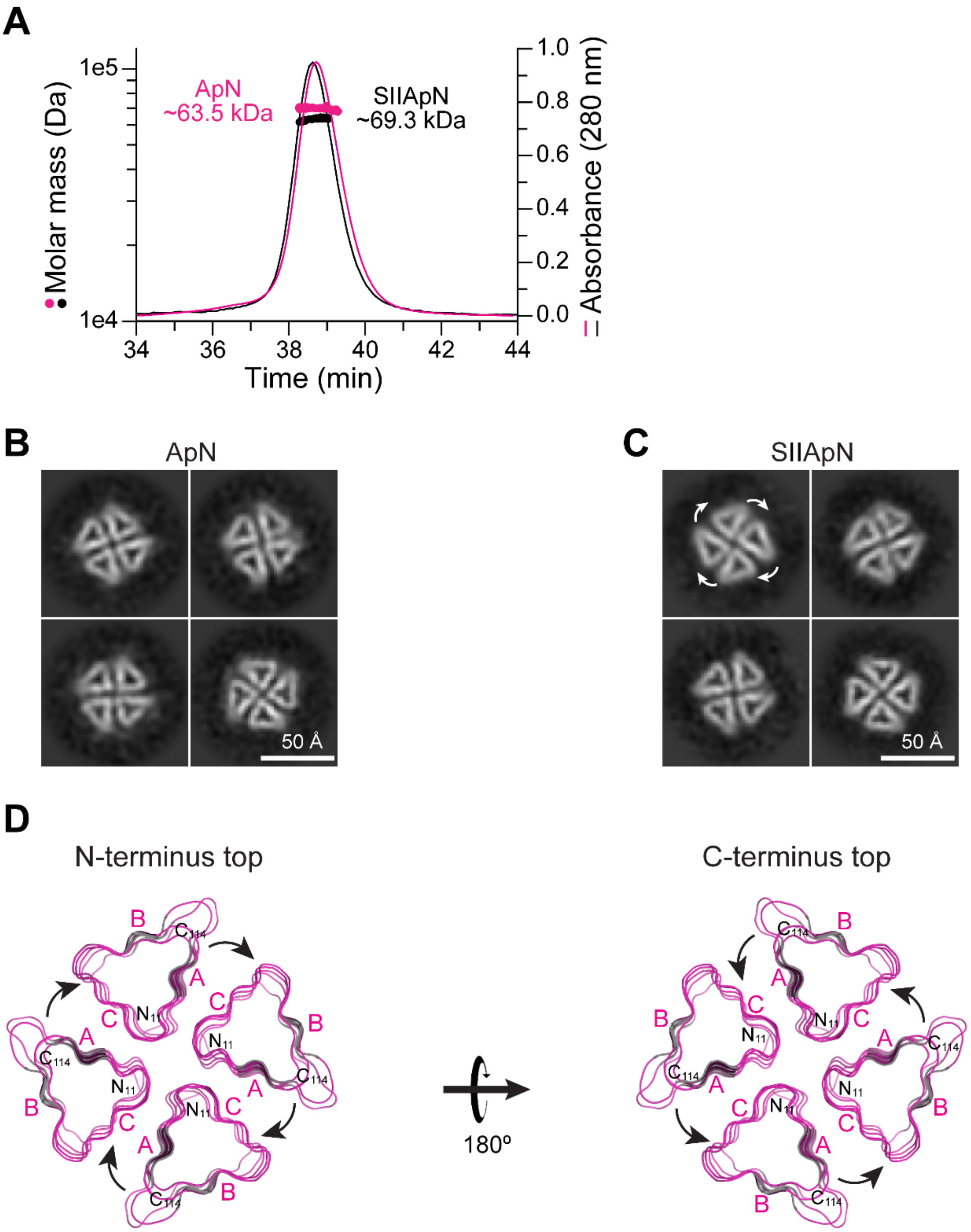


Figure 3.2 Oligomeric state of ApN.

(A) SEC-MALS analysis of purified ApN and SIIApN. The theoretical mass of tetrameric SIIApN is 70.85 kDa and tetrameric ApN is 65.31 kDa. The measured molar masses are indicated. Representative data from two independent experiments are shown (n=2). **(B-C)** Cryo-EM single-particle analysis of $(ApN)_4$ **(B)** and $(SIIApN)_4$ **(C)**. Shown are representative 2D class averages. Scale bar, 50 Å. The orientation of the wall A-wall C interface is indicated by white arrows. **(D)** Structural model of $(ApN)_4$ (residues 11-114) predicted by AlphaFold 3, is shown in end-on views with the ApN protomers in magenta in ribbon representation. The N- and C-terminal residues are indicated. Left: View of $(ApN)_4$ with N-termini on the top (clockwise orientation of wall A to wall C). Right: View of $(ApN)_4$ with C-termini on the top (anti-clockwise orientation of wall A to wall C). The three walls of the triangular-shaped ApN protomer are labeled as A, B, and C. The orientation of the wall A-wall C interface is indicated by black arrows.

A hydrophobicity moment plot analysis (Eisenberg et al., 1984) of the wall A to wall C interface revealed that several hydrophobic residues are exposed on the surfaces of walls A and C, stabilizing the interaction (Figure **3.3A,B**), while the surface-exposed residues of the four B walls surrounding the tetramer and the loops connecting to walls A and C, contain several charged residues (Figure **3.3A**). Notably, within the $(ApN)_4$ complex, wall A consistently packs against wall C (Figure **3.2D**), which contains two highly conserved cysteines, Cys49 and Cys88 (Figures **3.1A** and **3.3B**).

We note that the N-terminal domain of ApN (residues 1-118) adopts a five-turn left-handed β-helical barrel structure (Sun et al., 2021), which is similar to the seven-turn left-handed β-helical barrel structure of the N-terminal γCAL domain of CM (residues 1-181) (Pena et al., 2010; Zang et al., 2021) (Figure **3.3C**). Notably, no density corresponding to the C-terminal domains containing the EP helices was observed in the 2D class averages of $(ApN)_4$ or the reported crystal structure of the hetero-trimeric $(SUMO\text{-}ApN)_2{:}CM^{209}$ complex. The low confidence prediction by AlphaFold 3 of the C-terminal domains containing the EP helices in $(ApN)_4$ suggests conformational flexibility (see Appendix **A.5.1**), which is consistent with the absence of observable density.

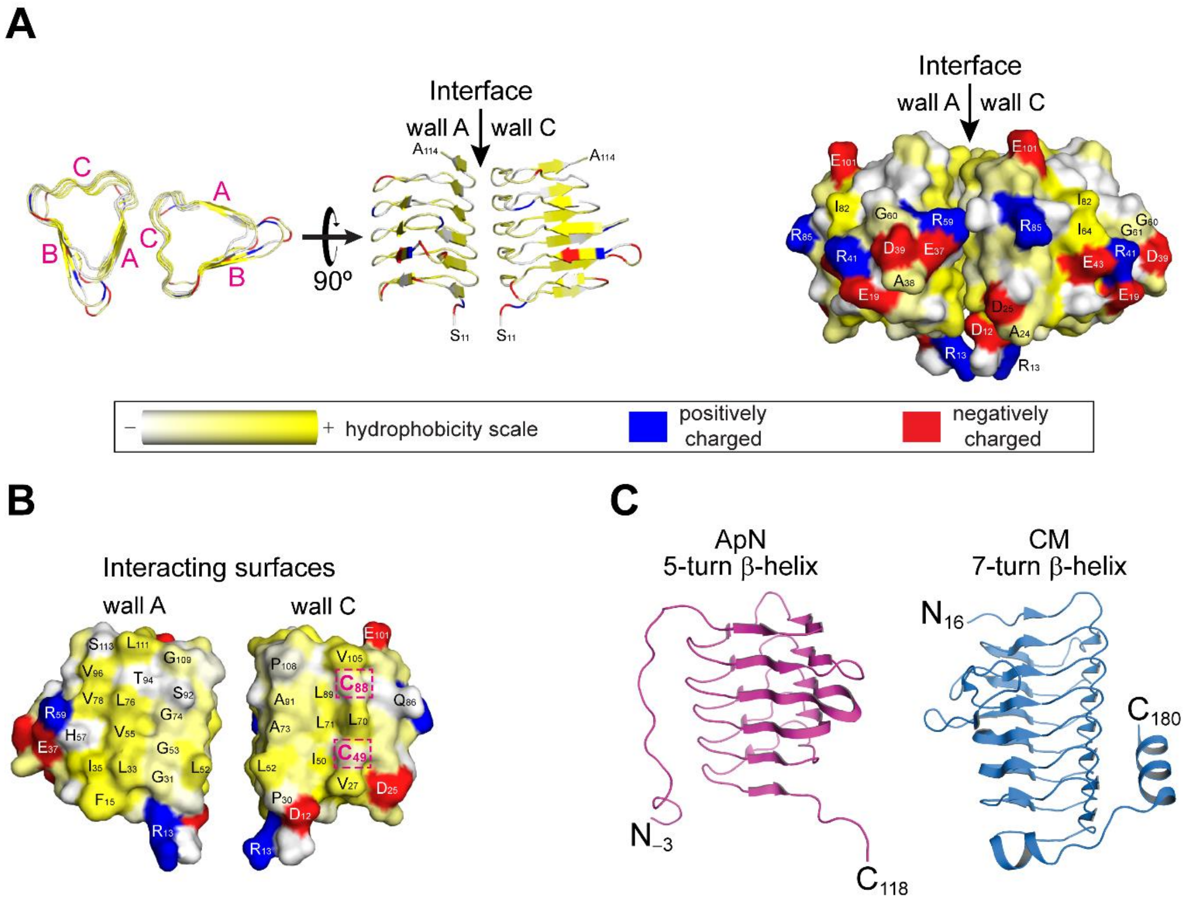


Figure 3.3 Structural analysis of ApN.
(A-B) The interaction surface between walls A and C of $(ApN)_4$. Two adjacent ApN protomers are shown in ribbon and surface representations **(A)**, and walls A and C are shown in surface representation, exposed to the viewer **(B)**. Two conserved cysteines, C49 and C88 are indicated by dashed boxes. Residue coloring is based on their hydrophobicity moment and charge. Hydrophobic, yellow; positive charge, blue; negative charge, red. **(C)** The N-terminal domains of ApN and CM have similar β-helical domain structures. Left: Crystal structure of the N-terminal 5-turn β-helix domain of ApN (residues -3-118; PDB: 7D6C). Right: Crystal structure of the N-terminal 7-turn β-helix γCAL domain of CM (residues 16-180; PDB: 7O4Z). Both are shown in ribbon representation.

3.2. Tetrameric ApN does not participate in pro-β-carboxysome formation

To investigate how $(ApN)_4$ interacts with the other components of the pro-β-carboxysome before shell assembly, we performed pull-down assays to analyze the possible interactions between $(ApN)_4$ and these components. $(SIIApN)_4$ was incubated with either $(CM)_3$, Rubisco ($RbcL_8S_8$), $(CA)_4$ or CM^{Ct}, followed by pull-down analysis. In all cases, only SIIApN was enriched in the pull-down fractions (Figure **3.4A**), indicating that $(ApN)_4$ does not directly interact with the other proteins of the pro-β-carboxysome.

We next took advantage of a sedimentation assay to test whether $(ApN)_4$ can be recruited into biomolecular condensates in which Rubisco and $(CA)_4$ were encapsulated by the scaffolding proteins $(CM)_3$ and CM^{Ct}. These four proteins are known to co-assemble into droplet-shaped condensates and sediment into the pellet fraction (Zang et al., 2021) (Figure **3.4B**, lanes 1-3). Given the low native concentration of ApN in *Se*7942 as previously reported (Liu et al., 2021; Sun et al., 2024), recruitment was first tested using 0.25 µM $(ApN)_4$. However, no ApN could be detected in the pellet fraction (Figure **3.4B**, lanes 4-6). Increasing the concentration of $(ApN)_4$ ten-fold to 2.5 µM also failed to show any co-sedimentation of ApN with the other four proteins (Figure **3.4B**, lanes 7-9). To exclude the possibility that the amount of $(ApN)_4$ recruited into the condensates might be present at levels below the detection limit of SDS-PAGE, we performed florescence microscopy analysis using Alexa Fluor (AF) labeled proteins: $Rubisco_{AF568}$, $(CM)_{3AF488}$, CM^{Ct}_{AF488}, $(CA)_{4AF647}$, and $(ApN)_{4AF405}$, each mixed at a 1:10 ratio with the respective unlabeled protein. No AF405 fluorescence signal of $(ApN)_4$ was detected (Figure **3.4C**), confirming that $(ApN)_4$ does not colocalize with Rubisco, $(CM)_3$, CM^{Ct} and $(CA)_4$ in phase-separated droplets.

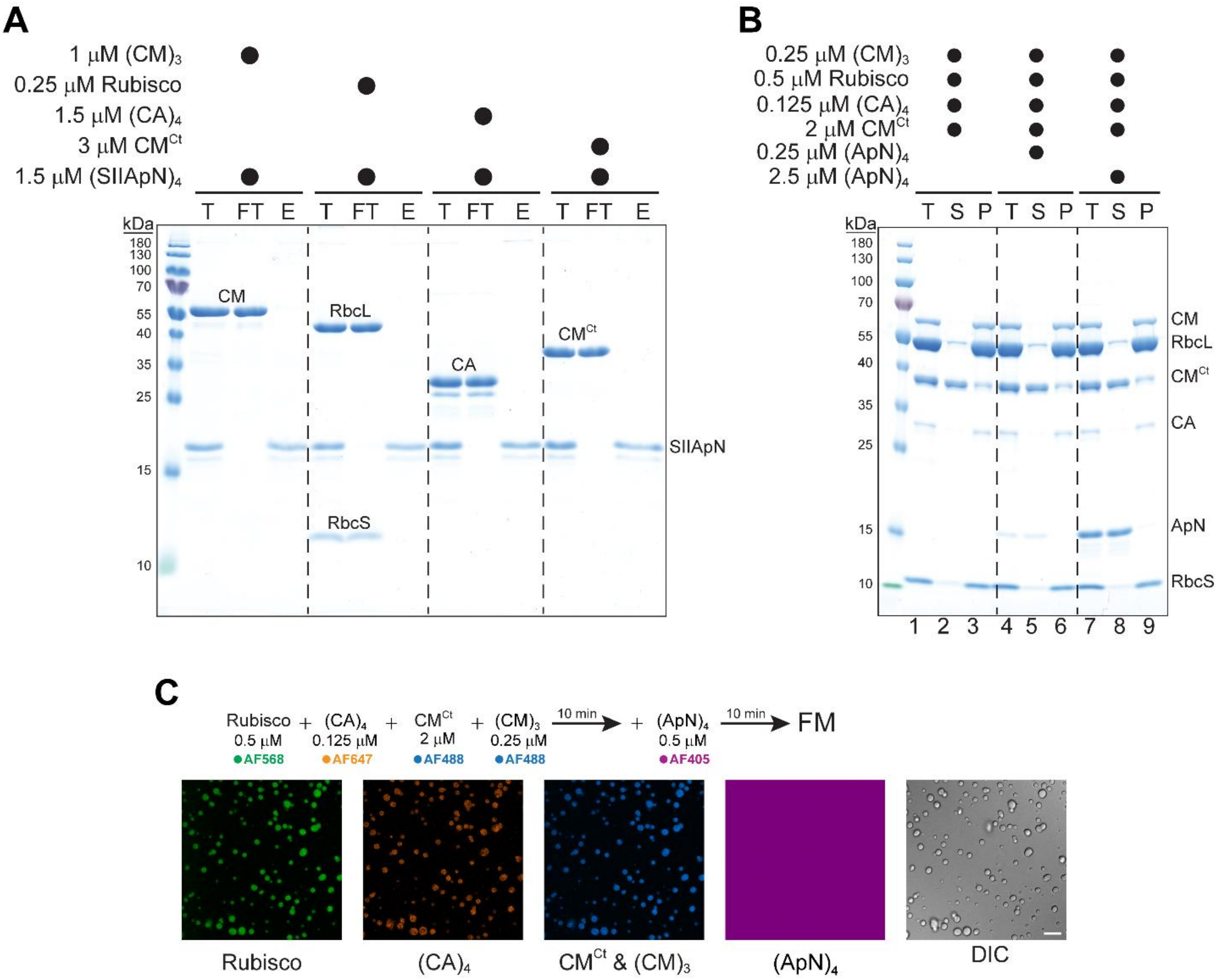


Figure 3.4 Undetected interactions between ApN and pro-carboxysome core proteins. **(A)** $(ApN)_4$ does not interact with other pro-carboxysome proteins. A pull-down assay was performed with 1.5 µM $(SIIApN)_4$ in the presence of the following pro-carboxysome proteins: 1 µM $(CM)_3$, 0.25 µM Rubisco, 1.5 µM $(CA)_4$, or 3 µM CM^{Ct}. The assay was analyzed by SDS-PAGE. T, Total; FT, Flow-through; E, Eluate. A representative Coomassie-blue stained gel from two independent experiments is shown (n=2). **(B)** $(ApN)_4$ fails to phase-separate with other pro-carboxysome proteins into condensate. A sedimentation assay was performed with 0.5 µM Rubisco, 2 µM CM^{Ct}, 0.25 µM $(CM)_3$, 0.125 µM $(CA)_4$, and 0.25 or 2.5 µM $(ApN)_4$. Samples were fractionated into total (T), supernatant (S) and pellet (P) and analyzed by SDS-PAGE and Coomassie staining. A representative Coomassie-blue stained gel from three independent experiments is shown (n=3). **(C)** Condensate formation analyzed by fluorescence microscopy. 0.5 µM Rubisco, 0.125 µM $(CA)_4$, 2 µM CM^{Ct}, 0.25 µM $(CM)_3$, and 0.5 µM $(ApN)_4$ were mixed as indicated. The respective labeled protein (fluorophore as indicated) was used at a 1:10 ratio with the unlabeled protein. DIC, differential interference contrast. Representative data from two independent experiments are shown (n=2). Scale bar, 10 µm.

3.3. Formation of CM-ApN hetero-complex

Since our findings suggested that tetrameric ApN does not interact with pro-carboxysome proteins $(CM)_3$, CM^{Ct}, Rubisco and $(CA)_4$, we sought to understand how the essential ApN is recruited into the pro-carboxysome to trigger shell formation and drive β-carboxysome maturation. Previous studies have shown that ApN interacts with the N-terminal γCAL domain of CM (Cot et al., 2008; Pena et al., 2010), as well as with the crystal structure of the hetero-trimeric complex $(SUMO\text{-}ApN)_2{:}CM^{209}$ (Sun et al., 2021). We therefore investigated if ApN participates in pro-β-carboxysome formation through such a hetero-complex formed with CM and, if so, how the hetero-complex contributes to the assembly process. Notably, the *ccmN* gene is typically located downstream of the *ccmM* gene in the *ccm* operon of β-cyanobacteria (Axen et al., 2014; Kerfeld & Melnicki, 2016; Rae et al., 2013) (Figure **3.5A**), suggesting a conserved genetic organization and potential functional linkage. CM protomers were reported to form trimers (Pena et al., 2010; Zang et al., 2021), and based on our findings, the ApN protomers can assemble into tetramers. To test if the subunits of these complexes can exchange to form CM-ApN hetero-complexes, a pull-down assay was performed using mixtures of $(SIIApN)_4$ and $(CM)_3$. The protein mixtures were incubated at either 25 °C or 37 °C for 30 min in the presence of DTT, but no formation of hetero-complex was observed (Figure **3.5B**, lanes 1-6). Given that the interior of mature carboxysomes is oxidizing (Chen et al., 2013; Pena et al., 2010; Price et al., 1992), we also performed pull-down assays using proteins oxidized by 2 mM hydrogen peroxide (H_2O_2) to test whether oxidative conditions might promote subunit exchange to form a hetero-complex, but again, no hetero-complexes were detected (Figure **3.5B**, lanes 7-12).

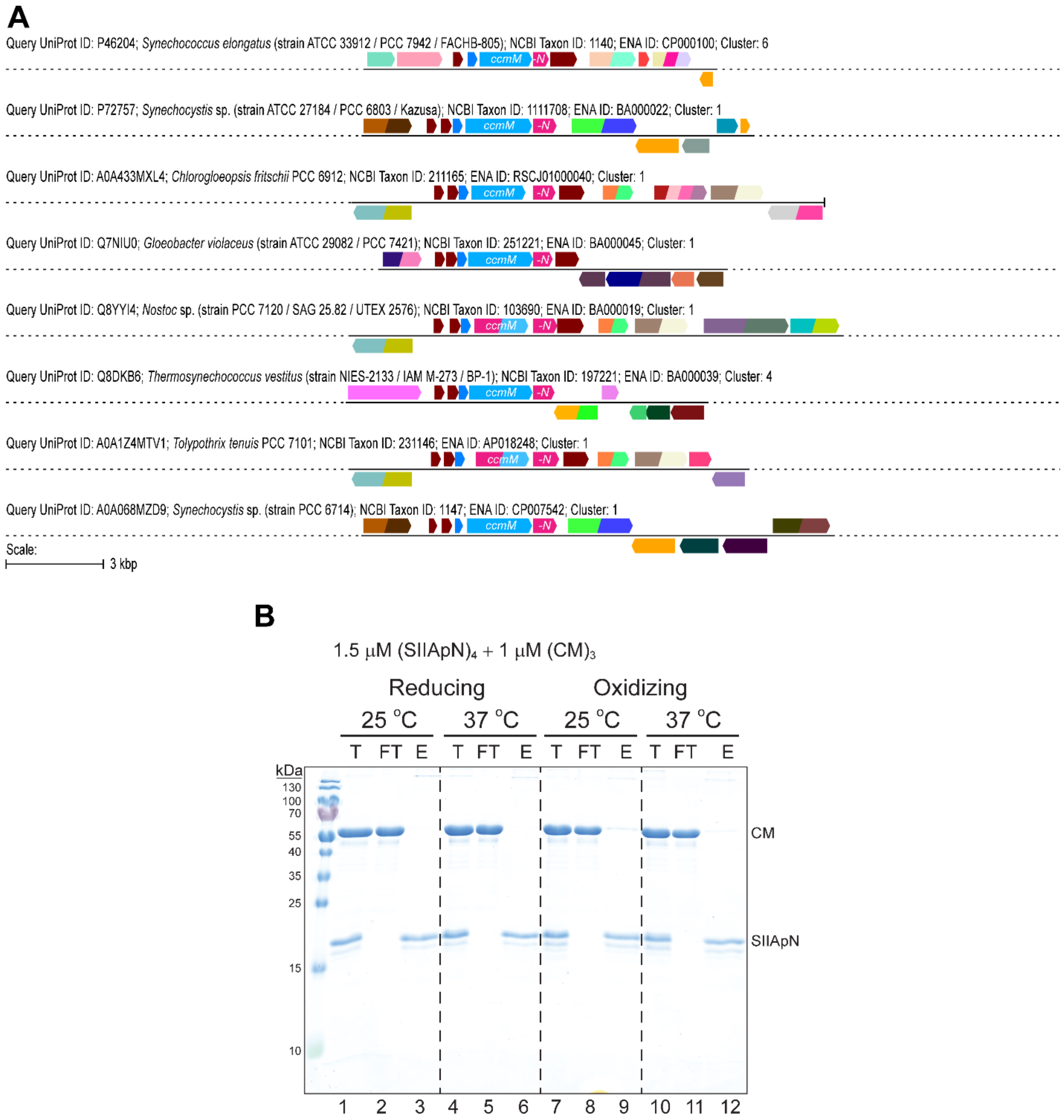


Figure 3.5 The possibility of hetero-complex formation by ApN and CM.
(A) Genome neighborhood of *ccmN* (ApN). Eight representative β-cyanobacterial species, referenced in the sequence alignment in Chapter **2.2** and selected from 543 genomes are shown to illustrate the localization of *ccmN* (ApN, magenta) in the genome. For each species, the five genes located upstream and downstream of *ccmN* are displayed. *ccmM* and *ccmN* are shown as indicated. In *Nostoc* sp. PCC 7120 and *Tolypothrix tenuis* PCC 7101, *ccmM* is shown with Pfam00101 (Rubisco_small) and Pfam00132 (Hexapep) domains. **(B)** No subunit exchange occurs between $(ApN)_4$ and $(CM)_3$. SDS-PAGE analysis of a pull-down assay using 1.5 µM $(SIIApN)_4$ and 1 µM $(CM)_3$ under reducing (lanes 1-6) or oxidizing (lanes 7-12) conditions following incubation at 25 °C or 37 °C for 30 min. The assay was analyzed by SDS-PAGE. T, Total; FT, Flow-through; E, Eluate. A representative Coomassie-blue stained gel from two independent experiments is shown (n=2).

The absence of subunit exchange between $(CM)_3$ and $(ApN)_4$ suggested that the hetero-complex likely forms in a manner coupled to protein synthesis (Shieh et al., 2015). We then used a bicistronic plasmid for the recombinant expression of the hetero-complexes CM-ApN, (CM-H_{10})-SIIApN, and (CM^{Nt}-H_{10})-SIIApN (CM^{Nt}: residues 1-198 of CM). For CM-ApN, a His6ubi tag was fused to the N terminus of the *ccmN* gene (*ApN*). For (CM-H_{10})-SIIApN and (CM^{Nt}-H_{10})-SIIApN, a Strep-tag II was fused to the N-terminus of the *ccmN* gene (*SIIApN*), and 10 histidines (H_{10}) were fused to the C-terminus of the full-length *ccmM* gene (*CM-H10*) or the truncated gene CM^{Nt} lacking the SSUL modules (CM^{Nt}-*H10*) (Figure **3.6A**). Strong ribosome binding sites (RBS) were inserted upstream of both genes to ensure efficient production of the hetero-complexes. This contrasts with the *in vivo* operon of *Se*7942, in which the RBS upstream of the *ccmN* gene is much weaker than that of the *ccmM* gene, resulting in high concentrations of $(CM)_3$ and CM^{Ct} and relatively low abundance of ApN. We successfully purified the hetero-complexes CM-ApN, (CM-H_{10})-SIIApN, and (CM^{Nt}-H_{10})-SIIApN, as demonstrated by SDS-PAGE. Details of the purification process are provided in Chapter **2.4.1.2**, and Appendix **A.2.3** and **A.2.4**.

SEC-MALS measurements in the presence of DTT revealed molar masses of ~103.8 kDa, ~111.6 kDa, and ~77.9 kDa for the hetero-complexes CM-ApN, (CM-H_{10})-SIIApN, and (CM^{Nt}-H_{10})-SIIApN, respectively (theoretical masses: 106.81 kDa, 112.69 kDa, and 75.83 kDa, respectively) (Figure **3.6B**), suggesting that these complexes contain three ApN or SIIApN protomers and one protomer of CM, CM-H_{10} or CM^{Nt}-H_{10}. This stoichiometry differs from that previously reported upon the co-expression of SUMO-ApN and CM^{209}, which formed a hetero-trimer with two SUMO-ApN protomers and one CM^{209} protomer (Sun et al., 2021).

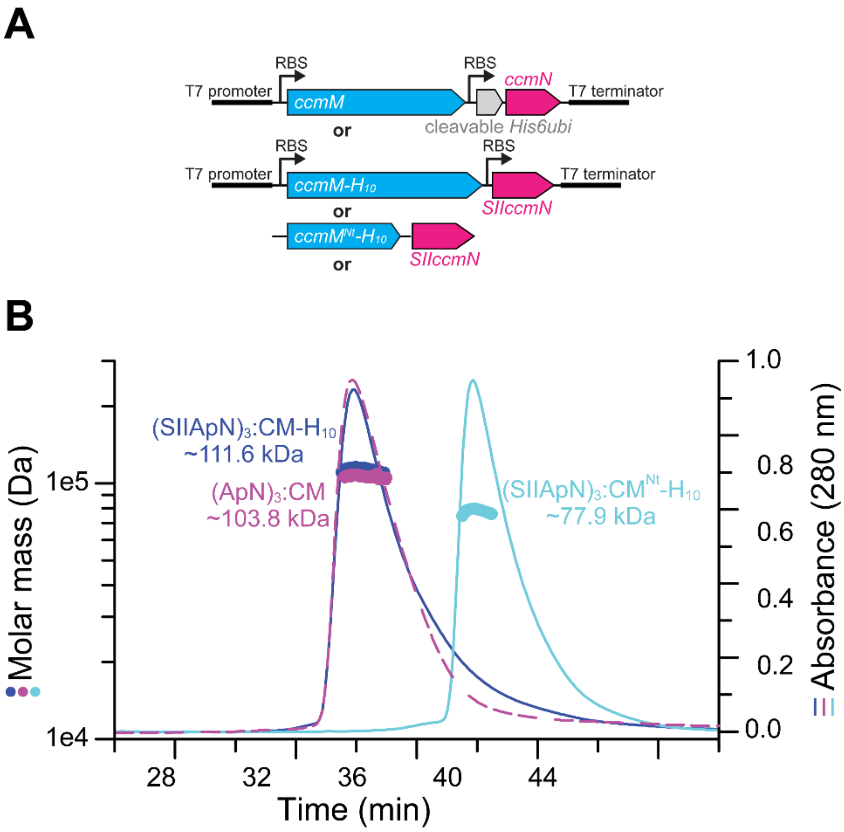


Figure 3.6 Formation of CM-ApN hetero-complex.
(A) Schematic of the bicistronic plasmids for recombinant expression of CM-ApN, (CM-H_{10})-SIIApN, and (CM^{Nt}-H_{10})-SIIApN hetero-complexes. For CM-ApN, a cleavable His6ubi tag was fused to the N-terminus of ApN. **(B)** SEC-MALS analysis of the hetero-complexes. The theoretical masses for complexes containing three ApN protomers and one CM protomer are 106.81 kDa for CM-ApN, 112.69 kDa for (CM-H_{10})-SIIApN, and 75.83 kDa for (CM^{Nt}-H_{10})-SIIApN. The measured molar masses are indicated. Representative data from two independent experiments are shown (n=2).

AlphaFold 3 was used to predict the structure of a hetero-tetramer consisting of three ApN protomers (ApN1 residues 1-114, ApN2 residues 11-114, and ApN3 residues 11-114) and one CM protomer (CM residues 16-182) (Figure **3.7A**), with pTM and ipTM values of 0.71 and 0.69, respectively. Notably, in the hetero-trimeric complex, wall C of ApN2 and wall D of CM are solvent-exposed (Sun et al., 2021). In the predicted hetero-tetramer model, wall A of the additional ApN (ApN3) interacts with the exposed wall C of ApN2. This interaction is similar to the wall A-wall C interaction observed in the $(ApN)_4$ homo-tetrameric complex (Figures **3.2D** and **3.3A**), whereas wall D of the CM protomer is covered by wall C of ApN3. However, there is a noticeable gap between wall C of ApN3 (which contains the conserved Cys49 and Cys88) and wall D of the CM protomer on its left, suggesting a weak interface (Figure **3.7A**) that possibly has a destabilizing effect on the interface between ApN3 wall A and ApN2 wall C, facilitating the dissociation of ApN3 from the $(ApN)_3$:CM hetero-tetramer and yielding the $(ApN)_2$:CM hetero-trimer observed in the crystal structure (Sun et al., 2021).

Another noteworthy interface observed in both the AlphaFold 3 model and the crystal structure of the hetero-trimer is the interaction of the loops between walls D and F of the CM protomer with wall A of ApN1 on its right (Figure **3.7A**). This interface is stabilized by an extensive "tongue and groove" interaction involving hydrophobic and polar interactions, including several hydrogen bonds (Figure **3.7B,C**). We suggest that this interface possibly has a stabilizing effect on the hydrophobic interface between ApN1 wall C and ApN2 wall A (Figure **3.7A**).

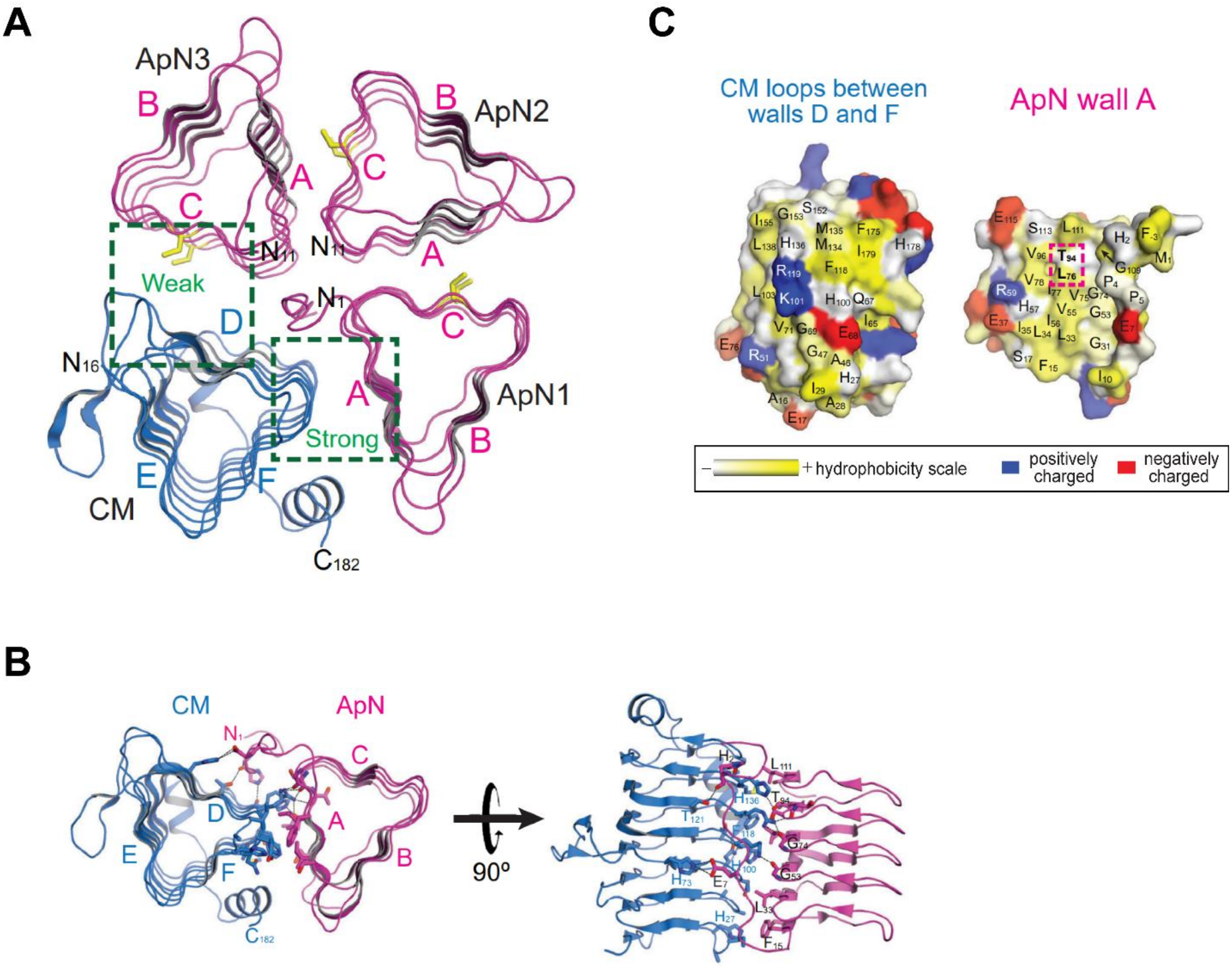


Figure 3.7 Structural analysis of the $(ApN)_3$:CM hetero-complex.
(A) Structural model of $(ApN)_3$:CM hetero-tetrameric complex (ApN1 residues 1-114, ApN2 and ApN3 residues 11-114, and CM residues 16-182) as predicted by AlphaFold 3 is shown in an end-on view, with the N-termini on the top. ApN protomers 1-3 are shown in magenta, and CM is shown in blue, both in ribbon representation. The N-terminal residues of the ApN protomers are indicated, while both N- and C-terminal residues of the CM protomer are

indicated. Walls of the ApN protomers are labeled A, B, and C, as shown in Figure **3.2D**, and the CM protomer walls labeled D, E, and F. The interacting surfaces between ApN protomers and CM are highlighted with dashed boxes and are labeled as either weak or strong. The highly conserved Cys49 and Cys88 on ApN wall C are shown as sticks in gold. **(B-C)** Structural analysis of the strong interface between ApN and CM. Side chains in stick representation showing the interlocking "tongue and groove" interaction of the loop residues between walls D and F of CM (in blue) with residues on wall A of ApN (in magenta) **(B)**. Surface representation showing the residues involved in the "tongue and groove" interaction exposed to the viewer **(C)**. The two conserved residues L76 and T94 on wall A of the ApN protomer are indicated by a dashed box. Residue coloring is based on their hydrophobicity moment and charge. Hydrophobic moment, shades of yellow; positively charged, blue; negatively charged, red.

Since the interface of CM loops between walls D and F and wall A of ApN1 likely plays a critical role in stabilizing the hetero-complex, we introduced point mutations targeting the two highly conserved residues Leu76 and Thr94 (Figures **3.1A** and **3.7C**). We mutated each residue individually to either the charged residues Asp or Arg (SIIApN-L76D, SIIApN-L76R, SIIApN-T94R) in the co-expression of (CM$^{\mathrm{Nt}}$-H$_{10}$)-SIIApN to impair hydrophobic interactions and introduce steric hindrance. Thr94 was found to be involved in a direct interaction between ApN1 and CM in the reported (SUMO-ApN)$_2$:CM hetero-trimer (Sun et al., 2021) (Figure **3.1A**). As expected, the ApN mutants failed to assemble into hetero-complexes when co-expressed with CM$^{\mathrm{Nt}}$-H$_{10}$, as analyzed by pull-down experiments (Figure **3.8**).

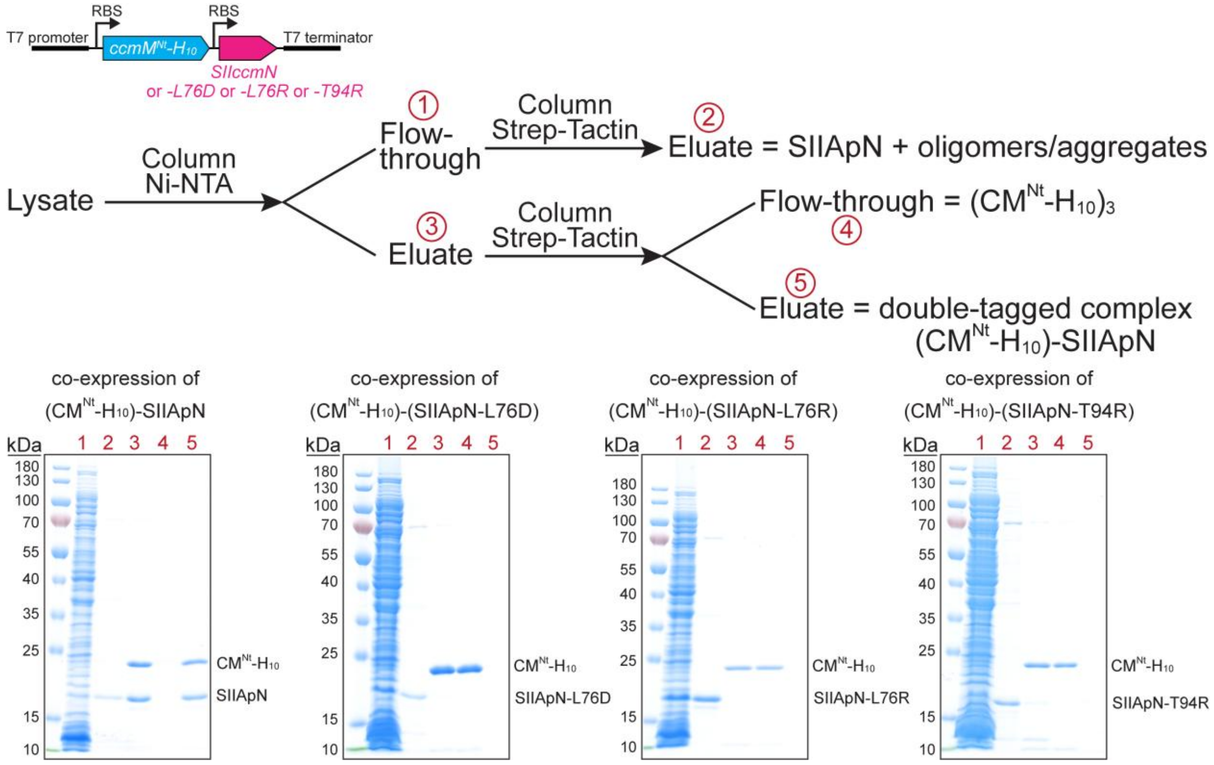


Figure 3.8 ApN point mutants at the strong interlocking interface.
The formation of hetero-complex between ApN and CM was assessed using ApN point mutants and analyzed by two-step purification with Ni-NTA column and Strep-Tactin columns, as indicated. Top: Schematic of the bicistronic plasmids used for recombinant co-expression of (CM^{Nt}-H_{10})-SIIApN (as control), (CM^{Nt}-H_{10})-(SIIApN-L76D), (CM^{Nt}-H_{10})-(SIIApN-L76R), and (CM^{Nt}-H_{10})-(SIIApN-T94R) hetero-complexes. Middle: Workflow of the pull-down experiments. Bottom: SDS-PAGE analysis of the pull-down results. The numbering of the lanes corresponds to the numbering in the workflow.

3.4. Cryo-EM single-particle analysis of the CM-ApN hetero-complex

To gain structural insights into the CM-ApN hetero-complexes and confirm their oligomeric state, we performed cryo-EM single-particle analysis on both the untagged $(ApN)_3$:CM and the double-tagged $(SIIApN)_3$:CM-H_{10}. Interestingly, the top twenty reference-free 2D class averages for both $(ApN)_3$:CM and $(SIIApN)_3$:CM-H_{10} hetero-complexes showed one major class of dimers of hetero-trimers in an S-shaped open conformation (Figure **3.9A,B**), which is similar to the S-shaped dimer-of-trimers $(SUMO\text{-}ApN)_2$:CM^{209} seen in the asymmetric unit of the crystal structure (Sun et al., 2021) (Figure **3.9C,D**). Representative 2D class averages of $(ApN)_3$:CM and $(SIIApN)_3$:CM-H_{10} are shown. All top twenty 2D class averages for both proteins are provided in Appendix **A.4.2**. Additional density for the N-terminal Strep-tag II at the N-terminus of ApN and the H_{10} at the N-terminus of CM, as well as the C-terminal domain (containing EP) of ApN and the C-terminal domain (containing the SSUL modules) of CM, was not observed (Figure **3.9A,B**). Some $(ApN)_4$ particles were also observed in the 2D class averages of $(ApN)_3$:CM (see Appendix **A.4.2**). However, no hetero-tetramers of either $(ApN)_3$:CM or $(SIIApN)_3$:CM-H_{10}, as identified in solution, were present in the 2D class averages.

In the S-shaped dimer-of-trimers, the CM protomers can easily be distinguished from the ApN protomers due to the presence of an additional density of its C-terminal α-helix that packs along wall F of the CM protomer (Figure **3.9A-C**). Note that the proteins used for crystallization are trimeric in solution (Sun et al., 2021). These hetero-trimers may arise from the dissociation of the labile ApN protomer from the hetero-tetramers, as we reasoned the weak interface between wall C of ApN3 and the wall D of CM could facilitate the dissociation (Figure **3.7A**). During crystallization (Vekilov & Vorontsova, 2014) or cryo-EM sample preparation (Schmidt-Krey & Rubinstein, 2011), the hetero-trimers may associate into the dimer of hetero-trimers due to high local concentrations of proteins. This association is likely driven by hydrophobic interactions between the juxtaposed 'C' walls of ApN protomers, which contain two conserved cysteines, Cys49 and Cys88 (Figures **3.1A**, **3.3B** and **3.9C,E**).

Two circular densities were observed in the center of each hetero-trimer in the 2D class averages (Figure **3.9A,B**). These densities may correspond to the N-terminal region of ApN1 (Figure **3.9A-D**), compared to the crystal structure (Sun et al., 2021), since the N-terminus of ApN2 is not resolved. Alternatively, the N-terminus of ApN2 and ApN3 appear in a disordered state in the predicted structural model by AlphaFold 3 (Figure **3.9F**). In order to test if these N-terminal residues contribute to the stability of the hetero-complex, we generated a truncated construct, (CM-H_{10})-ApNΔN10, in which the first ten amino acids following the initial methionine of ApN were deleted. SEC-MALS analysis revealed that the resulting complex exists as $(\text{ApN}\Delta\text{N10})_2$:CM-$H_{10}$ with a molar mass of ~87.4 kDa (theoretical mass: 90.04 kDa) in solution (see Appendix **A.3**). This suggests that the N-terminus of ApN1 is not necessary for stabilizing the hetero-trimers, but it may play a role in forming the hetero-tetramers.

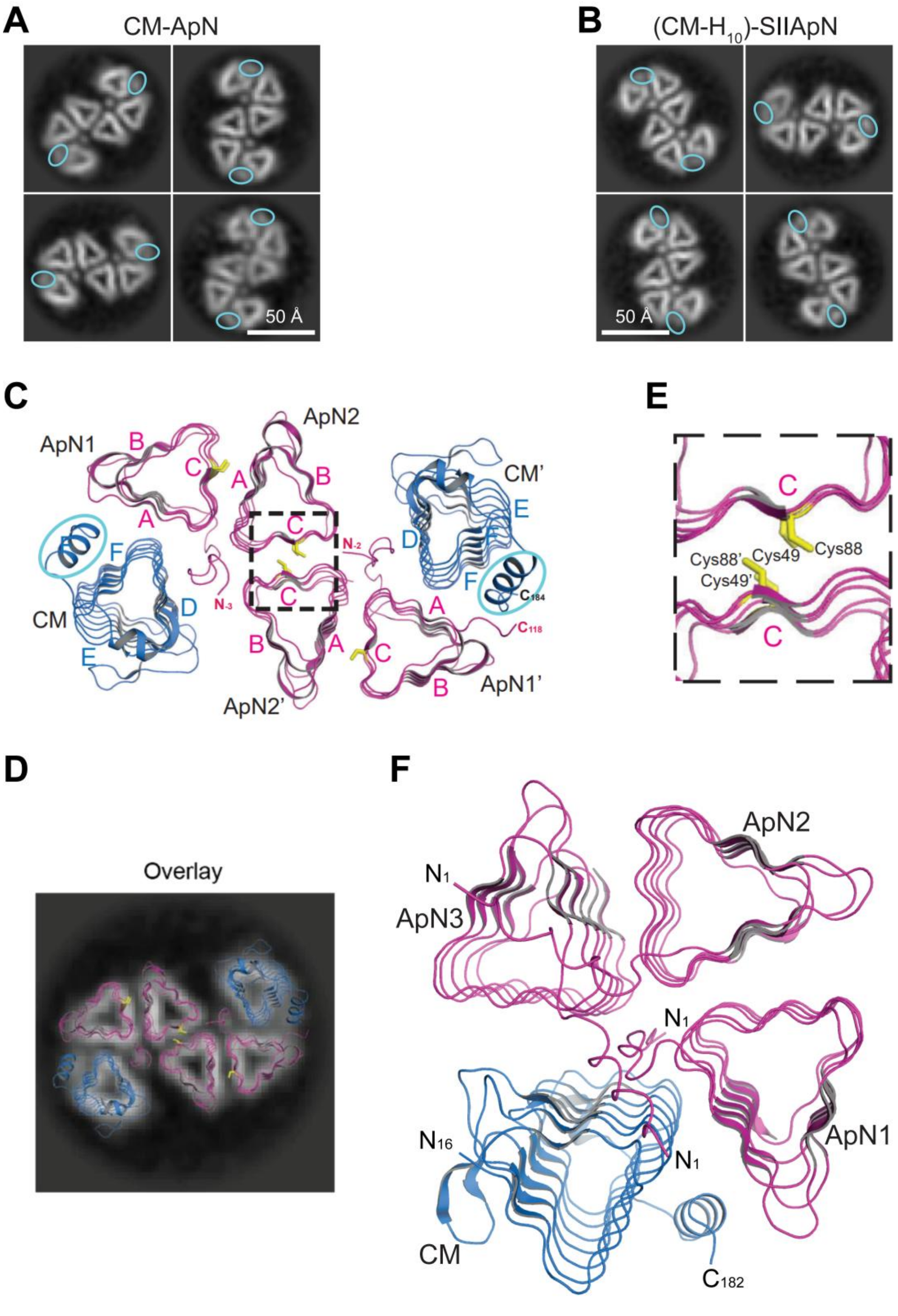


Figure 3.9 Lability of the hetero-tetrameric CM-ApN complex.
(A-B) Cryo-EM single-particle analysis of CM-ApN hetero-complexes **(A)** and (CM-H_{10})-SIIApN hetero-complexes **(B)**. Representative 2D class averages are shown. The C-terminal α-helix of the CM protomer, which distinguishes it from the ApN protomers, is highlighted by the blue oval. Scale bar, 50 Å. **(C)** End-on bottom view of the asymmetric unit of the (SUMO-ApN)$_2$:CM^{209} crystal structure (PDB: 7D6C) in ribbon representation. The CM protomers are shown in blue and the ApN protomers in magenta. The N- and C-termini are indicated when

possible. The C-terminal α-helices of the CM protomers are marked by blue ovals. The highly conserved Cys49 and Cys88 on ApN wall C are shown as sticks in gold. **(D)** Structural overlay of the crystal structure (PDB:7D6C) with a cryo-EM 2D class of CM-ApN dimer-of-trimers. **(E)** A close-up view within the black dashed box highlights the positions of the two conserved cysteines, Cys49 and Cys88, on the juxtaposed 'C' walls of ApN2 and ApN2'. **(F)** An end-on view of the hetero-complex of ApN (residues 1-114) and CM^{Nt} (residues 16-182) in ribbon representation highlights the position of the N-terminal tail of the ApN1 protomer in the center of the tetramer.

To stabilize the proteins and prevent the dissociation of one ApN protomer from the $(ApN)_3$:CM hetero-complex during cryo-EM sample preparation, we employed several strategies, including adding the non-ionic detergent Octyl-β-D-glucopyranoside (OG) (Benton et al., 2018) or the zwitterionic detergent CHAPSO (Chen et al., 2019) to proteins, and using graphene oxide-coated grids (Fan & Sun, 2022). These trials were conducted with $(SIIApN)_3$:CM-H_{10} due to its high purity. Results of these tests are provided in Appendix **A.4.2**. However, none of these methods successfully prevented the dissociation of the labile SIIApN protomer from the $(SIIApN)_3$:CM-H_{10} hetero-tetrameric complex and transformation into the dimer-of-trimers.

3.5. Cysteine oxidation induces the conversion of $(ApN)_3$:CM hetero-tetramer to $(ApN)_2$:CM hetero-trimer

We noticed the conversion of the $(ApN)_3$:CM hetero-tetramer to the $(ApN)_2$:CM trimer and sought to investigate whether there is a regulatory mechanism that drives this transition. Considering that the formation of CM-ApN hetero-complex occurs in the reducing environment of cytosol, whereas the carboxysome interior becomes oxidizing upon maturation (Chen et al., 2013; Pena et al., 2010; Price et al., 1992). This redox shift plays a key role in regulating the activities of the two essential enzymes, Rubisco and CA (Flecken et al., 2020; Pena et al., 2010). Additionally, the oxidizing condition promotes disulfide bond formation in the SSUL modules of CM and facilitates the homo-demixing of $(CM)_3$ (Zang et al., 2021). The thiol groups of the two conserved cysteines, Cys49 and Cys88, are surface-exposed on wall C of ApN3 in the hetero-tetramer. This raises the possibility that their oxidation could trigger the tetramer-to-trimer conversion. In order to test this hypothesis, we performed SEC-MALS analysis of $(SIIApN)_3$:CM^{Nt}-H_{10} following oxidation with 5 mM H_2O_2 on ice for 1 h, respectively. Surprisingly, the molar mass of the oxidized hetero-tetrameric complex decreased notably to ~60.1 kDa compared to the reduced complex, which has a molar mass of ~77.9 kDa (Figure **3.10A**). This corresponds to a hetero-trimer consisting of two SIIApN protomers and one CM^{Nt}-H_{10} protomer (theoretical mass: 58.12 kDa). This result was further supported by cryo-EM single-particle analysis of the $(SIIApN)_3$:CM-H_{10} hetero-tetramer upon oxidation with 2 mM H_2O_2 on ice for 30 min, which revealed 2D class averages corresponding to the hetero-trimer (Figure **3.10B**). Together, these findings suggest that cysteine oxidation interferes with the hydrophobic interactions between the juxtaposed 'C' walls of the ApN protomers that are necessary for the formation of the dimer-of trimers (Figures **3.3B** and **3.9A-C**).

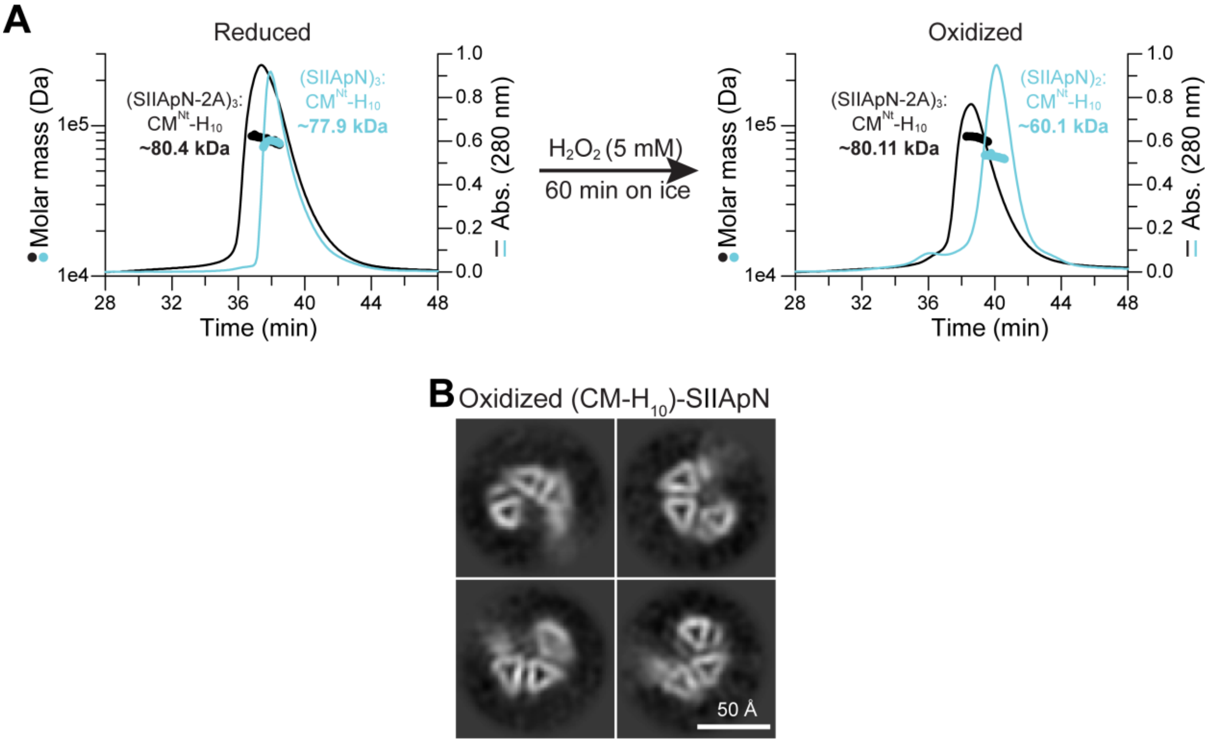


Figure 3.10 Oxidation of the (CM-H_{10})-SIIApN hetero-complexes.
(A) SEC-MALS analysis of (CM^{Nt}-H_{10})-SIIApN and (CM^{Nt}-H_{10})-(SIIApN-2A) with and without H_2O_2 oxidation treatment. Left: SEC-MALS analysis of proteins under reducing conditions in the presence of DTT. Right: SEC-MALS analysis of proteins under oxidizing conditions upon the treatment with H_2O_2. The theoretical masses for complexes containing three ApN protomers and one CM protomer are 75.83 kDa for (CM^{Nt}-H_{10})-SIIApN, 75.64 kDa for (CM^{Nt}-H_{10})-(SIIApN-2A), and for complexes containing two ApN protomers and one CM protomer are 58.12 kDa for (CM^{Nt}-H_{10})-SIIApN, 57.99 kDa for (CM^{Nt}-H_{10})-(SIIApN-2A). The measured molar masses are indicated. Representative data from two independent experiments are shown (n=2). **(B)** Cryo-EM single-particle analysis of (CM-H_{10})-SIIApN hetero-tetramers upon oxidation treatment with H_2O_2. Representative 2D class averages are shown. Scale bar, 50 Å.

To further confirm that the conversion is due to the oxidation of cysteines, we mutated the two cysteines Cys49 and Cys88 to alanine (ApN-2A) and serine (ApN-2S), respectively, and generated these proteins in the presence of DTT. We analyzed all CM-(ApN-2A) complexes by SEC-MALS and found that the molar masses were ~111.8 kDa for CM-(ApN-2A) (Figure **3.11A**), ~114.6 kDa for (CM-H_{10})-(SIIApN-2A), and ~80.4 kDa for (CM^{Nt}-H_{10})-(SIIApN-2A), respectively (theoretical masses: 106.62 kDa, 112.49 kDa, 75.64 kDa, respectively). These complexes were hetero-tetramers in solution. SEC-MALS analysis of (CM-H_{10})-(SIIApN-2A) and (CM^{Nt}-H_{10})-(SIIApN-2A) can be found in Appendix **A.3**. After undergoing the same treatment with H_2O_2 as $(SIIApN)_3$:CM^{Nt}-H_{10}, the molar mass of (SIIApN-2A)$_3$:CM^{Nt}-H_{10} remained almost unchanged at ~80.11 kDa (Figure **3.10A**), suggesting that the cysteine-to-alanine mutation does not affect the surface in the hetero-tetramers. SEC-MALS analysis shows that the CM-(ApN-2S) complexes are all hetero-trimers consisting of two ApN-2S protomers and one CM protomer, with molar masses of ~89.3 kDa for CM-(ApN-2S) (Figure **3.11B**), ~93.4 kDa for (CM-H_{10})-(SIIApN-2S), and ~57.2 kDa for (CM^{Nt}-H_{10})-(SIIApN-2S), respectively (theoretical masses: 90.42 kDa, 94.91 kDa, 58.05 kDa, respectively). SEC-MALS analysis of (CM-H_{10})-(SIIApN-2S) and (CM^{Nt}-H_{10})-(SIIApN-2S) can be found in Appendix **A.3**. These results indicate that changing cysteine to serine is sufficient to destabilize the weak ApN3-CM interface. Analysis by native-PAGE showed that the (SIIApN-2S)$_2$:CM^{Nt}-H_{10} complex migrated faster than $(SIIApN)_3$:CM^{Nt}-H_{10} and (SIIApN-2A)$_3$:CM^{Nt}-H_{10} (Figure **3.11C**), thus indicating that the Cys to Ser mutation induced a transition from hetero-tetramer to hetero-trimer. Interestingly, although both Ala and Ser are generally used to substitute for a redox-regulated Cys residue, they differ in their physicochemical properties. Ala is smaller than Cys and has a non-polar side chain (methyl group) that maintains the hydrophobicity as the reduced Cys. In contrast, Ser is similar in size to Cys but contains a polar hydroxyl (OH) group, making it more hydrophilic. Likewise, Cys oxidation would also increase hydrophilicity (Garrido Ruiz et al., 2022). This is consistent with the observation that Cys oxidation triggers a conformational switch from a hetero-tetramer to a hetero-trimer, similar to the Cys to Ser mutation.

Additionally, we performed cryo-EM single-particle analysis on untagged (ApN-2A)$_3$:CM and (ApN-2S)$_2$:CM hetero-complexes, and double-tagged (SIIApN-2A)$_3$:CM-H_{10} and (SIIApN-2S)$_2$:CM-H_{10} hetero-complexes. Reference-free 2D class averages of (ApN-2A)$_3$:CM and (SIIApN-2A)$_3$:CM-H_{10} show the presence of S-shaped dimer-of-trimers (Figure **3.11D,E**), which are similar to the dimer-of-trimers in the 2D class averages of $(ApN)_3$:CM and $(SIIApN)_3$:CM-H_{10} (Figure **3.9A,B**). This suggests that the conversion from hetero-tetramers to hetero-trimers can occur independently of cysteine oxidation due to weakening of the ApN3-CM interface caused by the air-water interface during cryo-EM sample preparation. Some $(SIIApN\text{-}2A)_2$:CM-H_{10} hetero-trimers are also observed (Figure **3.11E**). Additionally, some $(ApN\text{-}2A)_4$ also appear in the 2D classes of $(ApN\text{-}2A)_3$:CM (see Appendix **A.4.3**). Intriguingly, cryo-EM single-particle analysis of $(ApN\text{-}2S)_2$:CM hetero-complexes is complicated by the presence of 2D classes of hetero-trimers and some tetramer-like particles that cannot be identified unambiguously (Figure **3.11F**). In contrast, cryo-EM single-particle analysis of $(SIIApN\text{-}2S)_2$:CM-H_{10} shows dominant classes of hetero-trimers (Figure **3.11G**). The absence of S-shaped complexes in the 2D class averages suggests that the cysteine-to-serine mutation also altered the hydrophobic interactions between the juxtaposed 'C' walls required for forming the dimer-of-trimers (Figures **3.3B** and **3.9C**).

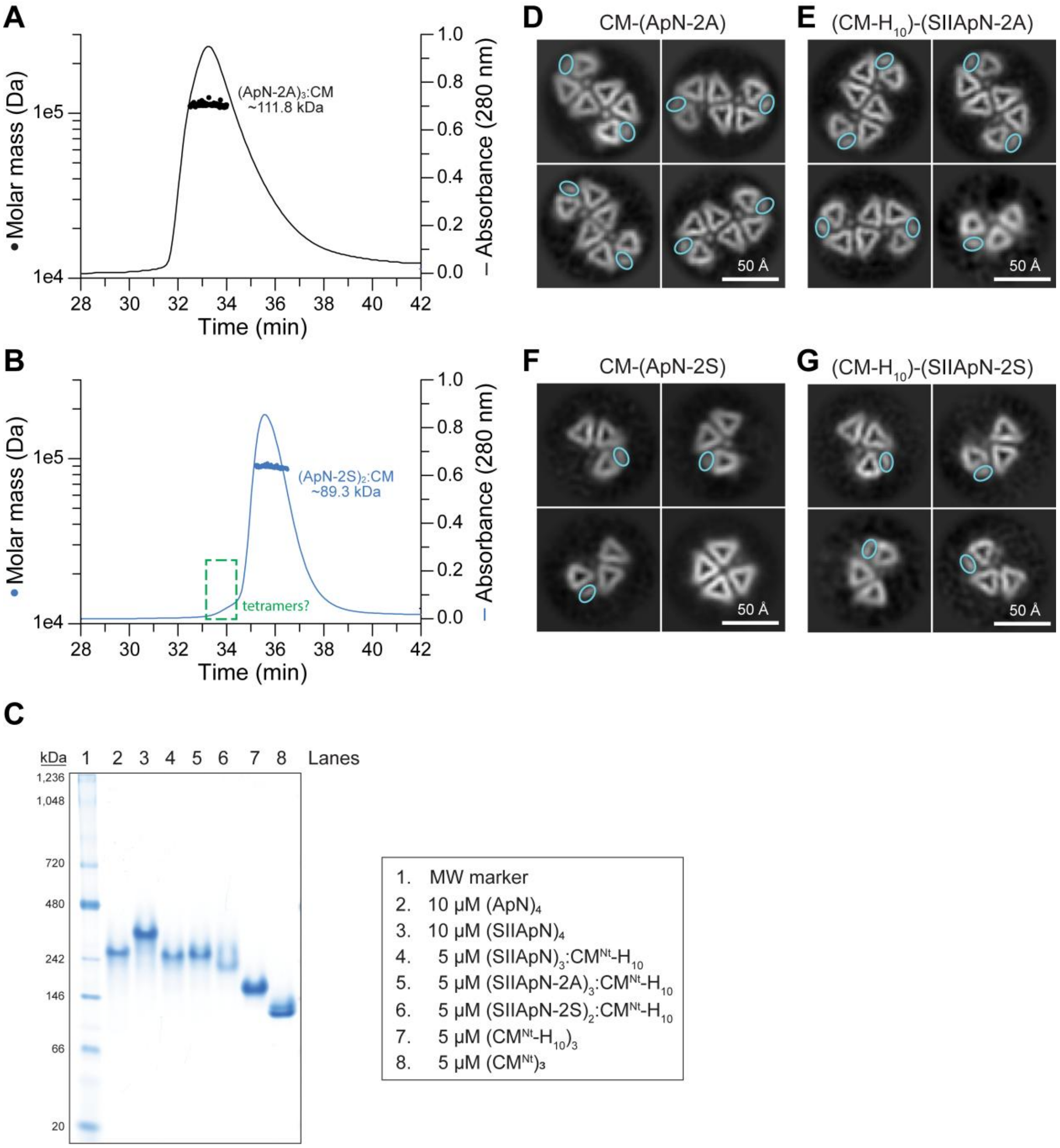
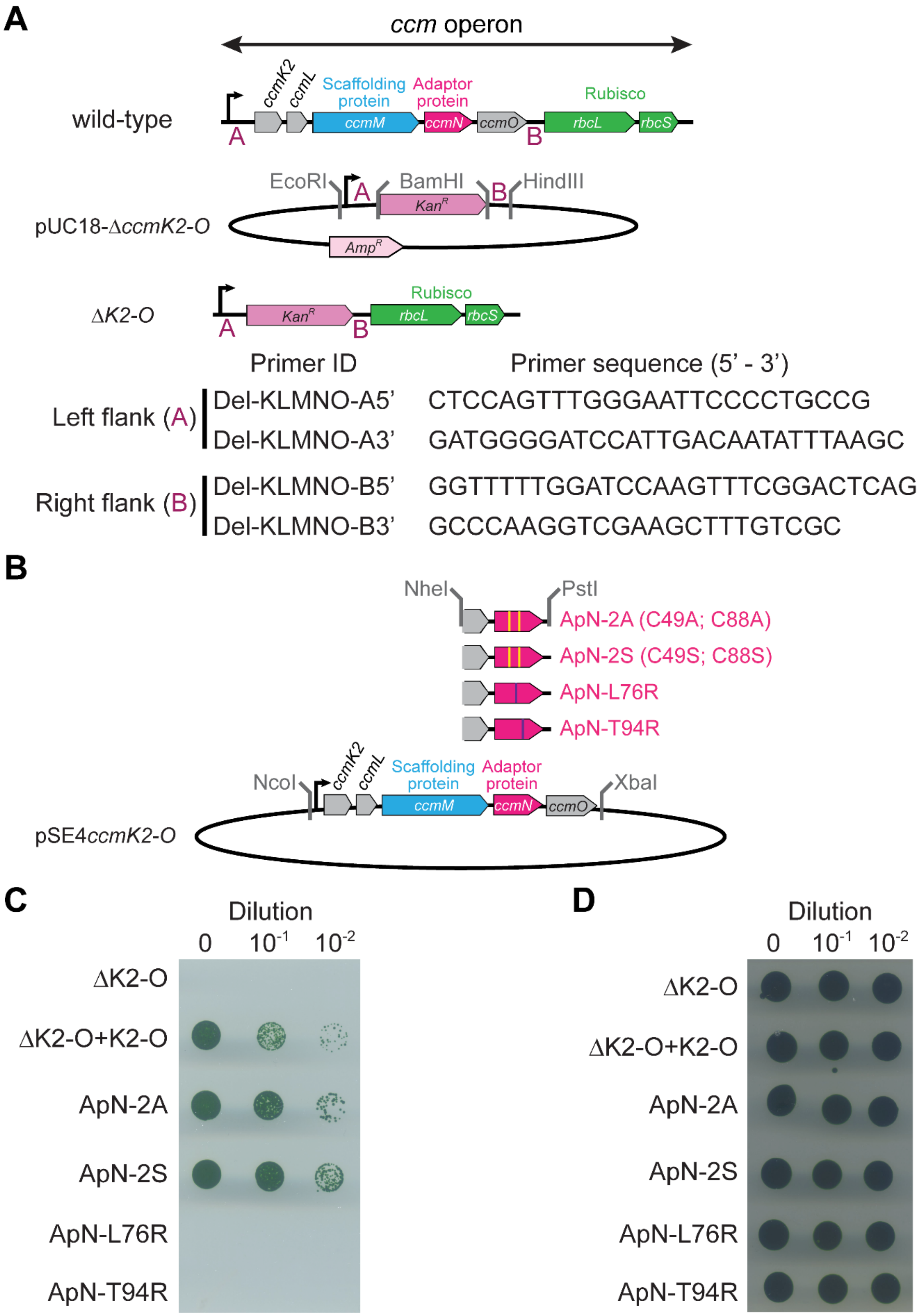


Figure 3.11 Biochemical and structural analysis of ApN-alanine- and ApN-cysteine-mutant CM-ApN hetero-complexes.

(A-B) SEC-MALS analysis of non-tagged ApN-cysteine-mutant CM-ApN hetero-complexes. The theoretical masses of $(ApN\text{-}2A)_3$:CM, and $(ApN\text{-}2S)_2$:CM are 106.62, and 90.42 kDa, respectively. The measured molar masses are indicated. The boxed area in **(B)** was determined as tetramer with low confidence. Representative data from two independent experiments are shown (n=2). **(C)** Native-PAGE analysis of CM-ApN hetero-complexes. The hetero-tetrameric $(SIIApN)_3$:CM^{Nt}-H_{10} and $(SIIApN\text{-}2A)_3$:CM^{Nt}-H_{10} complexes migrate at the same height (lanes 4 and 5, respectively), while the hetero-trimeric $(SIIApN\text{-}2S)_2$:CM^{Nt}-H_{10} complex migrates faster (lane 6). In comparison, the control proteins tetrameric ApN migrates at a similar height to the tetrameric hetero-complexes (lane 2), while tetrameric SIIApN migrates more slowly (lane 3). The trimeric CM^{Nt}-H_{10} and CM^{Nt} complexes migrate fast (lanes 7 and 8, respectively). **(D-G)** Representative 2D class averages from cryo-EM single-particle analysis of CM-(ApN-2A) **(D)**, (CM-H_{10})-(SIIApN-2A) **(E)**, CM-(ApN-2S) **(F)**, and (CM-H_{10})-(SIIApN-2S) **(G)**. The presumed C-terminal α-helices of the CM protomers are marked by blue ovals. Scale bar, 50 Å.

3.6. *In vivo* complementation in *Se*7942

In this study, we demonstrated that ApN alone does not interact with the pro-carboxysome core proteins or undergo phase separation to be recruited into pro-carboxysome condensates. Through the recombinant co-expression of ApN and CM, however, we obtained the CM-ApN hetero-complex. To investigate whether ApN requires hetero-complex formation with CM for functional carboxysome maturation and to evaluate the effect of ApN cysteine mutants on this process, we generated a series of ApN mutants for expression in a carboxysomeless mutant of *Se*7942 and performed *in vivo* complementation assays in *Se*7942. These mutants include the ApN-L76R and ApN-T94R, which disrupt the formation of CM-ApN hetero-complexes; and the ApN-2A and ApN-2S, which may affect redox regulation. The ApN-2S mutant, in particular, changes the composition of the CM-ApN hetero-complexes by lacking one ApN protomer, which may affect carboxysome maturation. We used the Δ*ccmKLMNO* mutant (ΔK2-O), in which these gene components of the *ccm* operon have been homologously replaced with the kanamycin-resistance marker, *Kan*^{R} (Figure **3.12A,B**). The ΔK2-O strain exhibits a high CO_2-requiring (HCR) phenotype, and its growth in air can be restored by reintroducing *ccmK2-O* with the native promoter on the pSE4 plasmid (pSE4*ccmK2-O*) (Maeda et al., 1998; Rae et al., 2012) (Figure **3.12B-D**). This design was created to verify the function and importance of each of the CcmK2, CcmL, CcmM, CcmN and CcmO proteins in carboxysome biogenesis by reintroducing their respective genes into the ΔK2-O strain.



	Primer ID	Primer sequence (5' - 3')
Left flank (A)	Del-KLMNO-A5'	CTCCAGTTTGGGAATTCCCCTGCCG
	Del-KLMNO-A3'	GATGGGGATCCATTGACAATATTTAAGC
Right flank (B)	Del-KLMNO-B5'	GGTTTTTGGATCCAAGTTTCGGACTCAG
	Del-KLMNO-B3'	GCCCAAGGTCGAAGCTTTGTCGC

Figure 3.12 Generation of the *Se*7942 ΔK2-O cell line and effect of ApN mutants on cell growth.

(A) Generation of the *Se*7942 ΔK2-O cell line. The ΔK2-O mutant cell line was obtained via homologous recombination of the *ccm* operon with a kanamycin resistance (*Kan*^{R}) cassette supplied in pUC18-*ΔKO*. Details of the generation of pUC18-*ΔKO* can be found in Chapter **2.3.1**. The primer sequences used to generate flanks A and B are listed. **(B)** Design of the plasmid for complementation. The *ccmK2-O* operon was reintroduced into competent *Se*7942 ΔK2-O cells on plasmid pSE4*ccmK2-O*. Variant ApN forms were cloned into the NheI and PstI sites of this construct to enable the expression of ApN-2A, ApN-2S, ApN-L76R, and ApN-T94R. **(C-D)** Growth tests of mutant strains. Cyanobacterial spot plate assays were undertaken to evaluate the ability of ΔK2-O cyanobacterial cells, as well as ΔK2-O cells complemented with wild-type K2-O or the ApN-2A, ApN-2S, ApN-L76R, or ApN-T94R mutants to grow in air **(C)** and at high 3% CO_2 (v/v in air) **(D)**, at 30 °C for 21 days. Growth was observed across all strains. A representative plate from three replicates is shown (n=3) **(C)**.

Complementation of the ΔK2-O strain with the plasmid pSE4*ccmK2-O* resulted in a moderately slower growth rate in air, with a doubling time of ~15 h, compared to ~7.5 h for the wild-type strain (Figure **3.13A**). Ultrastructure analysis of cyanobacterial cells revealed a smaller number of and slightly larger carboxysomes in ΔK2-O+pSE4*ccmK2-O* than in the wild-type (Figure **3.13B-D**). Cell expressing ApN-76R and ApN-94R were unable to grow at atmospheric CO_2 levels (Figure **3.12C**), indicating that these mutants cannot complement the HCR phenotype of ΔK2-O cells. Ultrastructure analysis showed that no carboxysomes formed in these cells; instead, polar aggregates significantly larger than the carboxysomes in the wild-type or ΔK2-O+pSE4*ccmK2-O* cells formed (Figure **3.13B-D**). Interestingly, expression of the ApN-2A and ApN-2S mutants rescued the HCR phenotype of ΔK2-O at atmospheric CO_2 levels (Figure **3.12C**), but they require a longer growth time of ~27 h, compared to ΔK2-O+pSE4*ccmK2-O* (~15 h) (Figure **3.13A**). Regularly shaped carboxysomes were observed in the ApN-2A and ApN-2S mutants (Figure **3.13B**), though they were slightly larger than those of the ΔK2-O+pSE4*ccmK2-O* cells (Figure **3.13C,D**). Together, these results suggest that the CM-ApN hetero-complexes are important for carboxysome maturation and that the conserved cysteines may play a role in fine-tuning the efficiency of carboxysome maturation.

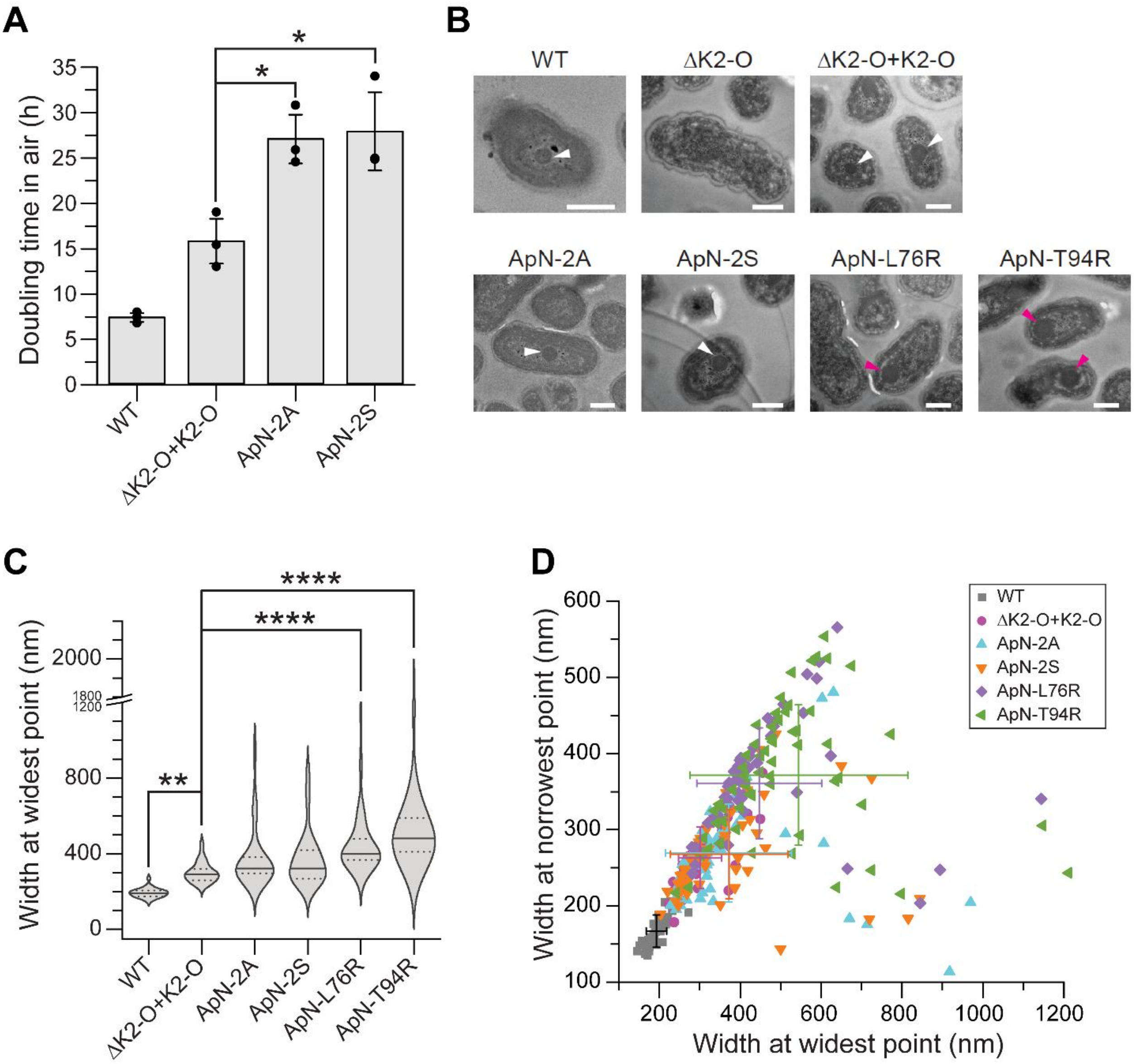


Figure 3.13 *In vivo* assessment of the role of ApN in carboxysome assembly and cyanobacterial cell growth in air.

(A) Cell growth analysis of wild-type (WT) *Se*7942 cells, ΔK2-O cells complemented with wild-type K2-O or with the ApN-2A or ApN-2S was performed in air at 30 °C for 8 h. The maximum growth rate was determined from the slope and converted to a doubling time. *, $p <$ 0.05. Data are mean ± s.d. of triplicate measurements (n=3). **(B)** Representative transmission electron micrographs of the following cells grown in modified BG-11 medium at 2% CO_2 and 30 °C for 7 days: wild-type (WT) *Se*7942 cells, ΔK2-O cells, and ΔK2-O complemented with wild-type K2-O or with the ApN-2A, ApN-2S, ApN-L76R or ApN-T94R. The images are representative of carboxysomes (indicated by white arrowheads) in WT, ΔK2-O+K2-O, ApN-2A, and ApN-2S cell lines. In the ApN-L76R and ApN-T94R mutant lines, polar bodies (indicated by magenta arrowheads) are observed; however, no internal ultrastructure is observed in the *ΔK2-O* knockout line. Scale bar, 500 nm. **(C-D)** Ultrastructure analysis of the dimensions of carboxysomes or polar bodies from images in **(B)**. Ultrastructure analysis shows the width at the widest point **(C)**, and the width at the widest point relative to the width at the narrowest point **(D)** of carboxysomes or polar bodies in the wild-type (WT) *Se*7942 cells (n=57), and ΔK2-O complemented with wild-type K2-O (n=53) or with the ApN-2A (n=58), ApN-2S (n=56), ApN-L76R (n=60) or ApN-T94R (n=65). In **(C)**, the median (solid line), interquartile range (dotted lines, 25%-75%), and the distribution of the data are shown. **, $p < 0.01$; ****, $p < 0.0001$. In **(D)**, each point represents the dimensions of a single carboxysome. The lines are standard deviation for each type of cell line. Color of lines corresponds to the symbol color of each cell type, with the exception for WT, where the line is black and the symbol is a grey square.

3.7. $(ApN)_3$:CM co-assembles with pro-carboxysome core proteins via multiple protein-protein interactions

The present study showed that ApN must form hetero-complexes with CM to participate in carboxysome maturation. However, the mechanism by which CM-ApN complexes are recruited into the pro-carboxysome biomolecular condensate remained unclear. Previous studies have shown that CM protomers mediate multiple dynamic protein-protein interactions in pro-carboxysome condensate formation (Wang et al., 2019; Zang et al., 2021) (Figure **3.14**). These interactions include (i) homo-demixing, mediated by interactions between the C-terminal SSUL modules and the N-terminal γCAL domain of adjacent CM protomers; (ii) interactions of CM with Rubisco via the SSUL modules; and (iii) interactions of CM with the C-termini of $(CA)_4$ via the γCAL domain, resulting in LLPS of each protein pair and of all four proteins (including Rubisco, $(CA)_4$, $(CM)_3$, and CM^{Nt}) in combination (Wang et al., 2019; Zang et al., 2021). We then analyzed the interaction between the $(ApN)_3$:CM hetero-complex and the other known components of the pro-carboxysome core under reducing conditions, using either DTT or TCEP as reducing agents, to mimic the cytosolic environment prior to carboxysome maturation (Chen et al., 2013; Pena et al., 2010; Price et al., 1992).

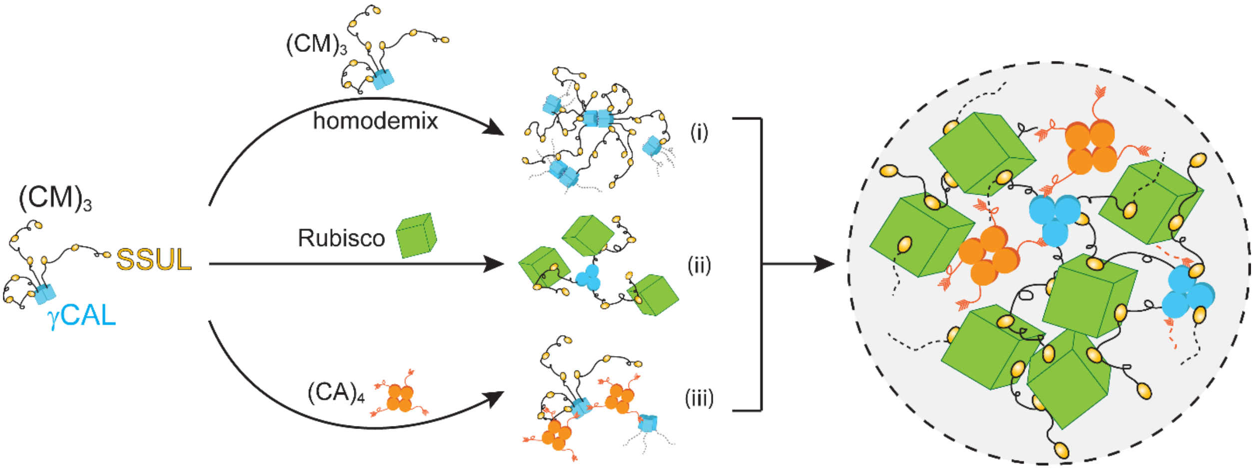


Figure 3.14 Participation of CM protomers in multiple protein-protein interactions (Zang et al., 2021).
Multiple protein-protein interactions involved in pro-β-carboxysome formation include: (i): Formation of a $(CM)_3$ homo-condensate via interactions between SSUL modules and γCAL domains of adjacent $(CM)_3$ complexes, as well as head-to-head association of γCAL domains between adjacent $(CM)_3$ trimers. Dotted lines indicate linker sequences between protein domains interacting with nearby molecules. γCAL domains are in blue and SSUL modules are in gold. (ii): Interactions of $(CM)_3$ with Rubisco (green) via SSUL modules binding in clefts between antiparallel RbcL dimers. (iii): Interactions of the C-termini of $(CA)_4$ (orange) with the γCAL domains of $(CM)_3$ together with the binding of SSUL modules of adjacent $(CM)_3$ complexes. Right: Condensate formation of $(CM)_3$ and CM^{Ct} with Rubisco and $(CA)_4$ via multiple protein-protein interactions.

The turbidity assay and fluorescence microscopy analysis were used to investigate the interaction of CM protomers in $(ApN)_3$:CM hetero-complexes with the other proteins. A turbidity assay of $(ApN)_3$:CM alone showed that $(ApN)_3$:CM underwent homo-demixing as well, though higher concentrations of proteins were required for aggregate formation than for $(CM)_3$ (Figure **3.15A**), which is due to the lower avidity of only three SSUL modules in $(ApN)_3$:CM compared to nine in $(CM)_3$. Florescence microscopy analysis confirmed the formation of phase-separated droplets of $(ApN)_3$:CM or $(CM)_3$ alone (Figure **3.15B,C**). Mixing $(ApN)_3$:CM with $(CM)_3$ enhanced phase separation, which occurred at lower concentrations of both proteins (Figure **3.15D**), where no homo-demixing was observed for the individual proteins (Figure **3.15B,C**). Turbidity assay also confirmed aggregate formation of the two proteins, while no absorbance signal was detected between $(CM)_3$ and $(ApN)_4$ (Figure **3.15E**). No homo-demixing or droplet formation was observed for $(ApN)_3$:CM^{Nt}, which lacks the SSUL modules (Figure **3.15A,F**), at a concentration that caused homo-demixing of $(ApN)_3$:CM. These results are consistent with previous findings that homo-demixing of CM protomers requires the presence of SSUL modules, and the interaction between SSUL modules and γCAL domains of adjacent CM protomers (Zang et al., 2021).

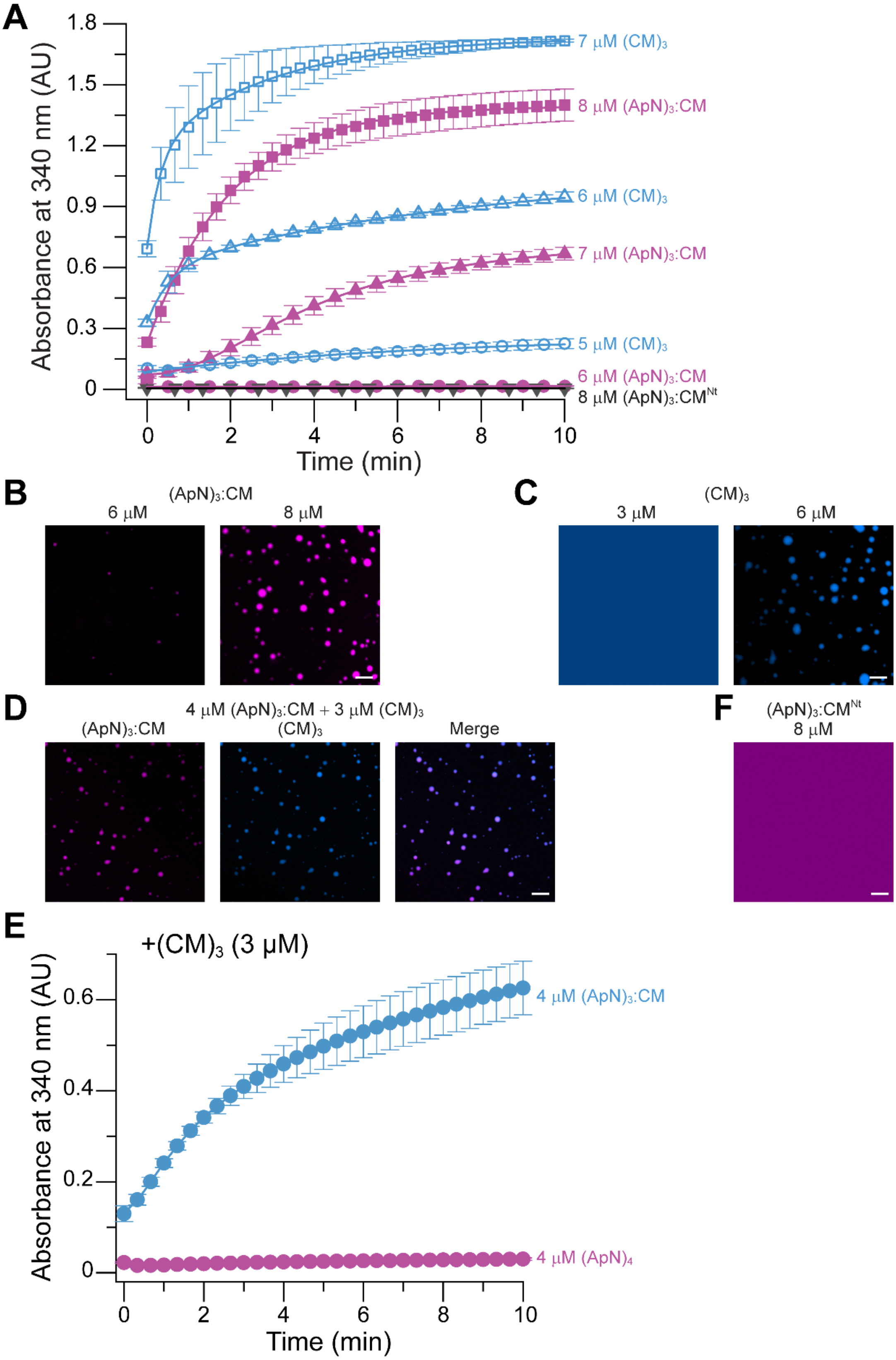


Figure 3.15 Concentration dependence of $(ApN)_3$:CM homo-condensate formation.
(A) Turbidity assays. Turbidity assays were performed in the presence of 5 mM DTT following the dilution of $(ApN)_3$:CM from a high concentration and salt level to reach final protein concentrations of 6-8 µM and a salt concentration of 50 mM KCl. $(CM)_3$ of 5-7 µM was also analyzed. Data are mean ± s.d. of triplicate measurements (n=3). **(B-D)** Analysis of LLPS by fluorescence microscopy. $(ApN)_3$:CM and $(CM)_3$ colocalize into droplet-shaped condensates. Efficient homo-condensate formation occurs for $(ApN)_3$:CM at 8 µM **(B)**, and for $(CM)_3$ at 6 µM **(C)**. When mixed at 4 µM and 3 µM, respectively, $(ApN)_3$:CM and $(CM)_3$ form a condensate **(D)** under concentrations where no homo-demixing occurs. $(ApN)_3$:CM was labeled with AF405, and $(CM)_3$ was labeled with AF488. Labeled proteins were used at a 1:10 ratio, mixed with respective unlabeled proteins and analyzed at the concentrations indicated. Representative data from two independent experiments are shown (n=2). Scale bar, 10 µm. **(E)** Turbidity assay of $(CM)_3$ (3 µM) with 4 µM of either $(ApN)_3$:CM or $(ApN)_4$ monitored at 340 nm under reducing conditions and 50 mM KCl. Data are mean ± s.d. of triplicate measurements (n=3). **(F)** Analysis of LLPS by fluorescence microscopy. No homo-condensate formation was detected with $(ApN)_3$:CM^{Nt} complex (8 µM), which lacks all SSUL modules. $(ApN)_3$:CM^{Nt} was labeled with AF405. Labeled proteins were used at a 1:10 ratio, mixed with respective unlabeled proteins and analyzed at the concentrations indicated. Representative data from two independent experiments are shown (n=2). Scale bar, 10 µm.

The SSUL modules of CM protomers interact with Rubisco by binding in a groove between antiparallel RbcL dimers (Wang et al., 2019; Zang et al., 2021). We performed a turbidity assay to confirm the interaction between $(ApN)_3$:CM and Rubisco, as well as to compare the binding affinities of $(ApN)_3$:CM, $(CM)_3$, and CM^{Ct} with Rubisco. $(ApN)_3$:CM has a ~6-fold lower apparent binding affinity for Rubisco (K_D^{app} ~0.4 µM) than $(CM)_3$ (K_D^{app} ~0.07 µM) (Figure **3.16A**). Fluorescence microscopy analysis confirmed that both proteins colocalized into phase-separated droplets (Figure **3.16B**). A similar K_D^{app} to Rubisco of ~0.4 µM was observed for CM^{Ct}, which also consists of only three SSUL modules. No interaction with Rubisco was observed in $(ApN)_3$:CM^{Nt} or in $(ApN)_4$ which lacks the SSUL modules (Figure **3.16C**) , consistent with the findings that the interaction between CM protomers and Rubisco requires the SSUL modules (Wang et al., 2019; Zang et al., 2021).

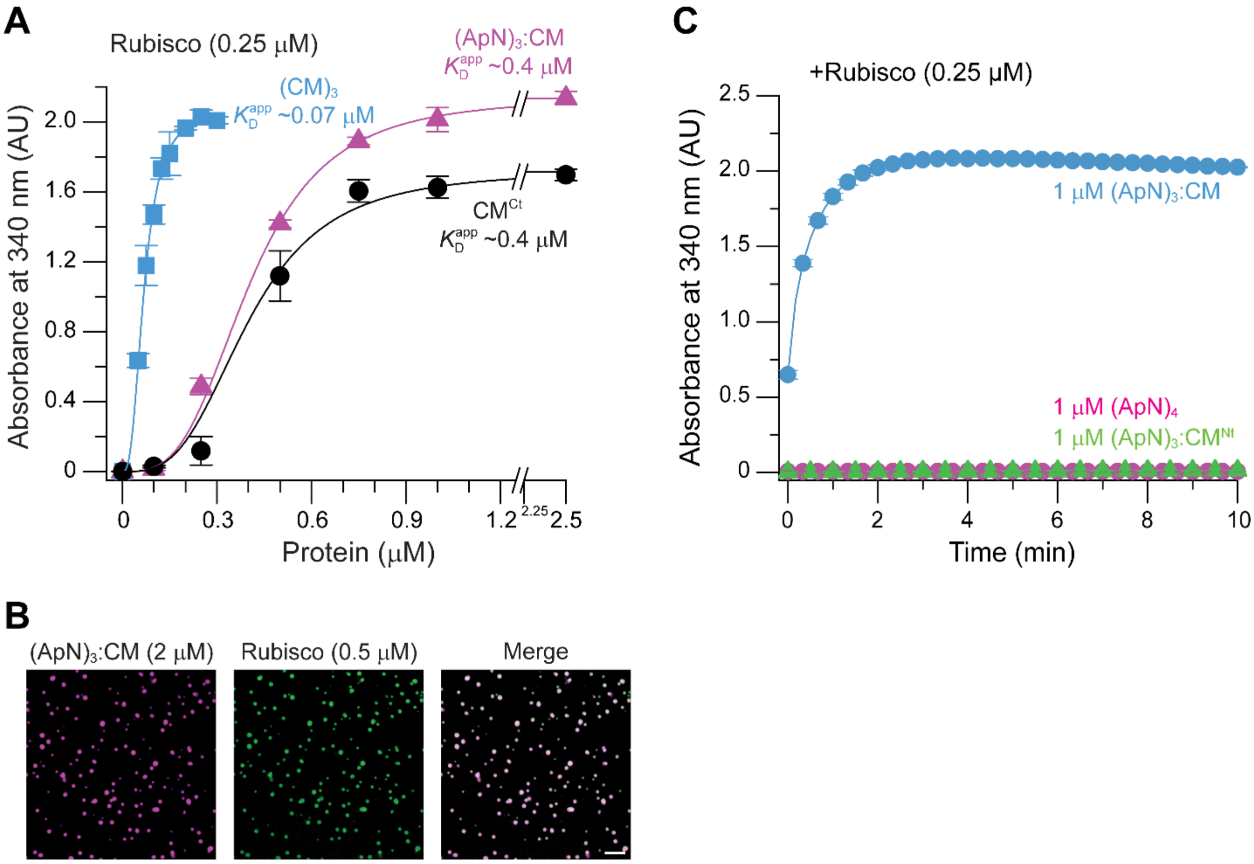


Figure 3.16 Interaction between $(ApN)_3$:CM and Rubisco.
(A) A comparison of binding affinities between $(CM)_3$/$(ApN)_3$:CM/CM^{Ct} with Rubisco. Turbidity was measured under reducing conditions with 100 mM KCl, using 0.25 µM Rubisco and 0 to 2.5 µM $(ApN)_3$:CM, 0 to 0.3 µM $(CM)_3$, or 0 to 2.5 µM CM^{Ct}. Absorbance values reached after 10 min are plotted. Data are mean ± s.d. of triplicate measurements (n=3). **(B)** Fluorescence microscopy analysis of $(ApN)_3$:CM-Rubisco condensate formation. Condensate formation was analyzed by fluorescence microscopy of $(ApN)_3$:CM and Rubisco under reducing conditions with 100 mM KCl. $(ApN)_3$:CM and Rubisco were labeled with the fluorophores AF405 and AF568, respectively. Labeled proteins were used at a 1:10 ratio, mixed with respective unlabeled proteins and analyzed at the concentrations indicated. Representative data from two independent experiments are shown (n=2). Scale bar, 10 µm. **(C)** Dependence of condensate formation on the presence of SSUL modules. Turbidity assays of Rubisco (0.25 µM) with 1 µM of either $(ApN)_3$:CM, $(ApN)_3$:CM^{Nt}, or $(ApN)_4$ was monitored at 340 nm in a buffer containing 100 mM KCl/5 mM DTT. Data are mean ± s.d. of triplicate measurements (n=3).

$(CM)_3$ interacts with $(CA)_4$ via the conserved C-terminal peptide of $(CA)_4$ which binds to a pocket in the γCAL domain of CM (Zang et al., 2021). This interaction results in a K_D^{app} of ~0.2 µM (Figure **3.17A**). For the interaction between $(ApN)_3$:CM and $(CA)_4$, we observed a K_D^{app} of ~1.4 µM, thus an affinity ~7-fold lower than that between $(CM)_3$ and $(CA)_4$ (Figure **3.17A**), consistent with the presence of only one CM protomer in the $(ApN)_3$:CM hetero-complex. Both proteins co-assembled into phase-separated droplets (Figure **3.17B**). Lacking of either the C-termini in (CAΔC)$_4$, the SSUL modules in (ApN)$_3$:CM^{Nt}, or the CM protomers in (ApN)$_4$, no absorbance signal was recorded (Figure **3.17C**), consistent with previous findings that phase separation involves the interaction of not only the C-terminal peptide of (CA)$_4$ and the γCAL domain of CM, but also the interaction of SSUL modules with γCAL domain of adjacent CM protomers (Zang et al., 2021).

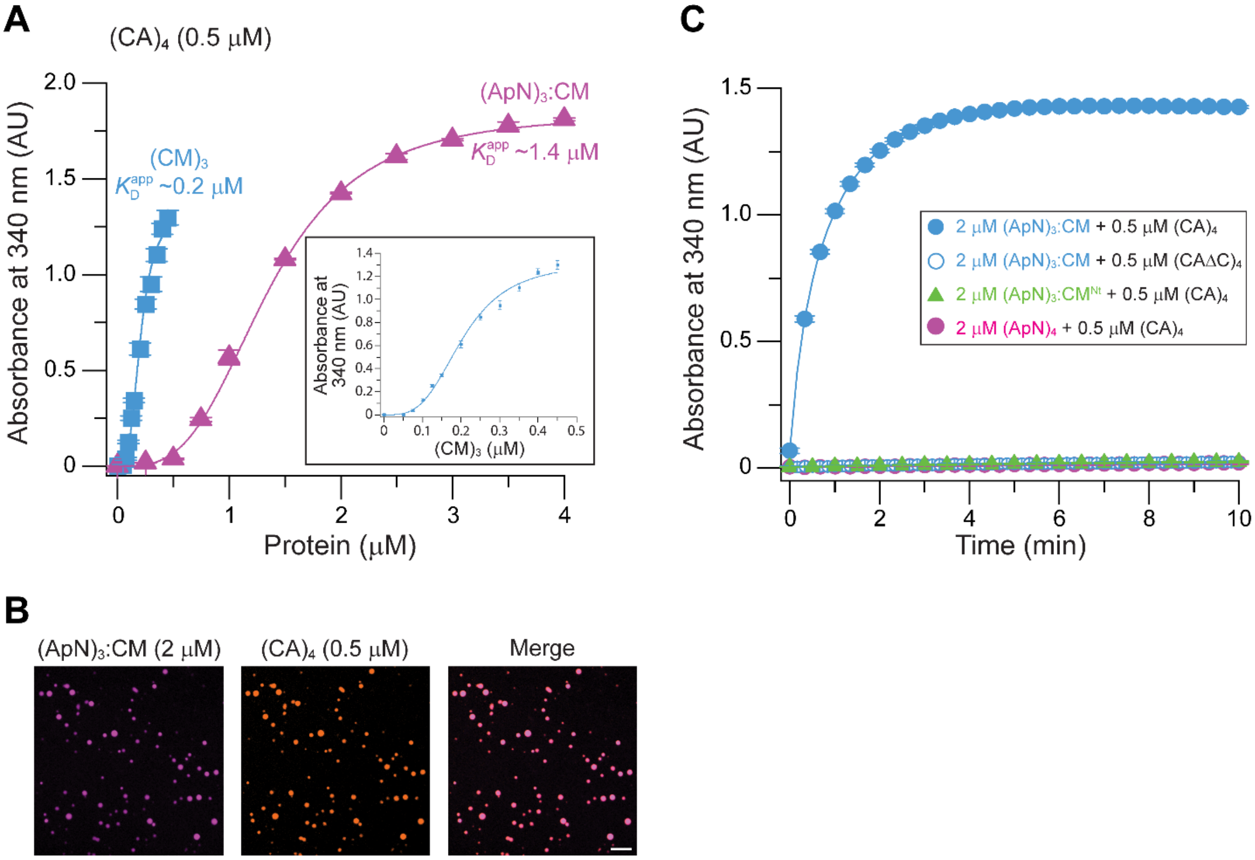


Figure 3.17 Interaction between (ApN)$_3$:CM and (CA)$_4$.
(A) A comparison of binding affinities between (CM)$_3$/(ApN)$_3$:CM with (CA)$_4$. Turbidity was measured under reducing conditions with 100 mM KCl, using 0.5 µM (CA)$_4$ and 0 to 4 µM (ApN)$_3$:CM, or 0 to 0.45 µM (CM)$_3$. Insert shows an expanded view of the titration of (CM)$_3$. Absorbance values reached after 10 min are plotted. Data are mean ± s.d. of triplicate measurements (n=3). **(B)** Fluorescence microscopy analysis of (ApN)$_3$:CM-(CA)$_4$ condensate formation. Condensate formation was analyzed by fluorescence microscopy of (ApN)$_3$:CM and (CA)$_4$ under reducing conditions with 100 mM KCl. (ApN)$_3$:CM and (CA)$_4$ were labeled with the fluorophores AF405 and AF647, respectively. Labeled proteins were used at a 1:10 ratio, mixed with respective unlabeled proteins and analyzed at the concentrations indicated. Representative data from two independent experiments are shown (n=2). Scale bar, 10 µm. **(C)** Dependence of condensate formation on the presence of γCAL domains and SSUL modules of CM protomers and the C-termini of (CA)$_4$. Co-assembly of either (CA)$_4$ (0.5 µM) or (CAΔC)$_4$ (0.5 µM) with (ApN)$_3$:CM (2 µM), and (CA)$_4$ (0.5 µM) with either (ApN)$_3$:CM^{Nt} (2 µM) or (ApN)$_4$ (2 µM) was monitored at 340 nm in a buffer containing 100 mM KCl/5 mM DTT. Data are mean ± s.d. of triplicate measurements (n=3).

To further analyze and confirm the stoichiometry of the subunits of $(ApN)_3$:CM and $(CA)_4$, we performed isothermal titration calorimetry (ITC) in the presence of the reducing agent TCEP (1 mM). We used $(ApN)_3:CM^{Nt}$, which lacks the SSUL modules, to avoid condensate formation during the ITC measurements (Figure **3.18A**). We performed an ITC assay with $(ApN)_3:CM^{Nt}$ and monomeric protein $E_{GFP}CA_{C17}$ (Zang et al., 2021), in which the enhanced green fluorescent protein (E_{GFP}) was fused to the N terminus of the last 17 residues of CA (CA_{C17}) via a flexible linker. The binding affinity (K_D) of $(ApN)_3:CM^{Nt}$ and $E_{GFP}CA_{C17}$ is ~24 µM and at a stoichiometry of 1:1, confirming that there is only one CM protomer in $(ApN)_3$:CM. Next, an ITC assay of $(ApN)_3:CM^{Nt}$ and $(CA)_4$ showed a K_D of ~2.25 µM and a binding ratio of 4:1 (Figure **3.18B**), supporting the previous finding that the *Se*7942 CA enzyme is tetrameric.

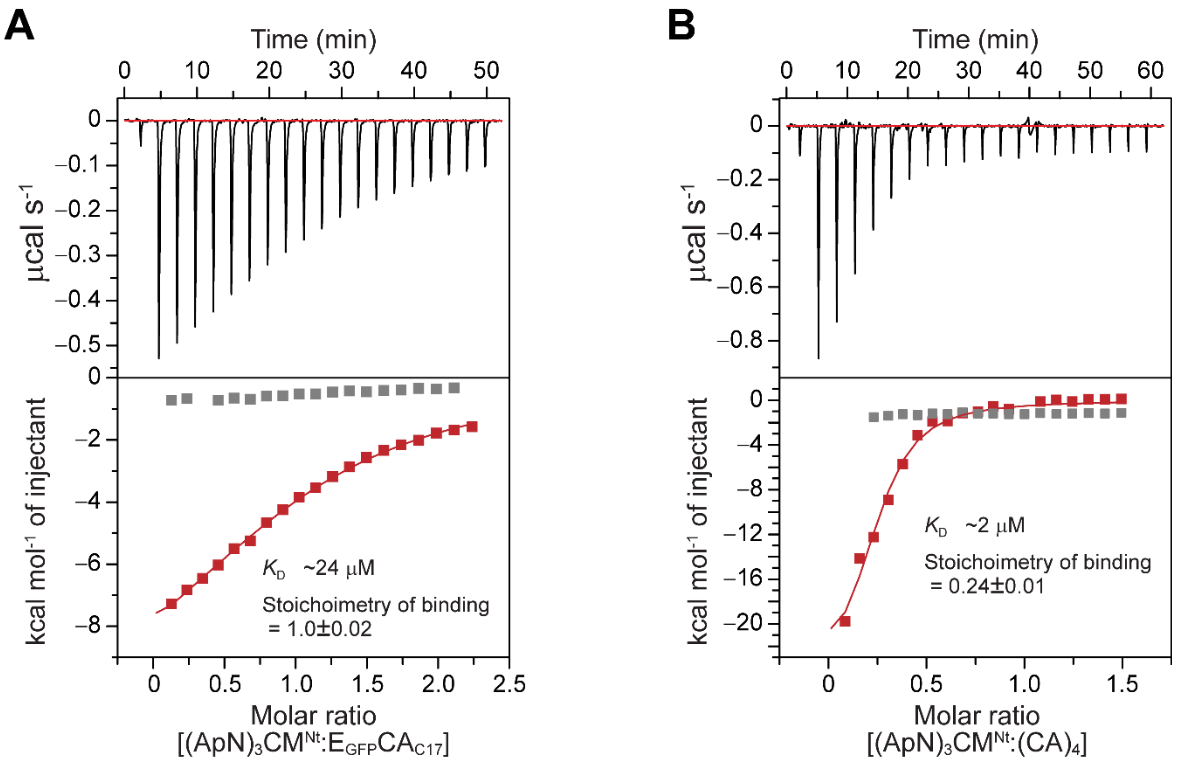


Figure 3.18 Analysis of affinity and stoichiometry of $(ApN)_3$:CM and $(CA)_4$ complex formation by isothermal titration calorimetry (ITC).

(A-B) ITC of the interaction of $E_{GFP}CA_{C17}$ with $(ApN)_3:CM^{Nt}$ **(A)**, and $(CA)_4$ with $(ApN)_3:CM^{Nt}$ **(B)**. Top: Change in enthalpy between sample and reference buffer upon titration of $E_{GFP}CA_{C17}$ **(A)** or $(CA)_4$ **(B)** into buffer containing $(ApN)_3:CM^{Nt}$. Bottom: Fitted integrated data generated in OriginPro 2020. Representative data from two independent experiments are shown (n=2).

3.8. $(ApN)_3$:CM incorporates into the periphery of the pro-β-carboxysome condensate

When the four proteins, including Rubisco, CM^{Ct}, $(CM)_3$, and $(CA)_4$ were mixed together, they co-assembled to form the pro-carboxysome biomolecular condensate, with all of these proteins distributing evenly throughout the droplet (Zang et al., 2021) (Figure **3.4C**). Studies have shown that β-carboxysome formation follows an 'inside-out' mechanism, with the aggregation of Rubisco via interactions with the SSUL modules of CM^{Ct} and $(CM)_3$ forming the initial condensate, followed by shell formation when shell proteins are recruited by the adaptor proteins ApN, to form the mature carboxysome (Cameron et al., 2013; Chen et al., 2013; Sun et al., 2021). The exact stoichiometry of the essential core components within a functional β-carboxysome remains unclear (Liu et al., 2021; Long et al., 2007; Long et al., 2011; Sun et al., 2024; Sun et al., 2019). However, current evidence suggests a hierarchical abundance: Rubisco and CM^{Ct} are the most abundant components, followed by $(CM)_3$, with $(CA)_4$ and ApN present in significantly lower quantities (Sun et al., 2024). In addition, genetic studies notably demonstrate that neither CA nor ApN are required for the initial formation of the Rubisco-enriched condensate with CM (Cameron et al., 2013; Price & Badger, 1989c; So et al., 2002; Suzuki et al., 1991). Based on these observations, we speculated that there must be a stepwise assembly pathway to ensure that ApN is positioned at the periphery of the pro-β-carboxysome, where it binds to shell proteins and initiates shell formation (Aussignargues et al., 2015; Cai et al., 2016; Cameron et al., 2013; Kinney et al., 2012; Lin et al., 2014). We would then like to reconstitute the pro-β-carboxysome *in vitro* following a stepwise mechanism with individually purified proteins to analyze at which step the $(ApN)_3$:CM hetero-complex is incorporated.

We first combined Rubisco, CM^{Ct}, and $(CM)_3$ for fluorescence microscopy analysis. Note that no condensate formation was observed with the same concentrations of CM^{Ct} and $(CM)_3$ in the absence of Rubisco (Figure **3.19A**). Fluorescence microscopy analysis of the mixture of Rubisco, CM^{Ct}, and $(CM)_3$ showed that the proteins co-assembled into droplet-shaped condensates and evenly distributed throughout the droplets (Figure **3.19B**). We carried out fluorescence recovery after photobleaching (FRAP) to monitor the mobility of Rubisco, CM^{Ct} or $(CM)_3$ in the condensate. We found that Rubisco and $(CM)_3$ were immobile in the condensate with no recovery observed in 5 min, while CM^{Ct} displayed some mobility with recovery within 5 min (Figure **3.19C-H**). Interestingly, when we mixed Rubisco and CM^{Ct}

together, followed by the addition of $(CM)_3$, $(CM)_3$ localized on the periphery of the homogeneous condensate formed by Rubisco and CM^{Ct} (Figure **3.19I**), indicating that the entry of $(CM)_3$ was restricted. Note that Rubisco is immobile in the Rubisco/CM^{Ct} condensate, while CM^{Ct} is mobile (Zang et al., 2021). In contrast, the Rubisco/CM^{Ct}/$(CM)_3$ condensate allowed CM^{Ct} to freely diffuse, but not $(CM)_3$ (Figure **3.19J**), consistent with the gel-like properties of the condensate.

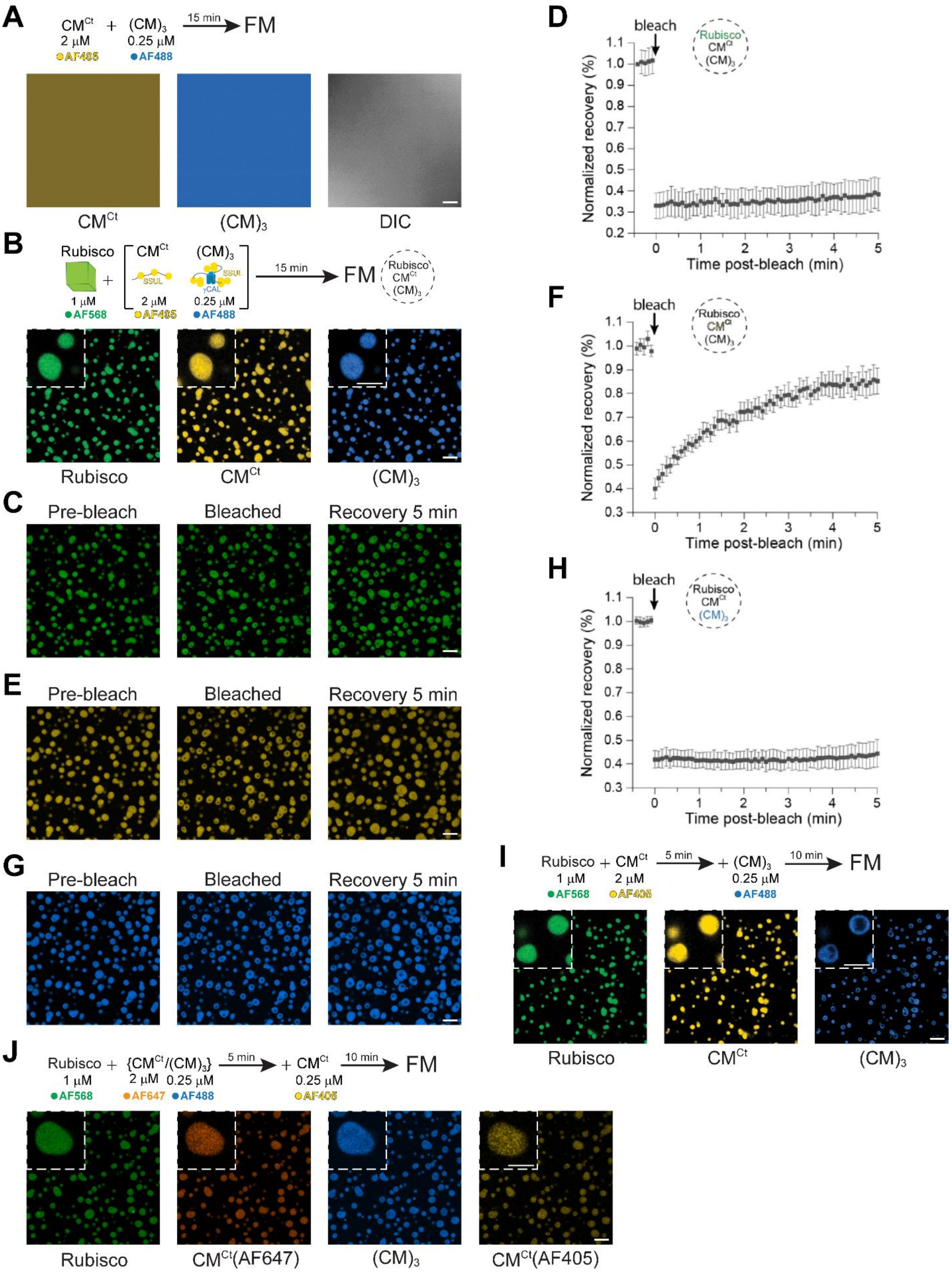


Figure 3.19 Fluorescence microscopy analysis of condensate formation by Rubisco, CM^{Ct}, and $(CM)_3$.

(A-B) Condensate formation analyzed by fluorescence microscopy under reducing conditions in the presence of 100 mM KCl. Mixtures of CM^{Ct} and $(CM)_3$ **(A)**, and condensate formation of Rubisco, CM^{Ct}, and $(CM)_3$ upon Rubisco and mixture of CM^{Ct} and $(CM)_3$ **(B)** analyzed by fluorescence microscopy. The experimental scheme and the concentrations of the proteins and fluorophores used are indicated for each respective co-assembly experiment. Labeled proteins were used at a 1:10 ratio with the respective unlabeled protein. Representative data from two or three independent experiments are shown (n=2 or 3). Scale bar, 10 µm. Inserts show magnified condensates; scale bar, 5 µm. **(C-H)** Mobility of Rubisco, CM^{Ct} and $(CM)_3$ in the initial condensate. Fluorescence recovery after photobleaching (FRAP) experiments are shown for condensates formed with Rubisco in the presence of both CM^{Ct} and $(CM)_3$. The final protein concentrations were 1 µM Rubisco, 0.25 µM $(CM)_3$, and 2 µM CM^{Ct}, respectively. FRAP analysis of Rubisco is shown in **(C-D)**, CM^{Ct} in **(E-F)** and $(CM)_3$ in **(G-H)**. Rubisco, CM^{Ct} and $(CM)_3$ are labeled with AF405, AF405 and AF488, respectively. Labeled proteins were used at a 1:10 ratio with the respective unlabeled protein. Pre-bleach fluorescence was set to 1. Mean ± s.d. from n=30 droplets. Scale bar, 10 µm. **(I)** Condensate formation analyzed by fluorescence microscopy under reducing conditions in the presence of 100 mM KCl by Rubisco, CM^{Ct}, and $(CM)_3$ upon subsequent addition of $(CM)_3$ to mixture of Rubisco and CM^{Ct}. The experimental schemes and the concentrations of the proteins and fluorophores used are indicated for each respective co-assembly experiment. Labeled proteins were used at a 1:10 ratio with the respective unlabeled protein. Representative data from two or three independent experiments are shown (n=2 or 3). Scale bar, 10 µm. Inserts show magnified condensates; scale bar, 5 µm. **(J)** Analysis of gel-like property of the initial condensate containing Rubisco, CM^{Ct} and $(CM)_3$ by fluorescence microscopy. The experimental scheme and the concentrations of the proteins and fluorophores used are indicated for each respective co-assembly experiment. Labeled proteins were used at a 1:10 ratio with the respective unlabeled protein. Representative data from three independent experiments are shown (n=3). Scale bar, 10 µm. Inserts show magnified condensates; scale bar, 5 µm.

Next, since $(ApN)_3$:CM is essential for recruiting shell proteins to initiate subsequent shell formation, we hypothesize that it incorporates into the pro-carboxysome simultaneously with or after $(CA)_4$. This sequential addition, whereby $(CA)_4$ joins the condensate simultaneously or before $(ApN)_3$:CM, would ensure correct spatial organization. When we added $(CA)_4$ to the condensate formed by Rubisco, CM^{Ct}, and $(CM)_3$, $(CA)_4$ consistently localized to the rim of the droplet independent of whether $(CM)_3$ was evenly distributed or localized to the periphery (Figure **3.20A-C**), suggesting that enough $(CM)_3$ molecules are exposed at the outer rim for $(CA)_4$ to bind via the interactions between the C-termini of $(CA)_4$ and the γCAL domains of $(CM)_3$. In contrast, when we mixed Rubisco and $(CA)_4$ together before adding CM^{Ct} and $(CM)_3$, the proteins distributed evenly (Figure **3.20D**), consistent with what was observed when all of these proteins were mixed together (Zang et al., 2021) (Figure **3.4C**).

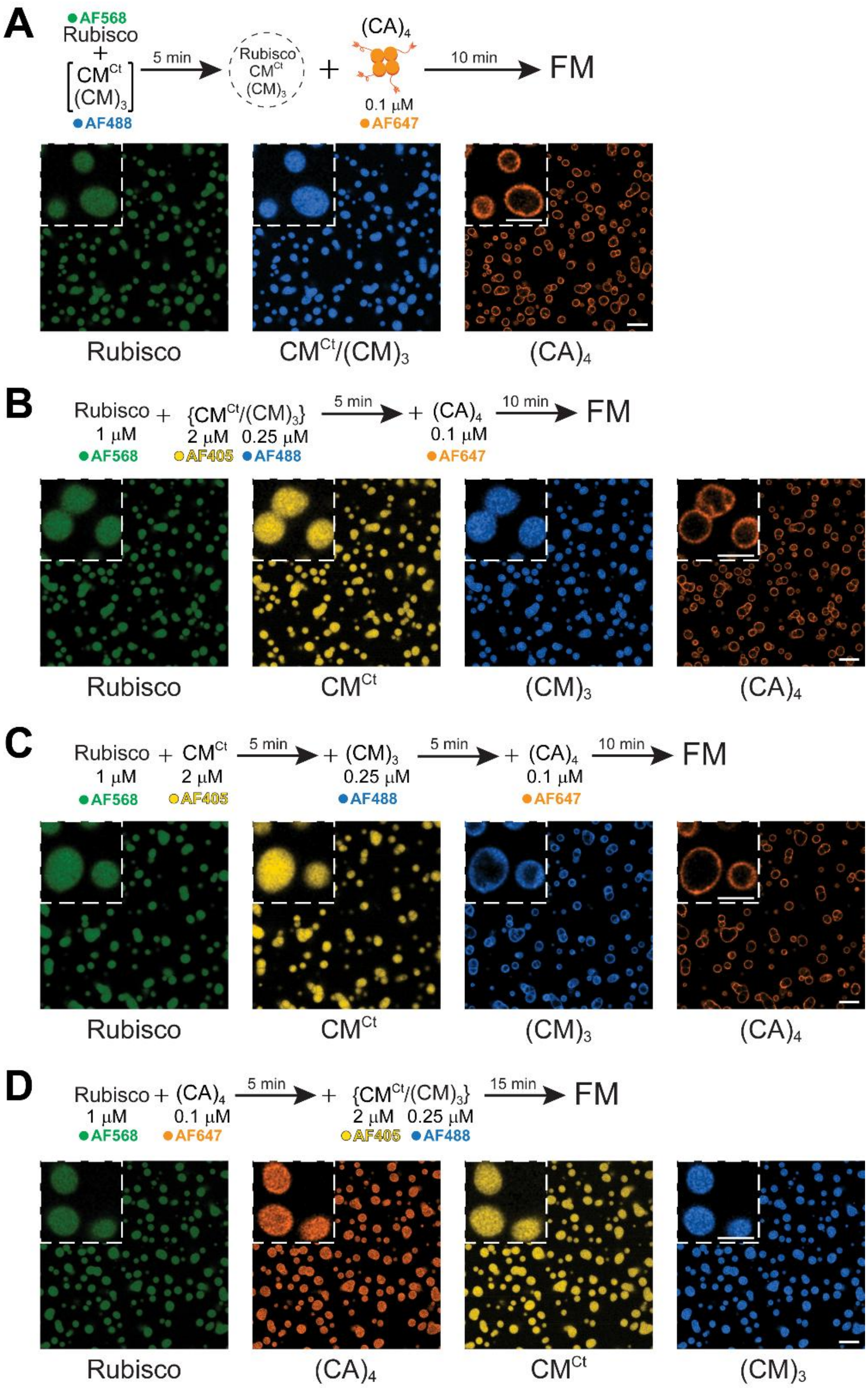


Figure 3.20 Fluorescence microscopy analysis of condensate formation by Rubisco, CM^{Ct}, (CM)$_3$, and (CA)$_4$.

(A-D) Condensate formation with Rubisco, CM^{Ct}, (CM)$_3$, and (CA)$_4$ analyzed by fluorescence microscopy under reducing conditions with 100 mM KCl. CM^{Ct} was labeled with AF488 **(A)** and AF405 **(B)**, respectively. The experimental schemes and the concentrations of the proteins and fluorophores used are indicated for each respective co-assembly experiment. Labeled proteins were used at a 1:10 ratio with the respective unlabeled protein. Representative data from two or three independent experiments are shown (n=2 or 3). Scale bar, 10 µm. Inserts show magnified condensates; scale bar, 5 µm.

Next, to explore the incorporation of the (ApN)$_3$:CM hetero-complex, we added (ApN)$_3$:CM to the condensate formed by Rubisco, CM^{Ct}, and (CM)$_3$. Regardless of whether (CM)$_3$ was evenly distributed or enriched at the periphery, (ApN)$_3$:CM consistently incorporated at the rim of the condensate (Figure **3.21A-C**). Analysis of the (ApN)$_3$:CM fluorescence profile across the droplets, clearly showed that the fluorescence intensity was higher at the condensate rim than at the core (Figure **3.21A-C**).

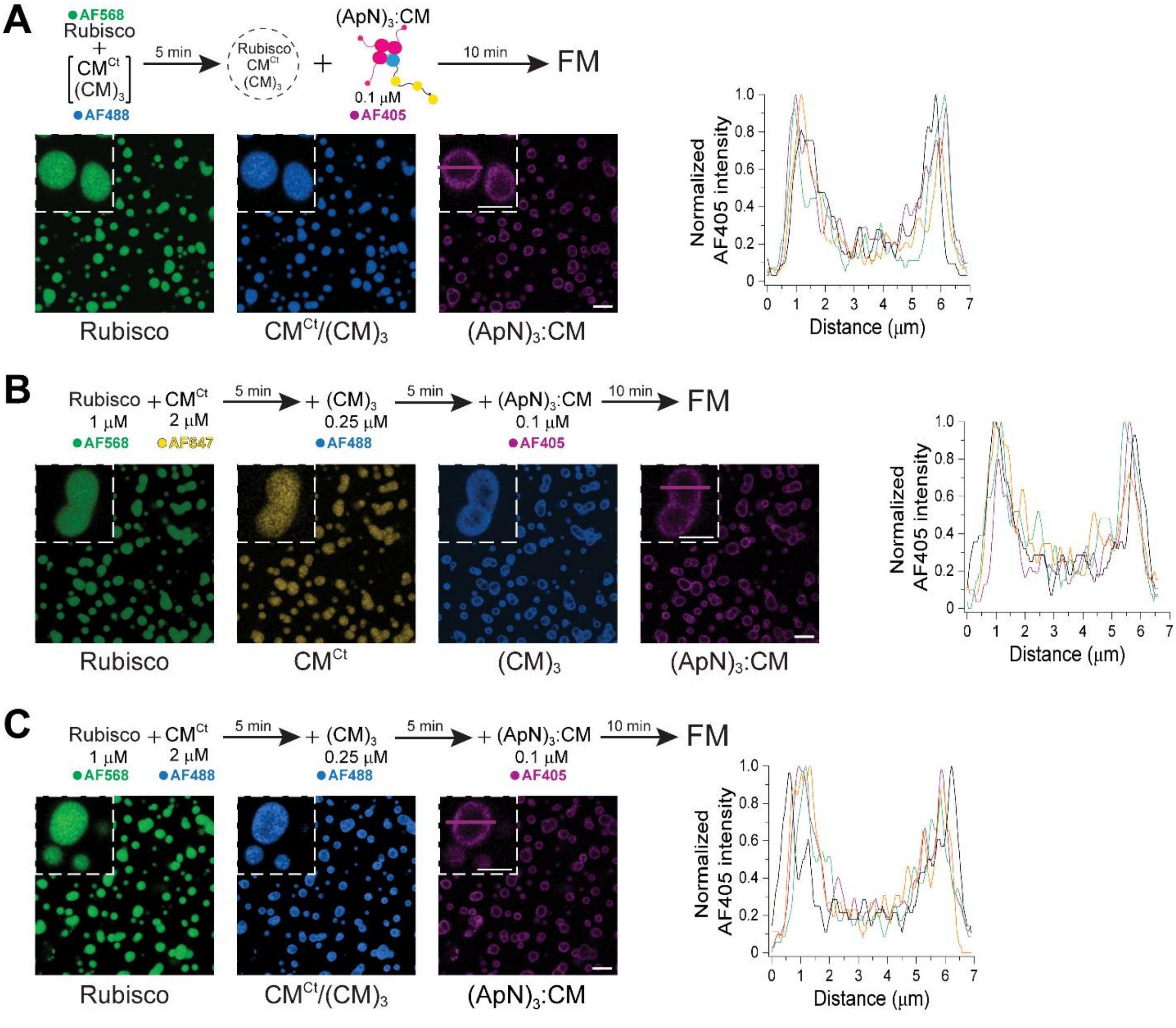


Figure 3.21 Fluorescence microscopy analysis of condensate formation by Rubisco, CM^{Ct}, (CM)$_3$, and (ApN)$_3$:CM.
(A-C) Condensate formation with Rubisco, CM^{Ct}, (CM)$_3$, and (ApN)$_3$:CM analyzed by fluorescence microscopy under reducing conditions with 100 mM KCl. CM^{Ct} was labeled with AF488 **(A) (C)** and AF405 **(B)**, respectively. The experimental schemes and the concentrations of the proteins and fluorophores used are indicated for each respective co-assembly experiment. Labeled proteins were used at a 1:10 ratio with the respective unlabeled protein. Representative data from two or three independent experiments are shown (n=2 or 3). Scale bar, 10 µm. Inserts show magnified condensates; scale bar, 5 µm. The distribution of (ApN)$_3$:CM within the

condensates was determined by measuring the AF405 fluorescence signal across 4 randomly selected droplets. The magenta profile shows the signal intensity along the magenta line drawn through each droplet.

Finally, we added $(ApN)_3$:CM to the condensate formed by Rubisco, CM^{Ct}, $(CM)_3$ and $(CA)_4$. The final reaction system included five proteins, but only four fluorophores were used to prevent interference from the absorbance of the dyes. Nevertheless, the involvement of the proteins in condensate formation was rigorously validated (Figures **3.20B,C** and **3.21B**). As expected, both $(CA)_4$ and $(ApN)_3$:CM localized to the periphery of the condensate, independent of the localization of $(CM)_3$ (Figure **3.22A,B**). However, when Rubisco and $(CA)_4$ were mixed together, followed by the addition of CM^{Ct} and $(CM)_3$, and then $(ApN)_3$:CM, $(ApN)_3$:CM was only weakly incorporated at the rim of the phase-separated droplets (Figure **3.22C**). These results suggest that the localization of $(CA)_4$ at the rim of the condensate allows for the efficient binding of $(ApN)_3$:CM, but an even distribution of $(CA)_4$ through the condensate weakens the preference of $(ApN)_3$:CM for the periphery, which may interfere with the subsequent sequestration of shell proteins and impact the efficiency of carboxysome formation. Interestingly, when $(CA)_4$ was added after $(ApN)_3$:CM, it efficiently localized to the rim and displaced $(ApN)_3$:CM to the condensate core (Figure **3.22D**), while this effect was not observed when $(CA)_4$ and $(ApN)_3$:CM were added simultaneously (Figure **3.22E**).

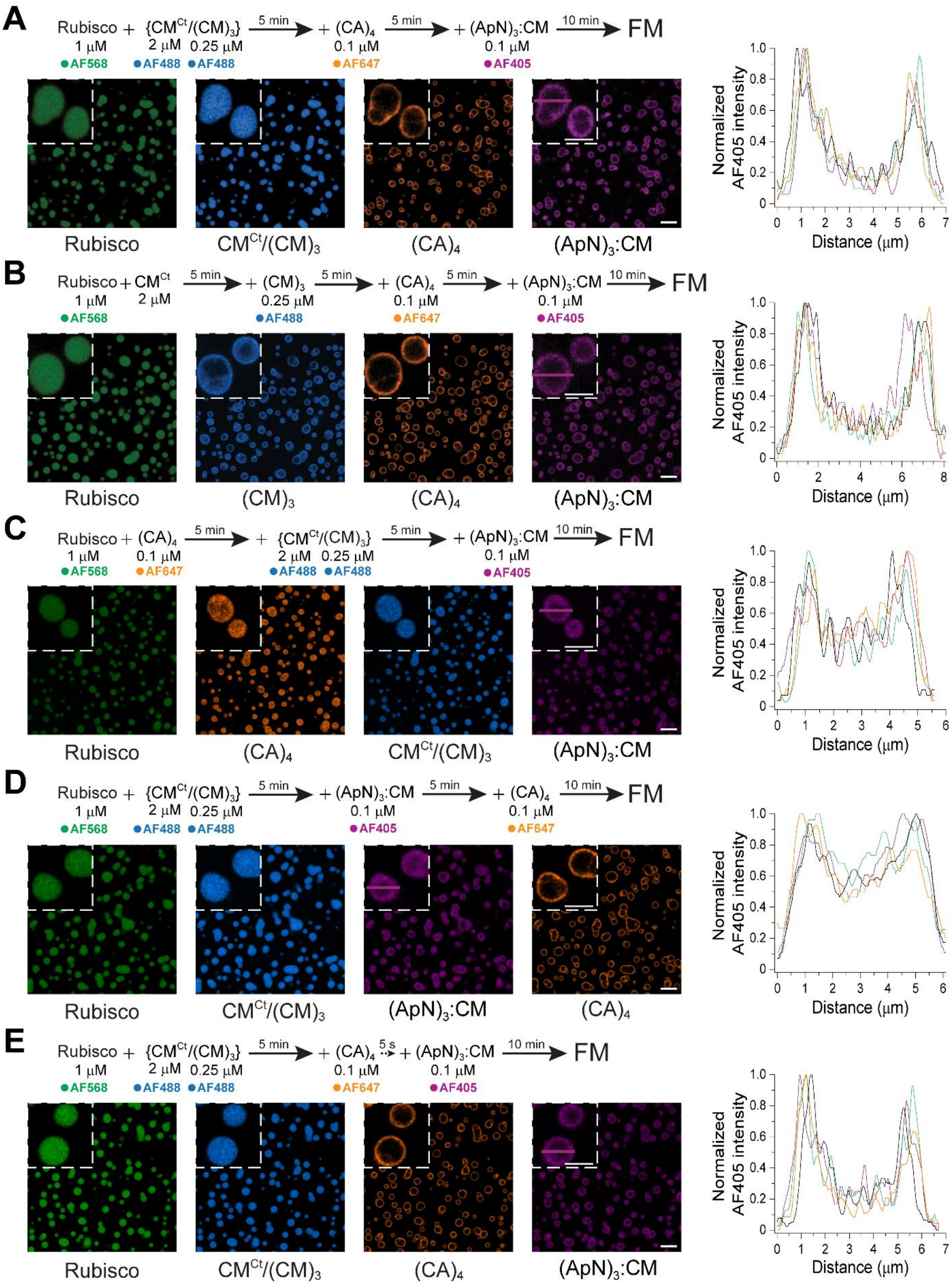


Figure 3.22 Fluorescence microscopy analysis of condensate formation by Rubisco, CM^{Ct}, (CM)$_3$, (CA)$_4$, and (ApN)$_3$:CM.

(A-E) Condensate formation with Rubisco, CM^{Ct}, (CM)$_3$, (CA)$_4$, and (ApN)$_3$:CM analyzed by fluorescence microscopy under reducing conditions with 100 mM KCl. The experimental schemes and the concentrations of the proteins and fluorophores used are indicated for each respective co-assembly experiment. Labeled proteins were used at a 1:10 ratio with the respective unlabeled protein. Representative data from two or three independent experiments are shown (n=2 or 3). Scale bar, 10 µm. Inserts show magnified condensates; scale bar, 5 µm. The distribution of $(ApN)_3$:CM within the condensates was determined by measuring the AF405 fluorescence signal across 4 randomly selected droplets. The magenta profile shows the signal intensity along the magenta line drawn through each droplet.

In summary, the results support a stepwise assembly mechanism for the pro-β-carboxysome. First, Rubisco, CM^{Ct}, and $(CM)_3$ form an initial condensate. Then, $(CA)_4$ is recruited, either simultaneously with or prior to $(ApN)_3$:CM, positioning $(ApN)_3$:CM at the periphery of the pro-carboxysome condensate. This arrangement will facilitate the subsequent recruitment of shell proteins and assembly of the shell.

4. Discussion

In this study, we investigated when and how the carboxysome shell adaptor protein ApN participates in pro-β-carboxysome assembly. We successfully obtained a tetrameric complex of ApN, $(ApN)_4$, from *Se*7942 through overexpression in *E.coli*. Surprisingly, pull-down assays and fluorescence microscopy analysis failed to reveal an interaction between $(ApN)_4$ and pro-carboxysome core proteins, including Rubisco, CM^{Ct}, $(CM)_3$, and $(CA)_4$. Given that *ccmN* is consistently located downstream of *ccmM* in the *ccm* operon (Figure **3.5A**), we reasoned that the two genes may be structurally linked for function. However, no subunit exchange was observed between $(ApN)_4$ and $(CM)_3$ under reducing or oxidizing conditions (Figure **3.5B**). Previously, a hetero-trimeric $(SUMO\text{-}ApN)_2{:}CM^{209}$ complex has been reported via the recombinant co-expression of the adaptor protein ApN and the scaffolding protein CM (Sun et al., 2021). These findings prompted us to employ a bicistronic plasmid to co-express ApN and CM. This approach yielded both a hetero-tetrameric $(ApN)_3{:}CM$ complex and a double-tagged hetero-tetrameric $(SIIApN)_3{:}CM\text{-}H_{10}$ complex in solution, different from the previously reported hetero-trimeric $(SUMO\text{-}ApN)_2{:}CM^{209}$ complex (Sun et al., 2021). To further investigate the oligomeric state of the CM-ApN hetero-complex, we performed cryo-EM single-particle analysis. However, 2D class averages did not reveal the presence of the hetero-tetrameric $(ApN)_3{:}CM$ complex. This will be discussed in Chapter **4.2**.

Based on the AlphaFold 3-predicted structural model of the $(ApN)_3{:}CM$ complex, which comprises ApN1 residues 1-114, ApN2 and ApN3 residues 11-114, and CM residues 16-182 (Figure **3.7A**), a strong interface is observed between wall A of ApN1 and the loops between walls D and F of CM, as well as a weak interface between wall C of ApN3 and wall D of CM. Mutating the conserved residues Leu76 or Thr94 on wall A of ApN to the charged residues Asp or Arg (SIIApN-L76D, SIIApN-L76R, SIIApN-T94R) abolished hetero-complex formation in the co-expression of $(CM^{Nt}\text{-}H_{10})$-SIIApN (Figure **3.8**). SEC-MALS analysis of the $(SIIApN)_3{:}CM^{Nt}\text{-}H_{10}$ complex and the $(SIIApN\text{-}2A)_3{:}CM^{Nt}\text{-}H_{10}$ complex under reducing and oxidizing conditions (Figure **3.10A**) revealed that cysteine oxidation could facilitate the release of one ApN protomer, converting the hetero-tetrameric $(SIIApN)_3{:}CM^{Nt}\text{-}H_{10}$ complex into the hetero-trimeric $(SIIApN)_2{:}CM^{Nt}\text{-}H_{10}$ complex. These results support the existence of the predicted weak interface and suggested that oxidation-induced dissociation of ApN3 from the complex is structurally plausible. *In vivo* complementation assays using the ΔK2-O *Se*7942 strain with pSE4*ccmK2-O* carrying a series of ApN mutants revealed that the ApN-L76R and ApN-T94R mutants could not rescue the HCR phenotype of the ΔK2-O *Se*7942 strain (Figure **3.12C**). Additionally, the ApN-2A and ApN-2S mutants resulted in slower growth compared to the ΔK2-O+pSE4*ccmK2-O* strain (Figure **3.13A**). Thus, ApN must form a hetero-complex with CM for carboxysome maturation and the conserved cysteines (Cys49 and Cys88) are essential for carboxysome function.

To investigate how the hetero-tetrameric $(ApN)_3$:CM complex is involved in pro-β-carboxysome formation, we performed fluorescence microscopy analysis. Our findings suggest that the adaptor protein ApN forms a hetero-complex with the scaffolding protein CM to participate in pro-β-carboxysome assembly via multiple protein-protein interactions. The stepwise assembly process of the pro-β-carboxysome begins with the SSUL modules of CM^{Ct} and $(CM)_3$ interacting with Rubisco to form the initial condensate (Figure **4.1**, stage 1). Next, the low-abundant carbonic anhydrase $(CA)_4$ and $(ApN)_3$:CM complexes are incorporated into the rim of the condensate (Figure **4.1**, stage 2). $(CA)_4$ binds either before or simultaneously with $(ApN)_3$:CM. These incorporations are mediated through multiple interactions: (i and ii) the C-terminal tail (C2) of $(CA)_4$ binding to the γCAL domains of $(CM)_3$ and to the γCAL domain of CM in $(ApN)_3$:CM, (iii) the SSUL modules of $(CM)_3$ binding to the γCAL domain of CM in $(ApN)_3$:CM or the SSUL modules of CM in $(ApN)_3$:CM binding to the γCAL domains of $(CM)_3$, and (iv) the SSUL modules of CM in $(ApN)_3$:CM binding to Rubisco. Together, these interactions drive the formation of the pro-β-carboxysome. The peripheral localization of the $(ApN)_3$:CM hetero-complexes exposes the encapsulation peptide of ApN, enabling recruitment of various shell proteins to form the mature carboxysomes (Figure **4.1**, stage 3). This assembly pathway is essential for establishing the proper spatial organization of the pro-β-carboxysome, positioning the hetero-tetrameric $(ApN)_3$:CM complex in the outer region and enabling the recruitment of shell proteins for carboxysome maturation. However, the precise mechanism by which shell proteins are recruited remains unclear.

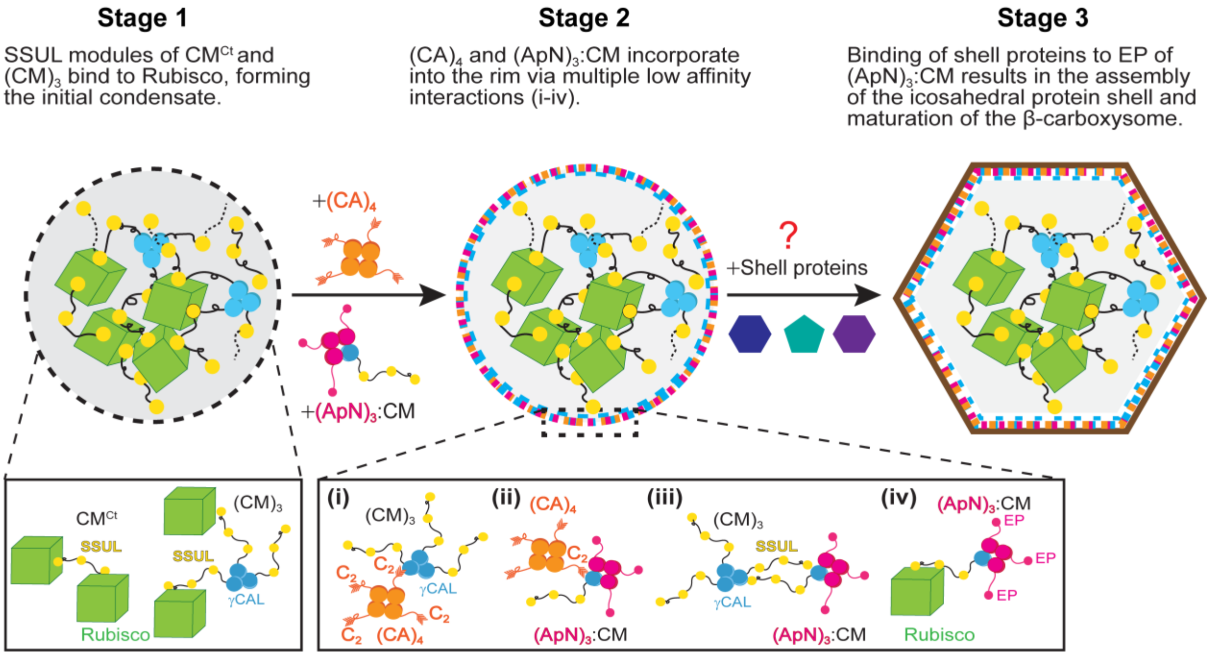


Figure 4.1 Model of the stages of pro-β-carboxysome assembly and carboxysome maturation.

In stage 1, Rubisco is sequestered via the interaction with the SSUL modules of CM^{Ct} and $(CM)_3$ to form the initial condensate. In stage 2, the low-abundance $(CA)_4$ and the $(ApN)_3$:CM hetero-tetramer are recruited to the rim of the condensate of stage 1 via multiple protein-protein interactions, involving binding of C2 peptides of $(CA)_4$ to the γCAL domains of $(CM)_3$ (i) and to the γCAL domain of CM in $(ApN)_3$:CM (ii), binding of SSUL modules of $(CM)_3$ to the γCAL domain of CM in $(ApN)_3$:CM or binding of SSUL modules of CM in $(ApN)_3$:CM to the γCAL domains of $(CM)_3$ (iii), binding of SSUL modules of CM in $(ApN)_3$:CM to Rubisco (iv). In stage 3, the peripheral localization of $(ApN)_3$:CM enables the recruitment of various shell proteins for carboxysome maturation.

Our findings clarify the role of the adaptor protein ApN in pro-β-carboxysome assembly, as well as the stage during assembly at which ApN becomes involved. Furthermore, our results shed light on the stepwise assembly mechanism of pro-β-carboxysome. These insights advance our understanding of β-carboxysome biogenesis and provide a valuable foundation for future efforts in carboxysome bioengineering. However, several important questions remain unanswered, underscoring the necessity of further investigation into the molecular mechanisms that govern carboxysome biogenesis.

4.1. Cryo-EM single-particle analysis of $(ApN)_4$ and $(ApN)_3$:CM

To mitigate the preferred orientation of $(ApN)_4$ in cryo-EM samples caused by its interactions with the air-water interface, we employed several strategies, including polylysine treatment of grids to enhance sample adherence and promote uniform distribution; graphene oxide-coated grids to absorb protein away from the air-water interface; and 45° tilt data collection to increase angular sampling (see Appendix **A.4.1** for details). However, none of these approaches were effective for such a small complex (~65 kDa; ~40 kDa of the rigid body). In particular, the 45° tilt resulted in a low signal-to-noise ratio. A similar challenge arose when we attempted to determine the structure of the hetero-tetrameric $(ApN)_3$:CM complex. Despite adding the detergent OG to form a protective layer at air-water interface during cryo-EM sample preparation, exposure still led to dissociation of the complex into the hetero-trimeric $(ApN)_2$:CM complex. This complex then reassociated into the hetero-hexameric $(ApN)_4$:CM_2 complex, which also displayed a strongly preferred orientation. Using CHAPSO for cryo-EM sample preparation, even at a working concentration lower than its critical micelle concentration (CMC), resulted in blurred particles. Even worse, when graphene oxide-coated grids were used, the particles were completely disrupted into separate subunits (see Appendix **A.4.2** for details).

It is estimated that protein particles in thin aqueous layers (less than 100 nm) diffuse to and are captured by the air-water interface in less than a millisecond (Liu & Wang, 2023). Since current vitrification devices operate on timescales longer than this, contact between particles and the air-water interface is inevitable. Recent advances in fast vitrification devices, which can freeze sample within a sub-second to millisecond range and proceed with no need of the blotting process, have helped mitigate preferred orientation in some cases; however, even these rapid methods cannot prevent protein particles from touching the air-water interface, even at sub-millisecond time intervals (Klebl et al., 2022).

Due to the small size of $(ApN)_4$ and the small size and instability of $(ApN)_3$:CM, we were unable to solve their structures, and further confirm the composition of the CM-ApN hetero-complex. The S-shaped dimer-of-trimers $(SUMO\text{-}ApN)_2$:CM^{209} observed in the asymmetric unit of the crystal structure (Sun et al., 2021) and the dominant classes of S-shaped dimer-of-trimers $(ApN)_2$:CM identified in our cryo-EM single-particle analysis may be artifacts due to the high local concentrations of proteins during crystallization (Vekilov & Vorontsova, 2014) or cryo-EM sample preparation (Schmidt-Krey & Rubinstein, 2011). Alternatively, the conformational change from the labile $(ApN)_3$:CM to the hetero-trimer and then to the dimer-of-trimers could play a functional role in stabilizing the complex. For instance, this change could increase the avidity of the EP peptide of ApN for recruiting shell proteins, thereby promoting carboxysome maturation. Further studies, perhaps aided by advances in high-resolution imaging techniques, may help us to solve the structure of the CM-ApN hetero-complex in the functional carboxysome.

4.2. The role of conserved cysteines in $(ApN)_3$:CM

Cysteine oxidation has important physiological relevance in various cellular processes, including redox regulation of protein function, oxidative stress sensing and protection, signal transduction, metal coordination and enzymatic activity, as well as protein-protein interactions and structural transitions (Garrido Ruiz et al., 2022). The mature carboxysome provides an oxidizing environment (Chen et al., 2013; Pena et al., 2010; Price et al., 1992), largely due to its selective permeability, which excludes the entry of thioredoxin and other redox equivalents into the interior (Pena et al., 2010). Within this microenvironment, CA is redox-regulated and becomes catalytically active only in the oxidizing environment of the carboxysomal lumen (Pena et al., 2010; Price et al., 1992). Similarly, disulfide bond formation under redox regulation in the SSUL modules of CM^{Ct} and $(CM)_3$ is essential for carboxysome biogenesis and CCM function (Wang et al., 2019; Zang et al., 2021). To evaluate the impact of cysteine oxidation within the $(ApN)_3$:CM complex, we treated the $(CM^{Nt}\text{-}H_{10})$-SIIApN hetero-complex with H_2O_2, which facilitated the conversion of the $(SIIApN)_3:CM^{Nt}\text{-}H_{10}$ complex to the $(SIIApN)_2:CM^{Nt}\text{-}H_{10}$ complex and resulted in the surface exposure of Cys49 and Cys88 in ApN. *In vivo* complementation with 2A and 2S mutants, in which the two conserved cysteines Cys49 and Cys88 are substituted to Ala or Ser, respectively, led to reduced growth rates compared to the complementation with wild-type K2-O strain. These results suggest that the conserved cysteines Cys49 and Cys88 in ApN may play a role *in vivo* to ensure timely carboxysome maturation, probably by mediating conformational changes of the CM-ApN hetero-complex. Further studies are needed to determine whether the two conserved cysteines have additional functions.

4.3. Localization of $(CM)_3$ and $(CA)_4$ in the β-carboxysome

Our study revealed that the order in which $(CM)_3$ is added significantly affects its spatial distribution within the Rubisco-CM^{Ct}/$(CM)_3$ condensate, as determined by fluorescence microscopy analysis. When $(CM)_3$ was mixed with CM^{Ct} prior to the addition of Rubisco, it distributed evenly throughout the condensate. However, when $(CM)_3$ was added to the preformed Rubisco-CM^{Ct} condensate, it localized in a ring-like pattern at the periphery. The position of $(CM)_3$ within the β-carboxysome is not determined yet.

CA catalyzes the conversion of HCO_3^- to CO_2 in the oxidizing environment of the carboxysome (Pena et al., 2010; Price et al., 1992). It is known that in some cyanobacterial species, the N-terminal γCAL domain of $(CM)_3$ is inactive as carbonic anhydrase, while interestingly, in some species that lack a separate CA, instead, they employ $(CM)_3$ as the active CA (Pena et al., 2010). Interestingly, deleting CA does not alter the size, shape, and abundance of carboxysomes in cells compared to the wild-type strain, but resulted in an HCR phenotype (Price & Badger, 1989c; So et al., 2002; Suzuki et al., 1991), suggesting that CA is not essential for the incorporation of the $(ApN)_3$:CM complex and the formation of carboxysomes but critical for their function. Our *in vitro* reconstitution of the pro-β-carboxysome supports this hypothesis that $(ApN)_3$:CM can be recruited to the biomolecular condensate formed by Rubisco, CM^{Ct}, and $(CM)_3$ in the absence of $(CA)_4$ (Figure **3.21**). In fluorescence microscopy analysis, the order in which $(CA)_4$ was added strongly influenced the organization of pro-β-carboxysome condensates. When $(CA)_4$ was added to Rubisco before condensate formation, this resulted in an even distribution of all proteins (Figure **3.22C**). However, adding $(CA)_4$ after the formation of Rubisco-CM^{Ct}/$(CM)_3$ condensates resulted in $(CA)_4$ localizing at the periphery (Figure **3.22A,B**). Additionally, when $(CA)_4$ was added to the preformed $(ApN)_3$:CM containing condensates, it efficiently localized to the rim and displaced $(ApN)_3$:CM to the condensate core (Figure **3.22D**). However, the precise sub-compartment localization of CA within β-carboxysomes remains unclear.

Some studies suggest that CA and CM are associated with the carboxysome shell to facilitate conversion of HCO_3^- to CO_2 (Cot et al., 2008; Long et al., 2011), while others propose that CM plays multiple structural and functional roles throughout the carboxysome and that CA interacts with CM throughout the internal matrix of Rubisco (Niederhuber et al., 2017). Immunogold labeling studies of α-carboxysomes revealed that CsoSCA is localized at the periphery of carboxysomes (Baker et al., 2000). More recently, cryo-ET analysis of α-carboxysomes has suggested that CA could be shell associated (Metskas et al., 2022). The theoretical structure-function model suggests that the cyanobacterial carboxysome operates as a discrete physico-metabolic unit, and that α- and β-carboxysomes operate similarly (Whitehead et al., 2014) (Figure **4.2A**). In this model, the subsurface-localized CA supplies CO_2 to the active sites of Rubisco to support optimal CO_2 fixation activity (Whitehead et al., 2014). From a functional point of view, CA, the relatively minor protein component of the carboxysome, may be strategically positioned throughout the carboxysome shell to access the intracellular HCO_3^- pool and efficiently channel CO_2 to Rubisco (So et al., 2004). Further investigations are required to clarify the sub-compartment localization and kinetic characteristics of $(CA)_4$ and/or $(CM)_3$, which would enhance our understanding of the assembly model of carboxysomes and the functional advantages conferred by their unique spatial organization.

Interestingly, the localization of CAH3 in the pyrenoid of *C. reinhardtii* has been determined by immunogold labeling experiments, indicating that it is primarily located in the intrapyrenoid thylakoids (pyrenoid tubules) and present in lower abundance in the chloroplast thylakoids (Sinetova et al., 2012) (Figure **4.2B**). This localization creates a concentrated source of CO_2 that diffuses into the surrounding pyrenoid matrix for capture by Rubisco (Catherall et al., 2025). The roles and localization of CAH3-like proteins in other organisms, such as diatoms and hornworts, require further investigation.

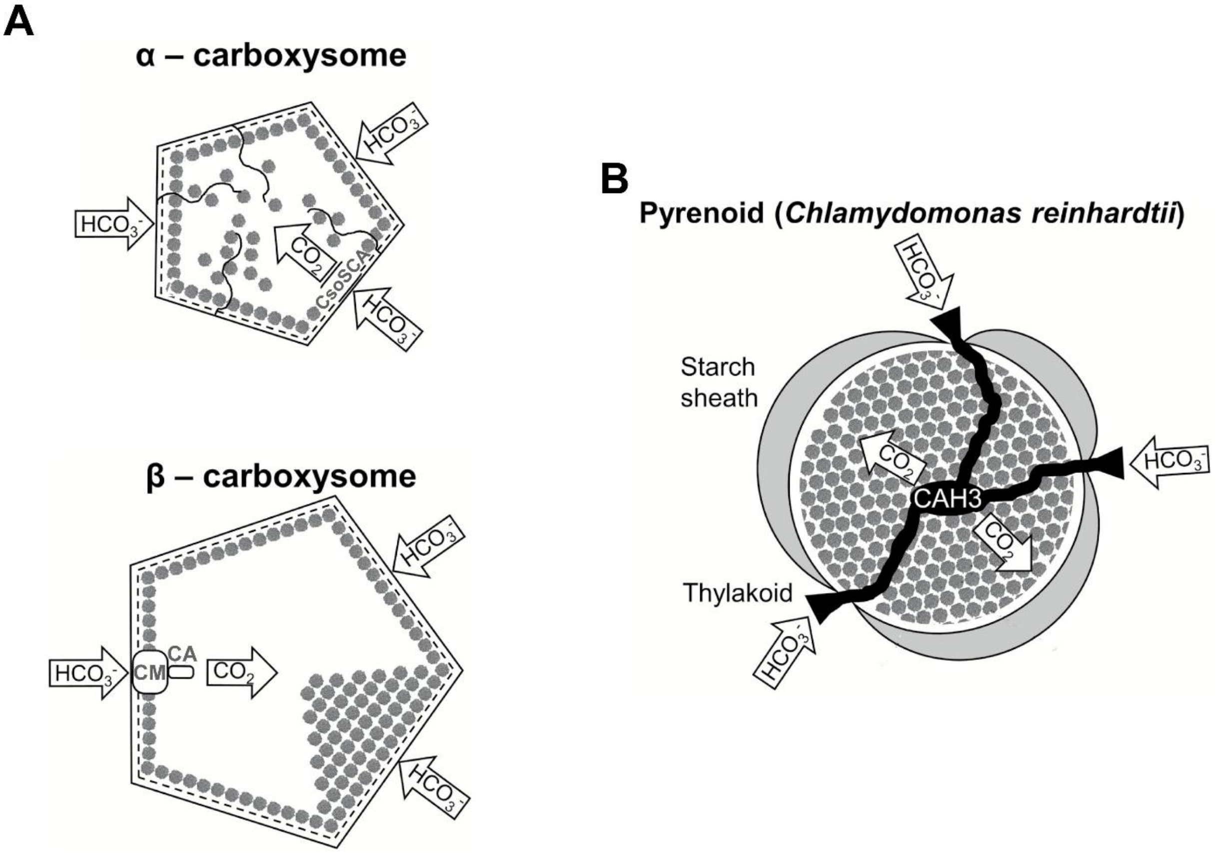


Figure 4.2 Schematic presentations of the substructural organization of the α- and β-carboxysomes, and the pyrenoid in *C. reinhardtii*.
(A-B) Carbon fixation proceeds by 'outside-in' and 'in-outside' concentrating mechanisms in carboxysomes and pyrenoids, respectively. **(A)** In both α- and β-carboxysomes, HCO_3^- is converted to CO_2 by one of several potential CA (CsoSCA in α-carboxysomes and CA and/or CM in β-carboxysomes), probably located at the subsurface of their carboxysome shell. **(B)** In *C. reinhardtii* pyrenoid, HCO_3^- is taken up into the thylakoid lumen and converted into CO_2 by CAH3 localized within the pyrenoid tubules. Figure modified from (Rae et al., 2017).

4.4. The interaction between $(ApN)_3$:CM and shell proteins

Our findings suggest that the $(ApN)_3$:CM complex is recruited to the rim of the pro-β-carboxysome condensate. This peripheral localization likely facilitates the recruitment of shell proteins to initiate shell assembly to form mature β-carboxysomes. However, given that only a few copies of ApN are observed in β-carboxysomes, this raises the question of how ApN can efficiently recruit sufficient shell proteins for carboxysome maturation. Notably, the interaction between ApN or CM and shell proteins is still being debated (Cai et al., 2016; Kinney et al., 2012; Lin et al., 2014; Sun et al., 2021). ApN has been reported to bind various combinations of shell proteins: CcmK2 (Kinney et al., 2012), CcmK2/L/O (Cai et al., 2016), CcmK2/O (Lin et al., 2014), and CcmL/O (Sun et al., 2021). The N-terminal γCAL domain of CM has also been shown to bind the abundant shell proteins CcmK1 (Cot et al., 2008; Long et al., 2007) and possibly also to CcmK2/K4/L (Cot et al., 2008), or CcmK2/L/O (Cai et al., 2016). However, the limited resolution of current cryo-ET combined with STA analyses has not resolved the interactions of ApN and/or CM with the shell proteins in β-carboxysomes. Additionally, the precise mechanism by which the EP of ApN interacts with the shell proteins remains unclear. Two models of EPs interacting with shell proteins have been proposed: (1) EPs bind to the C-terminal α-helices of hexameric shell proteins in a 1:1 ratio, and (2) EPs interact with the epitopes formed at the hexamer-hexamer interfaces (Figure **4.3**) (Aussignargues et al., 2015). Biochemical analysis combined with structural analysis will help to obtain more mechanistic information.

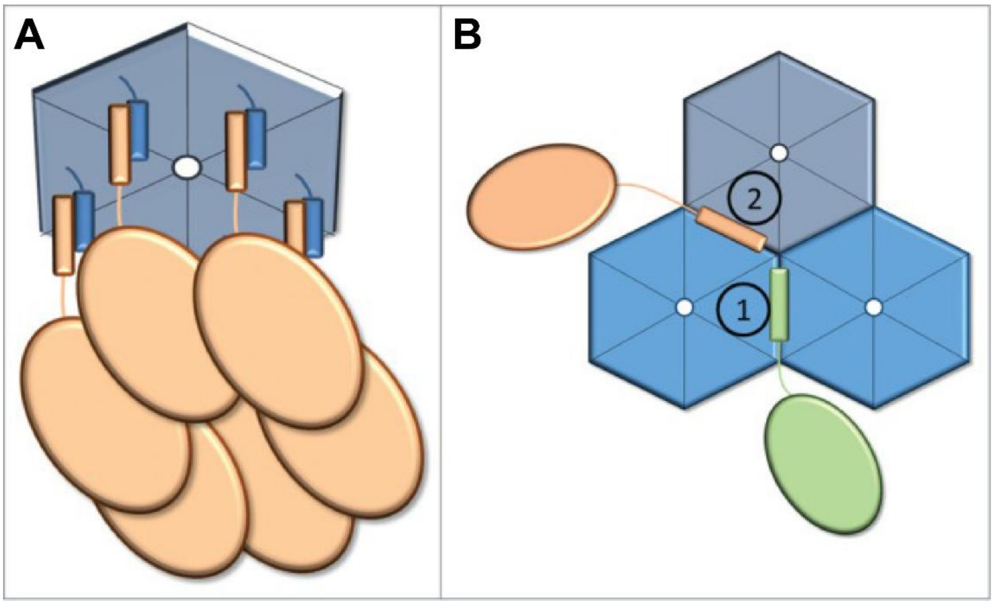


Figure 4.3 Model of EPs interacting with shell proteins.
(A) The proposed 1:1 model (Fan et al., 2012), in which the EPs (orange) interact with the C-terminal α-helices of a hexameric shell protein (blue). **(B)** An alternative model in which the EPs interact with an epitope formed by the hexamer-hexamer interfaces. Distinct epitopes (denoted as 1 and 2) are formed by the interactions of different combinations of shell protein paralogs (denoted by different shades of blue). Figure adapted from (Aussignargues et al., 2015).

4.5. Structural insights into β-carboxysome shell proteins

Shell proteins form the outer proteinaceous layer that encapsulates the internal enzymatic core of the carboxysome, playing critical roles in maintaining structural integrity, mediating selective permeability, and regulating metabolic flux. Structural studies of individual β-carboxysome shell proteins have provided important insights into their architecture and potential functions. The crystal structures of several hexameric CcmK isoforms have been resolved, including CcmK1 of *Syn*6803 (Tanaka et al., 2009), CcmK2, CcmK4 of *Syn*6803 (Kerfeld et al., 2005) and the hetero-hexameric CcmK3/K4 of *Syn*6803 (Garcia-Alles et al., 2019). The pentameric vertex protein CcmL of *Syn*6803 (Tanaka et al., 2008) and the pseudohexameric CcmP of *Se*7942 (Cai et al., 2013; Larsson et al., 2017) have also been structurally characterized. However, the structure of CcmO, a predicted tandem BMC domain protein, remains unknown. Notably, CcmO was not detected in cryo-EM single-particle analysis of reconstituted β-carboxysome shells (Sutter et al., 2019), leaving its structural role and incorporation into the shell uncertain. Furthermore, the ongoing challenges of overexpressing and the purifying certain shell proteins in *E. coli* continue to hinder *in vitro* structural and/or functional studies, limiting our understanding of their contributions to shell assembly and function.

Previously, it was debated whether carboxysome shells are composed of a single layer (Kerfeld et al., 2005; Tsai et al., 2007; Tsai et al., 2009) or a double layer (Samborska & Kimber, 2012) of hexameric BMC proteins. Cryo-EM analysis of recombinant β-carboxysome shells of *Ha*7418, shows a 3-nm-thick single layer, whereas the recombinant shell of *H. ochraceum* has some double-layered protrusions formed by BMC-T pseudohexamers (Sutter et al., 2017). More recently, cryo-ET and STA of native β-carboxysomes from *Se*7942 revealed that the shell is generally ~3 nm thick, with some scattered patches that are ~6 nm thick. Thus, the researchers proposed that the β-carboxysome shell is primarily single-layered, with scattered double-layered protrusions, most likely composed of CcmP pseudohexamers and/or CcmK3/K4 dodecamers (Kong et al., 2024). Despite these advances, many questions remain. The specific functional roles of different shell protein isoforms, their dynamic interactions during assembly, and how shell protein composition adapts to changing environmental conditions are still not fully understood.

4.6. Stoichiometry of carboxysome proteins

While the precise stoichiometry is unclear, the stoichiometry and expression levels of carboxysome proteins may affect their structure and function. Therefore, determining the precise stoichiometry of carboxysome components is essential to unraveling the underlying molecular mechanisms of carboxysome formation, and developing plausible methods to produce functional carboxysomes. This will allow us to modulate functional CO_2-fixing organelles and supercharge photosynthetic carbon fixation in synthetic biology applications, such as heterologously engineering in plants, with the overarching goal of boosting crop yields. However, the exact stoichiometry of all carboxysome building blocks in the functional carboxysome remains unclear. This is due to the complexity of their components, which include hundreds of proteins; the low abundance of certain components; variability in carboxysome size; and their dynamic composition and organization in response to environmental changes, such as CO_2 levels and light intensity (Sun et al., 2019).

Initial estimates of carboxysome protein stoichiometry were performed with either the whole cell lysates or the isolated carboxysome fractions, primarily using immunoblot and mass spectrometry to assess the relative abundance of individual components (Faulkner et al., 2017; Long et al., 2005; Long et al., 2011; Rae et al., 2012). However, these methods have limited specificity due to high sequence homology among carboxysome proteins, a lack of effective purification for intact carboxysomes, and the heterogeneity of carboxysome structures (Long et al., 2005). Furthermore, carboxysome protein stoichiometry based on the total amount of proteins in cell lysates may not directly reflect the composition and stoichiometry of carboxysome proteins in the organelle, given the possible free-standing carboxysome components in the cytosol (Dai et al., 2018). Additionally, the low abundance of proteins or their potential loss during isolation can result in the undetectability of several key components, including CcmO, ApN, CcmP, and RbcX of *Se*7942, by mass spectrometry (Faulkner et al., 2017).

Using a series of *Se*7942 mutants in which individual carboxysome components were C-terminally tagged with bright, fast-maturing enhanced yellow fluorescent protein (YFP), researchers applied real-time single-molecule fluorescence microscopy, along with confocal and electron microscopy, to perform the single-organelle analysis of the protein stoichiometry of eight carboxysome components, including CcmK3, CcmK4, CcmL, CM, ApN, CA, RbcS, and RbcX (Sun et al., 2019). Due to partial segregation, RbcL and CcmK2 could not be fully tagged with YFP; therefore, stoichiometric quantification of these two proteins was excluded

(Sun et al., 2019). Additionally, the system only allowed estimation of the total amounts of $(CM)_3$ and CM^{Ct} combined, without distinguishing between them, due to the C-terminal tagged system (Sun et al., 2019). The florescence tag on CcmP and CcmO interferes with the normal carboxysome assembly; thus, they cannot be analyzed (Sun et al., 2019).

Absolute quantification of protein components within native α- and β-carboxysome was achieved using Quantification conCATamer (QconCAT)-based quantitative high-resolution liquid chromatography-mass spectrometry, respectively (Sun et al., 2022; Sun et al., 2024). The reliability of QconCAT quantification was evident by the strong agreement between individual QconCAT peptides for the same carboxysome proteins (Sun et al., 2022). However, uncertainties remain regarding the exact quantities of several components. In the α-carboxysome, approximately 11 copies of CsoS4A/B pentamers (CsoS4A, 9; Csos4B, 2) were detected, which is slightly fewer than the theoretical estimation of 12 pentamers required to cap all vertices of a canonical icosahedron shell (Sun et al., 2022). Whether the absence of pentamers at some vertices contributes to structural heterogeneity or compromises the integrity of α-carboxysomes is still unknown. CsoS1D pseudohexamers were also found in low abundance, with only ~3 copies per carboxysome (Sun et al., 2022). In β-carboxysomes, only 6 copies of CcmL pentamers were quantified, which is also fewer than the expected theoretical number (Sun et al., 2024). Notably, only 8 copies of ApN were detected per β-carboxysome, which is a surprisingly low number given its proposed role in shell recruitment (Sun et al., 2024). Furthermore, RbcX, which is essential for Rubisco biogenesis and carboxylation activity, was present in only 3 copies per β-carboxysome (Sun et al., 2024).

4.7. The challenge of transplanting a carboxysome-based CCM into plant chloroplasts

In response to the growing global population and the challenges posed by climate change, increasing crop yield has become a critical goal. Many researches focus on improving photosynthetic efficiency, particularly by enhancing the performance of Rubisco in C_3 plants, which are the dominant crop species. Researchers have explored both direct and indirect strategies. Direct approaches include modifying *rbcL* by mutagenesis at or close to the active site (Andersson & Backlund, 2008); engineering Rubisco subunits by replacing plant Rubisco with more efficient variants from other organisms, including C_4 plants and some bacteria and algae (Bierbaumer et al., 2023; Chen, Riaz, et al., 2023; Long et al., 2018; Whitney, Sharwood, et al., 2011); and modifying RbcS to enhance heterologous protein accommodation and expression (Donovan et al., 2020; Khumsupan et al., 2020). Indirect strategies involve modifying Rubisco-interacting proteins by overexpressing molecular chaperones and regulatory proteins, including chaperonins Cpn60 and Cpn20, Raf1, Raf2, RbcX and BSD2 (Aigner et al., 2017; Carmo-Silva et al., 2015; Qu et al., 2021; Sugumar et al., 2024; Whitney et al., 2015) and optimizing the cellular environment in which Rubisco operates by optimizing Rubisco-related metabolic pathways, including mitigating photorespiration (Meacham-Hensold et al., 2024; South et al., 2019). Additionally, there are innovative efforts to increase local CO_2 concentrations around Rubisco to enhance carbon fixation by engineering either pyrenoid- (Atkinson et al., 2020; Tsuji et al., 2017) or carboxysome-based CCMs into crop plants (Chen, Hojka, et al., 2023; Long et al., 2018).

Reconstructing a fully functional CCM in C_3 plants remains a significant technical challenge, primarily due to the intricate coordination required among multiple enzymes and the need for precise organization to achieve synergistic function (Adler et al., 2022; An et al., 2023). Creating a robust CCM in C_3 plants requires assembling the carboxysome structure, ensuring proper protein stoichiometry and expression levels (Liu et al., 2021; Sun et al., 2022; Sun et al., 2019), and providing a compatible biochemical environment to support its function (Figure **4.4**). Specifically, activating the CCM pathway requires introducing bicarbonate transporters into the inner chloroplast membrane to facilitate the active uptake of inorganic carbon and suppressing native stromal CA activity to prevent the premature conversion of HCO_3^- to CO_2 and the diffusion of CO_2 (Chen, Hojka, et al., 2023; Long et al., 2018; Nguyen et al., 2024). These modifications will enable the efficient accumulation of HCO_3^- within the chloroplast stroma, ensuring a high local concentration of inorganic carbon for efficient CO_2 fixation within carboxysomes.

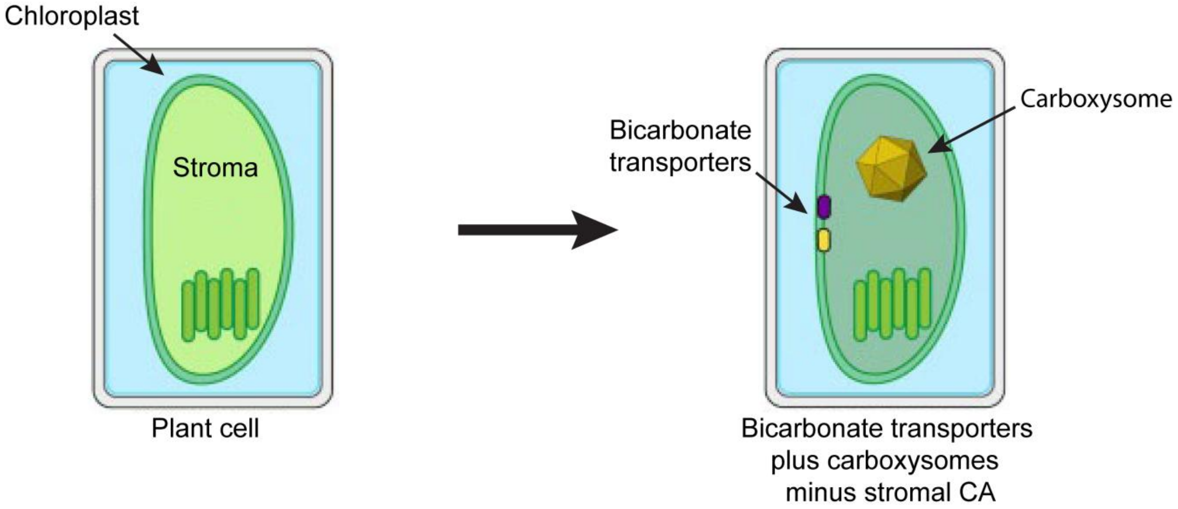


Figure 4.4 Cyanobacterial CCM components for improved photosynthesis.
One proposed strategy to improve C_3 photosynthesis in plant cells **(A)** is to introduce the cyanobacterial CCM components, including independent transfer of carboxysomes containing their native Rubisco into the chloroplast stroma and HCO_3^- transporters to the chloroplast inner-envelope membrane of the chloroplast, in combination with the elimination of stromal CA **(B)**. This approach is expected to promote HCO_3^- in the stroma and elevate CO_2 concentrations near Rubisco, resulting in enhanced CO_2 fixation efficiency. Figure modified from (Long et al., 2018).

Advances in synthetic biology and a deeper understanding of carboxysome architecture and the structural and functional roles of its components have enabled the creation of a chimeric protein that can replace the four native proteins required for carboxysome assembly: $(CM)_3$, CM^{Ct}, $(CA)_4$, and ApN (Gonzalez-Esquer et al., 2015). Remarkably, this chimeric construct can support the formation of smaller, ‘regular-looking’, and functional carboxysomes, albeit the content of this chimeric protein is stoichiometrically fixed, which may hinder its ability to respond flexibly to changes in metabolism (Gonzalez-Esquer et al., 2015). Concurrently, efforts are underway to construct hybrid carboxysomes that combine advantageous features from different carboxysome systems, such as reducing shell complexity and incorporating faster Rubiscos (Nguyen et al., 2023; Sun et al., 2025). For instance, the α-carboxysome of *Cy*7001 comprises a single hexameric shell protein isoform (CsoS1A) and one of the fastest Rubiscos identified in *Synechococcus* WH8102 (Nguyen et al., 2025). While these approaches are theoretically feasible, they require further optimization and validation to ensure that their performance and functionality closely mirror those of native systems.

4.8. Outlook

In our study, we successfully reconstituted the pro-β-carboxysome *in vitro* with Rubisco, CM^{Ct}, $(CM)_3$, $(CA)_4$, and the $(ApN)_3$:CM complex. The peripheral localization of the $(ApN)_3$:CM complex suggests that it is positioned to recruit shell proteins and initiate the formation of a mature β-carboxysome. Further work is needed to overexpress and purify *Se*7942 shell proteins, which would enable their incorporation into the pro-β-carboxysome condensate and the subsequent reconstitution of a mature β-carboxysome *in vitro*. Due to the difficulty of producing certain *Se*7942 shell proteins in *E. coli*, an alternative approach would be to use shell proteins from other organisms, but the compatibility of these proteins with ApN and/or CM must be carefully evaluated. One possible solution is to engineer chimeric proteins that retain essential interaction motifs from native partners by substituting corresponding residues in the target proteins to preserve essential interactions.

Studies have shown that hexameric BMC proteins can self-assemble into sheet- or strip-like higher-level oligomers, as observed in crystal structures (Dryden et al., 2009; Kerfeld et al., 2005; Tanaka et al., 2009). This self-assembly is the basis for the assembly of the facets in polyhedral BMC shells. However, it also implies the necessity of regulatory mechanism or assembly factors to properly incorporate BMC-H, BMC-T and BMC-P proteins into complete, functional shells. For instance, the reconstituted β-carboxysome shell of *Ha*7418 comprising BMC-H (CcmK1 and CcmK2), BMC-T (CcmO) and BMC-P (CcmL), which are recombinantly expressed in *E. coli*. However, CcmO was not detected, indicating that additional internal components, probably ApN and/or CM, may be required for the formation of an intact β-carboxysome shell (Sutter et al., 2019).

Carboxysomes typically range in diameter from 100 to 400 nm, with α-carboxysomes generally around 150 nm and β-carboxysomes between 200 and 400 nm. Due to their small size, carboxysome LLPS *in vivo* is almost impossible to observe using conventional microscopy, but current cytological evidence certainly suggests that carboxysomes are bacterial biomolecular condensates (Azaldegui et al., 2021). *In vivo*, Rubisco and CM nucleate into dense foci that resemble known condensates (Azaldegui et al., 2021), though it is difficult to determine whether these foci are indeed spherical in shape (Cameron et al., 2013; Chen et al., 2013). Interestingly, the adjacent foci interact dynamically, with one depleting as the other increases in size, alongside splitting events that demarcate carboxysome formation, consistent with liquid-like behavior (Azaldegui et al., 2021; Cameron et al., 2013). Note that the size of biomolecular condensates formed in our fluorescence microscopy analysis is much larger, with droplet sizes around 5 µm (Figure **3.22A,B,E**). This discrepancy likely arises from differences in the stoichiometry and concentration of the internal components that influence LLPS behavior *in vitro* (Figure **4.5**). These findings suggest that carboxysome biogenesis *in vivo* must have a mechanism to regulate the size of carboxysomes, which may be critical for substrate and product diffusion in carbon fixation. Understanding the underlying mechanism constitutes an important challenge for future research.

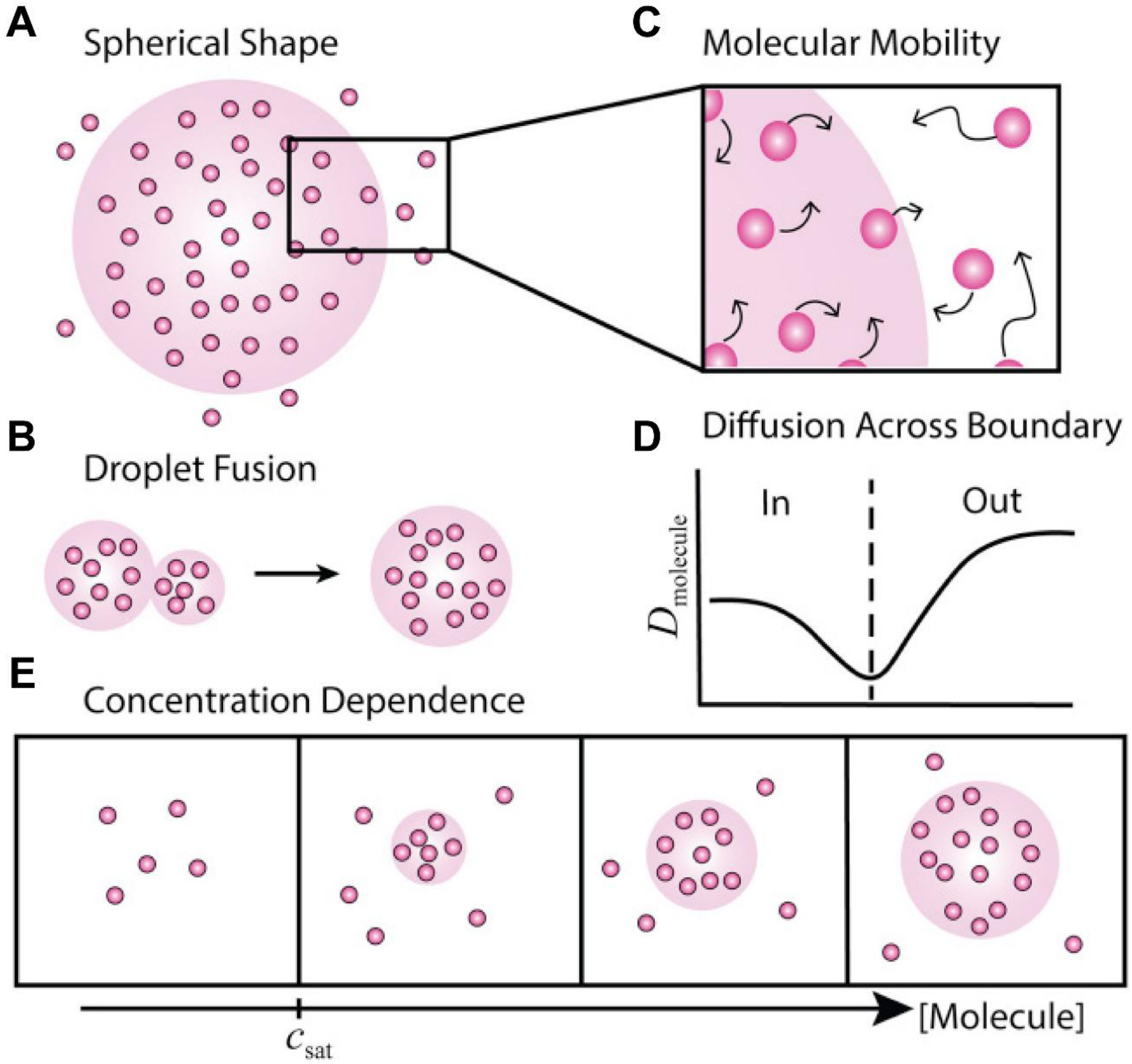


Figure 4.5 Proposed criteria to determine whether a cluster assembles via LLPS.
(A) Condensates are spherical due to surface tension. **(B)** Droplets can fuse upon contact. **(C)** Components within condensates remain mobile and can exchange with the surrounding environment. Arrow length denotes the rate of diffusion of the molecule. **(D)** Restricted diffusion across the membraneless boundary is shown. **(E)** Condensate formation and size scales with the concentration of components; cytoplasm component concentration is buffered. C_{sat}, a saturation threshold. Figure adapted from (Azaldegui et al., 2021).

5. Bibliography

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C. C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Zemgulyte, A., Arvaniti, E.,…Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, *630*(8016), 493–500.

Adler, L., Diaz-Ramos, A., Mao, Y., Pukacz, K. R., Fei, C., & McCormick, A. J. (2022). New horizons for building pyrenoid-based CO2-concentrating mechanisms in plants to improve yields. *Plant Physiol*, *190*(3), 1609–1627.

Aigner, H., Wilson, R. H., Bracher, A., Calisse, L., Bhat, J. Y., Hartl, F. U., & Hayer-Hartl, M. (2017). Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2. *Science*, *358*(6368), 1272–1278.

Alberti, S. (2017). The wisdom of crowds: regulating cell function through condensed states of living matter. *J Cell Sci*, *130*(17), 2789–2796.

Alberti, S., Gladfelter, A., & Mittag, T. (2019). Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell*, *176*(3), 419–434.

Allen, J. (2002). Photosynthesis of ATP-electrons, proton pumps, rotors, and poise. *Cell*, *110*(3), 273–276.

Allen, J. F. (2003). Cyclic, pseudocyclic and noncyclic photophosphorylation: new links in the chain. *Trends Plant Sci*, *8*(1), 15–19.

An, Y., Wang, D., Du, J., Wang, X., & Xiao, J. (2023). Pyrenoid: Organelle with efficient CO(2)-Concentrating mechanism in algae. *J Plant Physiol*, *287*, 154044.

Andersson, I. (2008). Catalysis and regulation in Rubisco. *J Exp Bot*, *59*(7), 1555–1568.

Andersson, I., & Backlund, A. (2008). Structure and function of Rubisco. *Plant Physiol Biochem*, *46*(3), 275–291.

Atkinson, N., Mao, Y., Chan, K. X., & McCormick, A. J. (2020). Condensation of Rubisco into a proto-pyrenoid in higher plant chloroplasts. *Nat Commun*, *11*(1), 6303.

Atkinson, N., Stringer, R., Mitchell, S. R., Seung, D., & McCormick, A. J. (2024). SAGA1 and SAGA2 promote starch formation around proto-pyrenoids in Arabidopsis chloroplasts. *Proc Natl Acad Sci U S A*, *121*(4), e2311013121.

Aussignargues, C., Paasch, B. C., Gonzalez-Esquer, R., Erbilgin, O., & Kerfeld, C. A. (2015). Bacterial microcompartment assembly: The key role of encapsulation peptides. *Commun Integr Biol*, *8*(3), e1039755.

Axen, S. D., Erbilgin, O., & Kerfeld, C. A. (2014). A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput Biol*, *10*(10), e1003898.

Azaldegui, C. A., Vecchiarelli, A. G., & Biteen, J. S. (2021). The emergence of phase separation as an organizing principle in bacteria. *Biophys J*, *120*(7), 1123–1138.

Badger, M. R., & Bek, E. J. (2008). Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle. *J Exp Bot*, *59*(7), 1525–1541.

Badger, M. R., Hanson, D., & Price, G. D. (2002). Evolution and diversity of CO2 concentrating mechanisms in cyanobacteria. *Funct Plant Biol*, *29*(3), 161–173.

Badger, M. R., & Price, G. D. (2003). CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J Exp Bot*, *54*(383), 609–622.

Baker, R. T., Catanzariti, A. M., Karunasekara, Y., Soboleva, T. A., Sharwood, R., Whitney, S., & Board, P. G. (2005). Using deubiquitylating enzymes as research tools. *Methods Enzymol*, *398*, 540–554.

Baker, S. H., Jin, S., Aldrich, H. C., Howard, G. T., & Shively, J. M. (1998). Insertion mutation of the form I cbbL gene encoding ribulose bisphosphate carboxylase/oxygenase (RuBisCO) in Thiobacillus neapolitanus results in expression of form II RuBisCO, loss of carboxysomes, and an increased CO2 requirement for growth. *J Bacteriol*, *180*(16), 4133–4139.

Baker, S. H., Williams, D. S., Aldrich, H. C., Gambrell, A. C., & Shively, J. M. (2000). Identification and localization of the carboxysome peptide Csos3 and its corresponding gene in Thiobacillus neapolitanus. *Arch Microbiol*, *173*(4), 278–283.

Banani, S. F., Lee, H. O., Hyman, A. A., & Rosen, M. K. (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol*, *18*(5), 285–298.

Bassham, J. A., Benson, A. A., & Calvin, M. (1950). The path of carbon in photosynthesis. *J Biol Chem*, *185*(2), 781–787.

Bauwe, H. (2023). Photorespiration - Rubisco's repair crew. *J Plant Physiol*, *280*, 153899.

Benton, D. J., Nans, A., Calder, L. J., Turner, J., Neu, U., Lin, Y. P., Ketelaars, E., Kallewaard, N. L., Corti, D., Lanzavecchia, A., Gamblin, S. J., Rosenthal, P. B., & Skehel, J. J. (2018). Influenza hemagglutinin membrane anchor. *Proc Natl Acad Sci U S A*, *115*(40), 10112–10117.

Bierbaumer, S., Nattermann, M., Schulz, L., Zschoche, R., Erb, T. J., Winkler, C. K., Tinzl, M., & Glueck, S. M. (2023). Enzymatic Conversion of CO(2): From Natural to Artificial Utilization. *Chem Rev*, *123*(9), 5702–5754.

Blanco-Rivero, A., Shutova, T., Roman, M. J., Villarejo, A., & Martinez, F. (2012). Phosphorylation controls the localization and activation of the lumenal carbonic anhydrase in Chlamydomonas reinhardtii. *PLoS One*, *7*(11), e49063.

Blankenship, R. E., & Wiley, I. (2002). *Molecular mechanisms of photosynthesis*. Blackwell Science.

Blikstad, C., Dugan, E. J., Laughlin, T. G., Turnsek, J. B., Liu, M. D., Shoemaker, S. R., Vogiatzi, N., Remis, J. P., & Savage, D. F. (2023). Identification of a carbonic anhydrase-Rubisco complex within the alpha-carboxysome. *Proc Natl Acad Sci U S A*, *120*(43), e2308600120.

Blikstad, C., Flamholz, A. I., Oltrogge, L. M., & Savage, D. F. (2019). Learning to Build a beta-Carboxysome. *Biochemistry*, *58*(16), 2091–2092.

Bobik, T. A., Lehman, B. P., & Yeates, T. O. (2015). Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways. *Mol Microbiol*, *98*(2), 193–207.

Bracher, A., Whitney, S. M., Hartl, F. U., & 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., Hoege, C., Gharakhani, J., Julicher, F., & Hyman, A. A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science*, *324*(5935), 1729–1732.

Bremer, A., Farag, M., Borcherds, W. M., Peran, I., Martin, E. W., Pappu, R. V., & Mittag, T. (2022). Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nat Chem*, *14*(2), 196–207.

Bryant, D. A., & Frigaard, N. U. (2006). Prokaryotic photosynthesis and phototrophy illuminated. *Trends Microbiol*, *14*(11), 488–496.

Busch, F. A., Tominaga, J., Muroya, M., Shirakami, N., Takahashi, S., Yamori, W., Kitaoka, T., Milward, S. E., Nishimura, K., Matsunami, E., Toda, Y., Higuchi, C., Muranaka, A., Takami, T., Watanabe, S., Kinoshita, T., Sakamoto, W., Sakamoto, A., & Shimada, H. (2020). Overexpression of BUNDLE SHEATH DEFECTIVE 2 improves the efficiency of photosynthesis and growth in Arabidopsis. *Plant J*, *102*(1), 129–137.

Cai, F., Bernstein, S. L., Wilson, S. C., & Kerfeld, C. A. (2016). Production and Characterization of Synthetic Carboxysome Shells with Incorporated Luminal Proteins. *Plant Physiol*, *170*(3), 1868–1877.

Cai, F., Dou, Z., Bernstein, S. L., Leverenz, R., Williams, E. B., Heinhorst, S., Shively, J., Cannon, G. C., & Kerfeld, C. A. (2015). Advances in Understanding Carboxysome Assembly in Prochlorococcus and Synechococcus Implicate CsoS2 as a Critical Component. *Life (Basel)*, *5*(2), 1141–1171.

Cai, F., Menon, B. B., Cannon, G. C., Curry, K. J., Shively, J. M., & 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.

Cai, F., Sutter, M., Cameron, J. C., Stanley, D. N., Kinney, J. N., & Kerfeld, C. A. (2013). The structure of CcmP, a tandem bacterial microcompartment domain protein from the beta-carboxysome, forms a subcompartment within a microcompartment. *J Biol Chem*, *288*(22), 16055–16063.

Cameron, J. C., Wilson, S. C., Bernstein, S. L., & Kerfeld, C. A. (2013). Biogenesis of a bacterial organelle: the carboxysome assembly pathway. *Cell*, *155*(5), 1131–1140.

Cannon, G. C., Heinhorst, S., & Kerfeld, C. A. (2010). Carboxysomal carbonic anhydrases: Structure and role in microbial CO2 fixation. *Biochim Biophys Acta*, *1804*(2), 382–392.

Carmo-Silva, E., Scales, J. C., Madgwick, P. J., & Parry, M. A. (2015). Optimizing Rubisco and its regulation for greater resource use efficiency. *Plant Cell Environ*, *38*(9), 1817–1832.

Carpenter, W. B., Lavania, A. A., Borden, J. S., Oltrogge, L. M., Perez, D., Dahlberg, P. D., Savage, D. F., & Moerner, W. E. (2022). Ratiometric Sensing of Redox Environments Inside Individual Carboxysomes Trapped in Solution. *J Phys Chem Lett*, *13*(20), 4455–4462.

Catanzariti, A. M., Soboleva, T. A., Jans, D. A., Board, P. G., & Baker, R. T. (2004). An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Sci*, *13*(5), 1331–1339.

Catherall, E., Musial, S., Atkinson, N., Walker, C. E., Mackinder, L. C. M., & McCormick, A. J. (2025). From algae to plants: understanding pyrenoid-based CO(2)-concentrating mechanisms. *Trends Biochem Sci*, *50*(1), 33–45.

Chaijarasphong, T., Nichols, R. J., Kortright, K. E., Nixon, C. F., Teng, P. K., Oltrogge, L. M., & Savage, D. F. (2016). Programmed Ribosomal Frameshifting Mediates Expression of the alpha-Carboxysome. *J Mol Biol*, *428*(1), 153–164.

Che, X., Wu, J., Liu, H., Su, J., & Chen, X. (2023). Cellular liquid-liquid phase separation: Concept, functions, regulations, and detections. *J Cell Physiol*, *238*(5), 847–865.

Chen, A. H., Robinson-Mosher, A., Savage, D. F., Silver, P. A., & Polka, J. K. (2013). The bacterial carbon-fixing organelle is formed by shell envelopment of preassembled cargo. *PLoS One*, *8*(9), e76127.

Chen, H., Cui, Y., Han, X., Hu, W., Sun, M., Zhang, Y., Wang, P. H., Song, G., Chen, W., & Lou, J. (2020). Liquid-liquid phase separation by SARS-CoV-2 nucleocapsid protein and RNA. *Cell Res*, *30*(12), 1143–1145.

Chen, J., Noble, A. J., Kang, J. Y., & Darst, S. A. (2019). Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and CHAPSO. *J Struct Biol X*, *1*.

Chen, T., Hojka, M., Davey, P., Sun, Y., Dykes, G. F., Zhou, F., Lawson, T., Nixon, P. J., Lin, Y., & Liu, L. N. (2023). Engineering alpha-carboxysomes into plant chloroplasts to support autotrophic photosynthesis. *Nat Commun*, *14*(1), 2118.

Chen, T., Riaz, S., Davey, P., Zhao, Z., Sun, Y., Dykes, G. F., Zhou, F., Hartwell, J., Lawson, T., Nixon, P. J., Lin, Y., & Liu, L. N. (2023). Producing fast and active Rubisco in tobacco to enhance photosynthesis. *Plant Cell*, *35*(2), 795–807.

Chowdhury, C., Sinha, S., Chun, S., Yeates, T. O., & Bobik, T. A. (2014). Diverse bacterial microcompartment organelles. *Microbiol Mol Biol Rev*, *78*(3), 438–468.

Correa, S. S., Schultz, J., Zahodnik-Huntington, B., Naschberger, A., & Rosado, A. S. (2025). Carboxysomes: The next frontier in biotechnology and sustainable solutions. *Biotechnol Adv*, *79*, 108511.

Cot, S. S., So, A. K., & Espie, G. S. (2008). A multiprotein bicarbonate dehydration complex essential to carboxysome function in cyanobacteria. *J Bacteriol*, *190*(3), 936–945.

Croce, R., & van Amerongen, H. (2014). Natural strategies for photosynthetic light harvesting. *Nat Chem Biol*, *10*(7), 492–501.

Cui, C., Shang, M., Li, Z., & Xiao, J. (2025). Synthetic biology approaches to improve Rubisco carboxylation efficiency in C(3) Plants: Direct and Indirect Strategies. *J Plant Physiol*, *307*, 154470.

Dai, W., Chen, M., Myers, C., Ludtke, S. J., Pettitt, B. M., King, J. A., Schmid, M. F., & Chiu, W. (2018). Visualizing Individual RuBisCO and Its Assembly into Carboxysomes in Marine Cyanobacteria by Cryo-Electron Tomography. *J Mol Biol*, *430*(21), 4156–4167.

de Araujo, C., Arefeen, D., Tadesse, Y., Long, B. M., Price, G. D., Rowlett, R. S., Kimber, M. S., & Espie, G. S. (2014). Identification and characterization of a carboxysomal gamma-carbonic anhydrase from the cyanobacterium Nostoc sp. PCC 7120. *Photosynth Res*, *121*(2-3), 135–150.

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., & Savage, D. F. (2019). DABs are inorganic carbon pumps found throughout prokaryotic phyla. *Nat Microbiol*, *4*(12), 2204–2215.

Deviri, D., & Safran, S. A. (2021). Physical theory of biological noise buffering by multicomponent phase separation. *Proc Natl Acad Sci U S A*, *118*(25).

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

Donovan, S., Mao, Y., Orr, D. J., Carmo-Silva, E., & McCormick, A. J. (2020). CRISPR-Cas9-Mediated Mutagenesis of the Rubisco Small Subunit Family in Nicotiana tabacum. *Front Genome Ed*, *2*, 605614.

Doron, L., & Kerfeld, C. A. (2024). Bacterial microcompartments as a next-generation metabolic engineering tool: utilizing nature's solution for confining challenging catabolic pathways. *Biochem Soc Trans*, *52*(3), 997–1010.

Doron, L., Segal, N., Gibori, H., & Shapira, M. (2014). The BSD2 ortholog in Chlamydomonas reinhardtii is a polysome-associated chaperone that co-migrates on sucrose gradients with the rbcL transcript encoding the Rubisco large subunit. *Plant J*, *80*(2), 345–355.

Dou, Z., Heinhorst, S., Williams, E. B., Murin, C. D., Shively, J. M., & Cannon, G. C. (2008). CO2 fixation kinetics of Halothiobacillus neapolitanus mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO2. *J Biol Chem*, *283*(16), 10377–10384.

Dryden, K. A., Crowley, C. S., Tanaka, S., Yeates, T. O., & Yeager, M. (2009). Two-dimensional crystals of carboxysome shell proteins recapitulate the hexagonal packing of three-dimensional crystals. *Protein Sci*, *18*(12), 2629–2635.

Duysens, L. N., Amesz, J., & Kamp, B. M. (1961). Two photochemical systems in photosynthesis. *Nature*, *190*, 510–511.

Dzuricky, M., Rogers, B. A., Shahid, A., Cremer, P. S., & Chilkoti, A. (2020). De novo engineering of intracellular condensates using artificial disordered proteins. *Nat Chem*, *12*(9), 814–825.

Eisenberg, D., Schwarz, E., Komaromy, M., & Wall, R. (1984). Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol*, *179*(1), 125–142.

Ellis, R. J. (2010). Biochemistry: Tackling unintelligent design. *Nature*, *463*(7278), 164–165.

Emlyn-Jones, D., Woodger, F. J., Price, G. D., & Whitney, S. M. (2006). RbcX can function as a rubisco chaperonin, but is non-essential in Synechococcus PCC7942. *Plant Cell Physiol*, *47*(12), 1630–1640.

Emrich-Mills, T. Z., Proctor, M. S., Degen, G. E., Jackson, P. J., Richardson, K. H., Hawkings, F. R., Buchert, F., Hitchcock, A., Hunter, C. N., Mackinder, L. C. M., Hippler, M., & Johnson, M. P. (2025). Tethering ferredoxin-NADP+ reductase to photosystem I promotes photosynthetic cyclic electron transfer. *Plant Cell*, *37*(3).

Engel, B. D., Schaffer, M., Kuhn Cuellar, L., Villa, E., Plitzko, J. M., & Baumeister, W. (2015). Native architecture of the Chlamydomonas chloroplast revealed by in situ cryo-electron tomography. *Elife*, *4*.

English, R. S., Jin, S., & Shively, J. M. (1995). Use of Electroporation To Generate a Thiobacillus neapolitanus Carboxysome Mutant. *Appl Environ Microbiol*, *61*(9), 3256–3260.

Erbilgin, O., McDonald, K. L., & Kerfeld, C. A. (2014). Characterization of a planctomycetal organelle: a novel bacterial microcompartment for the aerobic degradation of plant saccharides. *Appl Environ Microbiol*, *80*(7), 2193–2205.

Espie, G. S., & Kimber, M. S. (2011). Carboxysomes: cyanobacterial RubisCO comes in small packages. *Photosynth Res*, *109*(1-3), 7–20.

Evans, S. L., Al-Hazeem, M. M. J., Mann, D., Smetacek, N., Beavil, A. J., Sun, Y., Chen, T., Dykes, G. F., Liu, L. N., & Bergeron, J. R. C. (2023). Single-particle cryo-EM analysis of the shell architecture and internal organization of an intact alpha-carboxysome. *Structure*, *31*(6), 677–688 e674.

Falcioni, R., Antunes, W. C., Dematte, J. A. M., & Nanni, M. R. (2023). A Novel Method for Estimating Chlorophyll and Carotenoid Concentrations in Leaves: A Two Hyperspectral Sensor Approach. *Sensors (Basel)*, *23*(8).

Fan, C., Cheng, S., Liu, Y., Escobar, C. M., Crowley, C. S., Jefferson, R. E., Yeates, T. O., & Bobik, T. A. (2010). Short N-terminal sequences package proteins into bacterial microcompartments. *Proc Natl Acad Sci U S A*, *107*(16), 7509–7514.

Fan, C., Cheng, S., Sinha, S., & Bobik, T. A. (2012). Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *Proc Natl Acad Sci U S A*, *109*(37), 14995–15000.

Fan, H., & Sun, F. (2022). Developing Graphene Grids for Cryoelectron Microscopy. *Front Mol Biosci*, *9*, 937253.

Faulkner, M., Rodriguez-Ramos, J., Dykes, G. F., Owen, S. V., Casella, S., Simpson, D. M., Beynon, R. J., & Liu, L. N. (2017). Direct characterization of the native structure and mechanics of cyanobacterial carboxysomes. *Nanoscale*, *9*(30), 10662–10673.

Fei, C., Wilson, A. T., Mangan, N. M., Wingreen, N. S., & Jonikas, M. C. (2022). Modelling the pyrenoid-based CO(2)-concentrating mechanism provides insights into its operating principles and a roadmap for its engineering into crops. *Nat Plants*, *8*(5), 583–595.

Feric, M., Vaidya, N., Harmon, T. S., Mitrea, D. M., Zhu, L., Richardson, T. M., Kriwacki, R. W., Pappu, R. V., & Brangwynne, C. P. (2016). Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell*, *165*(7), 1686–1697.

Fisher, R. S., & Elbaum-Garfinkle, S. (2020). Tunable multiphase dynamics of arginine and lysine liquid condensates. *Nat Commun*, *11*(1), 4628.

Flecken, M., Wang, H., Popilka, L., Hartl, F. U., 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.

Freeman Rosenzweig, E. S., Xu, B., Kuhn Cuellar, L., Martinez-Sanchez, A., Schaffer, M., Strauss, M., Cartwright, H. N., Ronceray, P., Plitzko, J. M., Forster, F., Wingreen, N. S., Engel, B. D., Mackinder, L. C. M., & Jonikas, M. C. (2017). The Eukaryotic CO(2)-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization. *Cell*, *171*(1), 148–162 e119.

Friedman, J. R., & Nunnari, J. (2014). Mitochondrial form and function. *Nature*, *505*(7483), 335–343.

Gan, F., Zhang, S., Rockwell, N. C., Martin, S. S., Lagarias, J. C., & Bryant, D. A. (2014). Extensive remodeling of a cyanobacterial photosynthetic apparatus in far-red light. *Science*, *345*(6202), 1312–1317.

Garcia-Alles, L. F., Root, K., Maveyraud, L., Aubry, N., Lesniewska, E., Mourey, L., Zenobi, R., & Truan, G. (2019). Occurrence and stability of hetero-hexamer associations formed by beta-carboxysome CcmK shell components. *PLoS One*, *14*(10), e0223877.

Garrido Ruiz, D., Sandoval-Perez, A., Rangarajan, A. V., Gunderson, E. L., & Jacobson, M. P. (2022). Cysteine Oxidation in Proteins: Structure, Biophysics, and Simulation. *Biochemistry*, *61*(20), 2165–2176.

Ginell, G. M., & Holehouse, A. S. (2023). An Introduction to the Stickers-and-Spacers Framework as Applied to Biomolecular Condensates. *Methods Mol Biol*, *2563*, 95–116.

Gionfriddo, M., Zang, K., & Hayer-Hartl, M. (2023). The challenge of engineering Rubisco for improving photosynthesis. *FEBS Lett*, *597*(13), 1679–1680.

Gonzalez-Esquer, C. R., Shubitowski, T. B., & Kerfeld, C. A. (2015). Streamlined Construction of the Cyanobacterial CO2-Fixing Organelle via Protein Domain Fusions for Use in Plant Synthetic Biology. *Plant Cell*, *27*(9), 2637–2644.

Greber, B. J., Sutter, M., & Kerfeld, C. A. (2019). The Plasticity of Molecular Interactions Governs Bacterial Microcompartment Shell Assembly. *Structure*, *27*(5), 749–763 e744.

Guillen-Boixet, J., Kopach, A., Holehouse, A. S., Wittmann, S., Jahnel, M., Schlussler, R., Kim, K., Trussina, I., Wang, J., Mateju, D., Poser, I., Maharana, S., Ruer-Gruss, M., Richter, D., Zhang, X., Chang, Y. T., Guck, J., Honigmann, A., Mahamid, J.,…Franzmann, T. M. (2020). RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell*, *181*(2), 346–361 e317.

Hanson, T. E., & Tabita, F. R. (2001). A ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO)-like protein from Chlorobium tepidum that is involved with sulfur metabolism and the response to oxidative stress. *Proc Natl Acad Sci U S A*, *98*(8), 4397–4402.

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

Hartman, F. C., & Harpel, M. R. (1994). Structure, function, regulation, and assembly of D-ribulose-1,5-bisphosphate carboxylase/oxygenase. *Annu Rev Biochem*, *63*, 197–234.

Hauser, T., Bhat, J. Y., Milicic, G., Wendler, P., Hartl, F. U., Bracher, A., & Hayer-Hartl, M. (2015). Structure and mechanism of the Rubisco-assembly chaperone Raf1. *Nat Struct Mol Biol*, *22*(9), 720–728.

Hayer-Hartl, M., & Hartl, F. U. (2020). Chaperone Machineries of Rubisco - The Most Abundant Enzyme. *Trends Biochem Sci*, *45*(9), 748–763.

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., He, G., Chen, V. K., Hughson, F. M., McCormick, A. J., Mueller-Cajar, O., Engel, B. D., Yu, Z., & Jonikas, M. C. (2020). The structural basis of Rubisco phase separation in the pyrenoid. *Nat Plants*, *6*(12), 1480–1490.

Heinhorst, S., Williams, E. B., Cai, F., Murin, C. D., Shively, J. M., & Cannon, G. C. (2006). Characterization of the carboxysomal carbonic anhydrase CsoSCA from Halothiobacillus neapolitanus. *J Bacteriol*, *188*(23), 8087–8094.

Hennacy, J. H., Atkinson, N., Kayser-Browne, A., Ergun, S. L., Franklin, E., Wang, L., Eicke, S., Kazachkova, Y., Kafri, M., Fauser, F., Vilarrasa-Blasi, J., Jinkerson, R. E., Zeeman, S. C., McCormick, A. J., & Jonikas, M. C. (2024). SAGA1 and MITH1 produce matrix-traversing membranes in the CO(2)-fixing pyrenoid. *Nat Plants*, *10*(12), 2038–2051.

Hines, K. M., Chaudhari, V., Edgeworth, K. N., Owens, T. G., & Hanson, M. R. (2021). Absence of carbonic anhydrase in chloroplasts affects C(3) plant development but not photosynthesis. *Proc Natl Acad Sci U S A*, *118*(33).

Ho, M. Y., Soulier, N. T., Canniffe, D. P., Shen, G., & Bryant, D. A. (2017). Light regulation of pigment and photosystem biosynthesis in cyanobacteria. *Curr Opin Plant Biol*, *37*, 24–33.

Holehouse, A. S., & Pappu, R. V. (2018). Functional Implications of Intracellular Phase Transitions. *Biochemistry*, *57*(17), 2415–2423.

Hsin, J., Chandler, D. E., Gumbart, J., Harrison, C. B., Sener, M., Strumpfer, J., & Schulten, K. (2010). Self-assembly of photosynthetic membranes. *Chemphyschem*, *11*(6), 1154–1159.

Huang, J., Jiang, Q., Yang, M., Dykes, G. F., Weetman, S. L., Xin, W., He, H. L., & Liu, L. N. (2022). Probing the Internal pH and Permeability of a Carboxysome Shell. *Biomacromolecules*, *23*(10), 4339–4348.

Huang, Q., & Szebenyi, D. M. E. (2023). Crystal structure of a type III Rubisco in complex with its product 3-phosphoglycerate. *Proteins*, *91*(3), 330–337.

Hyman, A. A., Weber, C. A., & Julicher, F. (2014). Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol*, *30*, 39–58.

Iancu, C. V., Ding, H. J., Morris, D. M., Dias, D. P., Gonzales, A. D., Martino, A., & Jensen, G. J. (2007). The structure of isolated Synechococcus strain WH8102 carboxysomes as revealed by electron cryotomography. *J Mol Biol*, *372*(3), 764–773.

Iancu, C. V., Morris, D. M., Dou, Z., Heinhorst, S., Cannon, G. C., & Jensen, G. J. (2010). Organization, structure, and assembly of alpha-carboxysomes determined by electron cryotomography of intact cells. *J Mol Biol*, *396*(1), 105–117.

Iserman, C., Roden, C. A., Boerneke, M. A., Sealfon, R. S. G., McLaughlin, G. A., Jungreis, I., Fritch, E. J., Hou, Y. J., Ekena, J., Weidmann, C. A., Theesfeld, C. L., Kellis, M., Troyanskaya, O. G., Baric, R. S., Sheahan, T. P., Weeks, K. M., & Gladfelter, A. S. (2020). Genomic RNA Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid. *Mol Cell*, *80*(6), 1078–1091 e1076.

Ishikawa-Ankerhold, H. C., Ankerhold, R., & Drummen, G. P. (2012). Advanced fluorescence microscopy techniques--FRAP, FLIP, FLAP, FRET and FLIM. *Molecules*, *17*(4), 4047–4132.

Itakura, A. K., Chan, K. X., Atkinson, N., Pallesen, L., Wang, L., Reeves, G., Patena, W., Caspari, O., Roth, R., Goodenough, U., McCormick, A. J., Griffiths, H., & Jonikas, M. C. (2019). A Rubisco-binding protein is required for normal pyrenoid number and starch sheath morphology in Chlamydomonas reinhardtii. *Proc Natl Acad Sci U S A*, *116*(37), 18445–18454.

Johnson, M. P. (2025). Structure, regulation and assembly of the photosynthetic electron transport chain. *Nat Rev Mol Cell Biol*.

Kaplan, A., Schwarz, R., Lieman-Hurwitz, J., & Reinhold, L. (1991). Physiological and molecular aspects of the inorganic carbon-concentrating mechanism in cyanobacteria. *Plant Physiol*, *97*(3), 851–855.

Kellogg, E. A. (2013). C4 photosynthesis. *Curr Biol*, *23*(14), R594–599.

Kennedy, N. W., Mills, C. E., Nichols, T. M., Abrahamson, C. H., & Tullman-Ercek, D. (2021). Bacterial microcompartments: tiny organelles with big potential. *Curr Opin Microbiol*, *63*, 36–42.

Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F., & Sutter, M. (2018). Bacterial microcompartments. *Nat Rev Microbiol*, *16*(5), 277–290.

Kerfeld, C. A., & Erbilgin, O. (2015). Bacterial microcompartments and the modular construction of microbial metabolism. *Trends Microbiol*, *23*(1), 22–34.

Kerfeld, C. A., Heinhorst, S., & Cannon, G. C. (2010). Bacterial microcompartments. *Annu Rev Microbiol*, *64*, 391–408.

Kerfeld, C. A., & 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., & Yeates, T. O. (2005). Protein structures forming the shell of primitive bacterial organelles. *Science*, *309*(5736), 936–938.

Khumsupan, P., Kozlowska, M. A., Orr, D. J., Andreou, A. I., Nakayama, N., Patron, N., Carmo-Silva, E., & McCormick, A. J. (2020). Generating and characterizing single- and multigene mutants of the Rubisco small subunit family in Arabidopsis. *J Exp Bot*, *71*(19), 5963–5975.

Kimata-Ariga, Y., Chikuma, Y., Saitoh, T., Miyata, M., Yanagihara, Y., Yamane, K., & Hase, T. (2019). NADP(H) allosterically regulates the interaction between ferredoxin and ferredoxin-NADP(+) reductase. *FEBS Open Bio*, *9*(12), 2126–2136.

Kimber, M. S. (2014). Carboxysomal carbonic anhydrases. *Subcell Biochem*, *75*, 89–103.

Kinney, J. N., Axen, S. D., & Kerfeld, C. A. (2011). Comparative analysis of carboxysome shell proteins. *Photosynth Res*, *109*(1-3), 21–32.

Kinney, J. N., Salmeen, A., Cai, F., & Kerfeld, C. A. (2012). Elucidating essential role of conserved carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly. *J Biol Chem*, *287*(21), 17729–17736.

Kirchhoff, H. (2019). Chloroplast ultrastructure in plants. *New Phytol*, *223*(2), 565–574.

Klebl, D. P., Kay, R. W., Sobott, F., Kapur, N., & Muench, S. P. (2022). Towards sub-millisecond cryo-EM grid preparation. *Faraday Discuss*, *240*(0), 33–43.

Klein, M. G., Zwart, P., Bagby, S. C., Cai, F., Chisholm, S. W., Heinhorst, S., Cannon, G. C., & Kerfeld, C. A. (2009). Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. *J Mol Biol*, *392*(2), 319–333.

Kolesinski, P., Belusiak, I., Czarnocki-Cieciura, M., & Szczepaniak, A. (2014). Rubisco Accumulation Factor 1 from Thermosynechococcus elongatus participates in the final stages of ribulose-1,5-bisphosphate carboxylase/oxygenase assembly in Escherichia coli cells and in vitro. *FEBS J*, *281*(17), 3920–3932.

Kong, W. W., Zhu, Y., Zhao, H. R., Du, K., Zhou, R. Q., Li, B., Yang, F., Hou, P., Huang, X. H., Chen, Y., Wang, Y. C., Sun, F., Jiang, Y. L., & Zhou, C. Z. (2024). Cryo-electron tomography reveals the packaging pattern of RuBisCOs in Synechococcus beta-carboxysome. *Structure*, *32*(8), 1110–1120 e1114.

Kono, T., Mehrotra, S., Endo, C., Kizu, N., Matusda, M., Kimura, H., Mizohata, E., Inoue, T., Hasunuma, T., Yokota, A., Matsumura, H., & Ashida, H. (2017). A RuBisCO-mediated carbon metabolic pathway in methanogenic archaea. *Nat Commun*, *8*, 14007.

Krainer, G., Welsh, T. J., Joseph, J. A., Espinosa, J. R., Wittmann, S., de Csillery, E., Sridhar, A., Toprakcioglu, Z., Gudiskyte, G., Czekalska, M. A., Arter, W. E., Guillen-Boixet, J., Franzmann, T. M., Qamar, S., George-Hyslop, P. S., Hyman, A. A., Collepardo-Guevara, R., Alberti, S., & Knowles, T. P. J. (2021). Reentrant liquid condensate phase of proteins is stabilized by hydrophobic and non-ionic interactions. *Nat Commun*, *12*(1), 1085.

Lander, G. C., Estrin, E., Matyskiela, M. E., Bashore, C., Nogales, E., & Martin, A. (2012). Complete subunit architecture of the proteasome regulatory particle. *Nature*, *482*(7384), 186–191.

Larsson, A. M., Hasse, D., Valegard, K., & Andersson, I. (2017). Crystal structures of beta-carboxysome shell protein CcmP: ligand binding correlates with the closed or open central pore. *J Exp Bot*, *68*(14), 3857–3867.

Leister, D. (2023). Enhancing the light reactions of photosynthesis: Strategies, controversies, and perspectives. *Mol Plant*, *16*(1), 4–22.

Li, P., Banjade, S., Cheng, H. C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J. V., King, D. S., Banani, S. F., Russo, P. S., Jiang, Q. X., Nixon, B. T., & Rosen, M. K. (2012). Phase transitions in the assembly of multivalent signalling proteins. *Nature*, *483*(7389), 336–340.

Li, T., Jiang, Q., Huang, J., Aitchison, C. M., Huang, F., Yang, M., Dykes, G. F., He, H. L., Wang, Q., Sprick, R. S., Cooper, A. I., & Liu, L. N. (2020). Reprogramming bacterial protein organelles as a nanoreactor for hydrogen production. *Nat Commun*, *11*(1), 5448.

Liberton, M., Austin, J. R., 2nd, Berg, R. H., & Pakrasi, H. B. (2011). Unique thylakoid membrane architecture of a unicellular N2-fixing cyanobacterium revealed by electron tomography. *Plant Physiol*, *155*(4), 1656–1666.

Lin, M. T., Occhialini, A., Andralojc, P. J., Devonshire, J., Hines, K. M., Parry, M. A., & Hanson, M. R. (2014). beta-Carboxysomal proteins assemble into highly organized structures in Nicotiana chloroplasts. *Plant J*, *79*(1), 1–12.

Lindahl, M., & Florencio, F. J. (2003). Thioredoxin-linked processes in cyanobacteria are as numerous as in chloroplasts, but targets are different. *Proc Natl Acad Sci U S A*, *100*(26), 16107–16112.

Liu, C., Young, A. L., Starling-Windhof, A., Bracher, A., Saschenbrecker, S., Rao, B. V., Rao, K. V., Berninghausen, O., Mielke, T., Hartl, F. U., Beckmann, R., & Hayer-Hartl, M. (2010). Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature*, *463*(7278), 197–202.

Liu, L. N. (2022). Advances in the bacterial organelles for CO(2) fixation. *Trends Microbiol*, *30*(6), 567–580.

Liu, L. N., Yang, M., Sun, Y., & Yang, J. (2021). Protein stoichiometry, structural plasticity and regulation of bacterial microcompartments. *Curr Opin Microbiol*, *63*, 133–141.

Liu, N., & Wang, H. W. (2023). Better Cryo-EM Specimen Preparation: How to Deal with the Air-Water Interface? *J Mol Biol*, *435*(9), 167926.

Long, B. M., Badger, M. R., Whitney, S. M., & Price, G. D. (2007). Analysis of carboxysomes from Synechococcus PCC7942 reveals multiple Rubisco complexes with carboxysomal proteins CcmM and CcaA. *J Biol Chem*, *282*(40), 29323–29335.

Long, B. M., Forster, B., Pulsford, S. B., Price, G. D., & Badger, M. R. (2021). Rubisco proton production can drive the elevation of CO(2) within condensates and carboxysomes. *Proc Natl Acad Sci U S A*, *118*(18).

Long, B. M., Hee, W. Y., Sharwood, R. E., Rae, B. D., Kaines, S., Lim, Y. L., Nguyen, N. D., Massey, B., Bala, S., von Caemmerer, S., Badger, M. R., & Price, G. D. (2018). Carboxysome encapsulation of the CO(2)-fixing enzyme Rubisco in tobacco chloroplasts. *Nat Commun*, *9*(1), 3570.

Long, B. M., Price, G. D., & Badger, M. R. (2005). Proteomic assessment of an established technique for carboxysome enrichment from PCC7942. *Canadian Journal of Botany-Revue Canadienne De Botanique*, *83*(7), 746–757.

Long, B. M., Rae, B. D., Badger, M. R., & Price, G. D. (2011). Over-expression of the beta-carboxysomal CcmM protein in Synechococcus PCC7942 reveals a tight co-regulation of carboxysomal carbonic anhydrase (CcaA) and M58 content. *Photosynth Res*, *109*(1-3), 33–45.

Long, B. M., Tucker, L., Badger, M. R., & 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*(1), 285–293.

Luzio, J. P., Pryor, P. R., & Bright, N. A. (2007). Lysosomes: fusion and function. *Nat Rev Mol Cell Biol*, *8*(8), 622–632.

Maeda, S., Kawaguchi, Y., Ohe, T. A., & Omata, T. (1998). cis-acting sequences required for NtcB-dependent, nitrite-responsive positive regulation of the nitrate assimilation operon in the cyanobacterium Synechococcus sp. strain PCC 7942. *J Bacteriol*, *180*(16), 4080–4088.

Malone, L. A., Proctor, M. S., Hitchcock, A., Hunter, C. N., & Johnson, M. P. (2021). Cytochrome b(6)f - Orchestrator of photosynthetic electron transfer. *Biochim Biophys Acta Bioenerg*, *1862*(5), 148380.

Marco, E., Martinez, I., Ronen-Tarazi, M., Orus, M. I., & Kaplan, A. (1994). Inactivation of ccmO in Synechococcus sp. Strain PCC 7942 Results in a Mutant Requiring High Levels of CO(2). *Appl Environ Microbiol*, *60*(3), 1018–1020.

Mastronarde, D. N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol*, *152*(1), 36–51.

Mata-Cabana, A., Florencio, F. J., & Lindahl, M. (2007). Membrane proteins from the cyanobacterium Synechocystis sp. PCC 6803 interacting with thioredoxin. *Proteomics*, *7*(21), 3953–3963.

McGrath, J. M., & Long, S. P. (2014). Can the cyanobacterial carbon-concentrating mechanism increase photosynthesis in crop species? A theoretical analysis. *Plant Physiol*, *164*(4), 2247–2261.

McGurn, L. D., Moazami-Goudarzi, M., White, S. A., Suwal, T., Brar, B., Tang, J. Q., Espie, G. S., & Kimber, M. S. (2016). The structure, kinetics and interactions of the beta-carboxysomal beta-carbonic anhydrase, CcaA. *Biochem J*, *473*(24), 4559–4572.

Meacham-Hensold, K., Cavanagh, A. P., Sorensen, P., South, P. F., Fowler, J., Boyd, R., Jeong, J., Burgess, S., Stutz, S., Dilger, R. N., Lee, M., Ferrari, N., Larkin, J., & Ort, D. R. (2024). Shortcutting Photorespiration Protects Potato Photosynthesis and Tuber Yield Against Heatwave Stress. *Glob Chang Biol*, *30*(12), e17595.

Menon, B. B., Heinhorst, S., Shively, J. M., & Cannon, G. C. (2010). The carboxysome shell is permeable to protons. *J Bacteriol*, *192*(22), 5881–5886.

Metskas, L. A., Ortega, D., Oltrogge, L. M., Blikstad, C., Lovejoy, D. R., Laughlin, T. G., Savage, D. F., & Jensen, G. J. (2022). Rubisco forms a lattice inside alpha-carboxysomes. *Nat Commun*, *13*(1), 4863.

Mitchell, P. (1961). Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature*, *191*, 144–148.

Mitchell, P. (1972). Chemiosmotic coupling in energy transduction: a logical development of biochemical knowledge. *J Bioenerg*, *3*(1), 5–24.

Miziorko, H. M., & Lorimer, G. H. (1983). Ribulose-1,5-bisphosphate carboxylase-oxygenase. *Annu Rev Biochem*, *52*, 507–535.

Naponelli, V., Noiriel, A., Ziemak, M. J., Beverley, S. M., Lye, L. F., Plume, A. M., Botella, J. R., Loizeau, K., Ravanel, S., Rebeille, F., de Crecy-Lagard, V., & Hanson, A. D. (2008). Phylogenomic and functional analysis of pterin-4a-carbinolamine dehydratase family (COG2154) proteins in plants and microorganisms. *Plant Physiol*, *146*(4), 1515–1527.

Nelson, N., & Ben-Shem, A. (2004). The complex architecture of oxygenic photosynthesis. *Nat Rev Mol Cell Biol*, *5*(12), 971–982.

Newman, J., Branden, C. I., & Jones, T. A. (1993). Structure determination and refinement of ribulose 1,5-bisphosphate carboxylase/oxygenase from Synechococcus PCC6301. *Acta Crystallogr D Biol Crystallogr*, *49*(Pt 6), 548–560.

Nguyen, N. D., Pulsford, S. B., Forster, B., Rottet, S., Rourke, L., Long, B. M., & Price, G. D. (2024). A carboxysome-based CO(2) concentrating mechanism for C(3) crop chloroplasts: advances and the road ahead. *Plant J*, *118*(4), 940–952.

Nguyen, N. D., Pulsford, S. B., Hee, W. Y., Rae, B. D., Rourke, L. M., Price, G. D., & Long, B. M. (2023). Towards engineering a hybrid carboxysome. *Photosynth Res*, *156*(2), 265–277.

Nguyen, N. D., Rourke, L. M., Cleaver, A., Brock, J., Long, B. M., & Price, D. G. (2025). Understanding carboxysomes to enhance carbon fixation in crops. *Biochem Soc Trans*, *53*(3), 671–685.

Ni, T., Jiang, Q., Ng, P. C., Shen, J., Dou, H., Zhu, Y., Radecke, J., Dykes, G. F., Huang, F., Liu, L. N., & Zhang, P. (2023). Intrinsically disordered CsoS2 acts as a general molecular thread for alpha-carboxysome shell assembly. *Nat Commun*, *14*(1), 5512.

Ni, T., Sun, Y., Burn, W., Al-Hazeem, M. M. J., Zhu, Y., Yu, X., Liu, L. N., & Zhang, P. (2022). Structure and assembly of cargo Rubisco in two native alpha-carboxysomes. *Nat Commun*, *13*(1), 4299.

Niederhuber, M. J., Lambert, T. J., Yapp, C., Silver, P. A., & Polka, J. K. (2017). Superresolution microscopy of the beta-carboxysome reveals a homogeneous matrix. *Mol Biol Cell*, *28*(20), 2734–2745.

Oberg, N., Zallot, R., & Gerlt, J. A. (2023). EFI-EST, EFI-GNT, and EFI-CGFP: Enzyme Function Initiative (EFI) Web Resource for Genomic Enzymology Tools. *J Mol Biol*, *435*(14), 168018.

Occhialini, A., Lin, M. T., Andralojc, P. J., Hanson, M. R., & Parry, M. A. (2016). Transgenic tobacco plants with improved cyanobacterial Rubisco expression but no extra assembly factors grow at near wild-type rates if provided with elevated CO2. *Plant J*, *85*(1), 148–160.

Ochoa, J. M., Nguyen, V. N., Nie, M., Sawaya, M. R., Bobik, T. A., & Yeates, T. O. (2020). Symmetry breaking and structural polymorphism in a bacterial microcompartment shell protein for choline utilization. *Protein Sci*, *29*(11), 2201–2212.

Oltrogge, L. M., Chaijarasphong, T., Chen, A. W., Bolin, E. R., Marqusee, S., & Savage, D. F. (2020). Multivalent interactions between CsoS2 and Rubisco mediate alpha-carboxysome formation. *Nat Struct Mol Biol*, *27*(3), 281–287.

Pak, C. W., Kosno, M., Holehouse, A. S., Padrick, S. B., Mittal, A., Ali, R., Yunus, A. A., Liu, D. R., Pappu, R. V., & Rosen, M. K. (2016). Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol Cell*, *63*(1), 72–85.

Parry, M. A., Andralojc, P. J., Scales, J. C., Salvucci, M. E., Carmo-Silva, A. E., Alonso, H., & Whitney, S. M. (2013). Rubisco activity and regulation as targets for crop improvement. *J Exp Bot*, *64*(3), 717–730.

Parry, M. A., Keys, A. J., Madgwick, P. J., Carmo-Silva, A. E., & Andralojc, P. J. (2008). Rubisco regulation: a role for inhibitors. *J Exp Bot*, *59*(7), 1569–1580.

Pearce, F. G. (2006). Catalytic by-product formation and ligand binding by ribulose bisphosphate carboxylases from different phylogenies. *Biochem J*, *399*(3), 525–534.

Pena, K. L., Castel, S. E., de Araujo, C., Espie, G. S., & Kimber, M. S. (2010). Structural basis of the oxidative activation of the carboxysomal gamma-carbonic anhydrase, CcmM. *Proc Natl Acad Sci U S A*, *107*(6), 2455–2460.

Peterhansel, C., Horst, I., Niessen, M., Blume, C., Kebeish, R., Kurkcuoglu, S., & Kreuzaler, F. (2010). Photorespiration. *Arabidopsis Book*, *8*, e0130.

Portis, A. R., Jr. (2003). Rubisco activase - Rubisco's catalytic chaperone. *Photosynth Res*, *75*(1), 11–27.

Portis, A. R., Jr., & Parry, M. A. (2007). Discoveries in Rubisco (Ribulose 1,5-bisphosphate carboxylase/oxygenase): a historical perspective. *Photosynth Res*, *94*(1), 121–143.

Price, G. D., & Badger, M. R. (1989a). Ethoxyzolamide Inhibition of CO(2) Uptake in the Cyanobacterium Synechococcus PCC7942 without Apparent Inhibition of Internal Carbonic Anhydrase Activity. *Plant Physiol*, *89*(1), 37–43.

Price, G. D., & Badger, M. R. (1989b). Expression of Human Carbonic Anhydrase in the Cyanobacterium Synechococcus PCC7942 Creates a High CO(2)-Requiring Phenotype : Evidence for a Central Role for Carboxysomes in the CO(2) Concentrating Mechanism. *Plant Physiol*, *91*(2), 505–513.

Price, G. D., & Badger, M. R. (1989c). Isolation and Characterization of High CO(2)-Requiring-Mutants of the Cyanobacterium Synechococcus PCC7942 : Two Phenotypes that Accumulate Inorganic Carbon but Are Apparently Unable to Generate CO(2) within the Carboxysome. *Plant Physiol*, *91*(2), 514–525.

Price, G. D., Badger, M. R., & von Caemmerer, S. (2011). The prospect of using cyanobacterial bicarbonate transporters to improve leaf photosynthesis in C3 crop plants. *Plant Physiol*, *155*(1), 20–26.

Price, G. D., Coleman, J. R., & Badger, M. R. (1992). Association of Carbonic Anhydrase Activity with Carboxysomes Isolated from the Cyanobacterium Synechococcus PCC7942. *Plant Physiol*, *100*(2), 784–793.

Price, G. D., Howitt, S. M., Harrison, K., & Badger, M. R. (1993). Analysis of a genomic DNA region from the cyanobacterium Synechococcus sp. strain PCC7942 involved in carboxysome assembly and function. *J Bacteriol*, *175*(10), 2871–2879.

Price, G. D., Shelden, M. C., & Howitt, S. M. (2011). Membrane topology of the cyanobacterial bicarbonate transporter, SbtA, and identification of potential regulatory loops. *Mol Membr Biol*, *28*(5), 265–275.

Pulsford, S. B., Outram, M. A., Forster, B., Rhodes, T., Williams, S. J., Badger, M. R., Price, G. D., Jackson, C. J., & Long, B. M. (2024). Cyanobacterial alpha-carboxysome carbonic anhydrase is allosterically regulated by the Rubisco substrate RuBP. *Sci Adv*, *10*(19), eadk7283.

Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods*, *14*(3), 290–296.

Qu, Y., Sakoda, K., Fukayama, H., Kondo, E., Suzuki, Y., Makino, A., Terashima, I., & Yamori, W. (2021). Overexpression of both Rubisco and Rubisco activase rescues rice photosynthesis and biomass under heat stress. *Plant Cell Environ*, *44*(7), 2308–2320.

Rae, B. D., Long, B. M., Badger, M. R., & Price, G. D. (2012). Structural determinants of the outer shell of beta-carboxysomes in Synechococcus elongatus PCC 7942: roles for CcmK2, K3-K4, CcmO, and CcmL. *PLoS One*, *7*(8), e43871.

Rae, B. D., Long, B. M., Badger, M. R., & 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*(3), 357–379.

Rae, B. D., Long, B. M., Forster, B., Nguyen, N. D., Velanis, C. N., Atkinson, N., Hee, W. Y., Mukherjee, B., Price, G. D., & McCormick, A. J. (2017). Progress and challenges of engineering a biophysical CO2-concentrating mechanism into higher plants. *J Exp Bot*, *68*(14), 3717–3737.

Raines, C. A. (2003). The Calvin cycle revisited. *Photosynth Res*, *75*(1), 1–10.

Raines, C. A. (2011). Increasing photosynthetic carbon assimilation in C3 plants to improve crop yield: current and future strategies. *Plant Physiol*, *155*(1), 36–42.

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

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*, *42*(Web Server issue), W320–324.

Roberts, E. W., Cai, F., Kerfeld, C. A., Cannon, G. C., & Heinhorst, S. (2012). Isolation and characterization of the Prochlorococcus carboxysome reveal the presence of the novel shell protein CsoS1D. *J Bacteriol*, *194*(4), 787–795.

Ronen-Tarazi, M., Lieman-Hurwitz, J., Gabay, C., Orus, M. I., & Kaplan, A. (1995). The genomic region of rbcLS in Synechococcus sp. PCC 7942 contains genes involved in the ability to grow under low CO2 concentration and in chlorophyll biosynthesis. *Plant Physiol*, *108*(4), 1461–1469.

Saha, S., Weber, C. A., Nousch, M., Adame-Arana, O., Hoege, C., Hein, M. Y., Osborne-Nishimura, E., Mahamid, J., Jahnel, M., Jawerth, L., Pozniakovski, A., Eckmann, C. R., Julicher, F., & Hyman, A. A. (2016). Polar Positioning of Phase-Separated Liquid Compartments in Cells Regulated by an mRNA Competition Mechanism. *Cell*, *166*(6), 1572–1584 e1516.

Salesse, C., Sharwood, R., Sakamoto, W., & Stern, D. (2017). The Rubisco Chaperone BSD2 May Regulate Chloroplast Coverage in Maize Bundle Sheath Cells. *Plant Physiol*, *175*(4), 1624–1633.

Samborska, B., & Kimber, M. S. (2012). A dodecameric CcmK2 structure suggests beta-carboxysomal shell facets have a double-layered organization. *Structure*, *20*(8), 1353–1362.

Sanders, D. W., Kedersha, N., Lee, D. S. W., Strom, A. R., Drake, V., Riback, J. A., Bracha, D., Eeftens, J. M., Iwanicki, A., Wang, A., Wei, M. T., Whitney, G., Lyons, S. M., Anderson, P., Jacobs, W. M., Ivanov, P., & Brangwynne, C. P. (2020). Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. *Cell*, *181*(2), 306–324 e328.

Sarewicz, M., Pintscher, S., Pietras, R., Borek, A., Bujnowicz, L., Hanke, G., Cramer, W. A., Finazzi, G., & Osyczka, A. (2021). Catalytic Reactions and Energy Conservation in the Cytochrome bc(1) and b(6)f Complexes of Energy-Transducing Membranes. *Chem Rev*, *121*(4), 2020–2108.

Saschenbrecker, S., Bracher, A., Rao, K. V., Rao, B. V., Hartl, F. U., & Hayer-Hartl, M. (2007). Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell*, *129*(6), 1189–1200.

Sato, T., Atomi, H., & Imanaka, T. (2007). Archaeal type III RuBisCOs function in a pathway for AMP metabolism. *Science*, *315*(5814), 1003–1006.

Sawaya, M. R., Cannon, G. C., Heinhorst, S., Tanaka, S., Williams, E. B., Yeates, T. O., & Kerfeld, C. A. (2006). The structure of beta-carbonic anhydrase from the carboxysomal shell reveals a distinct subclass with one active site for the price of two. *J Biol Chem*, *281*(11), 7546–7555.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods*, *9*(7), 676–682.

Schmidt-Krey, I., & Rubinstein, J. L. (2011). Electron cryomicroscopy of membrane proteins: specimen preparation for two-dimensional crystals and single particles. *Micron*, *42*(2), 107–116.

Schopf, J. W., & Packer, B. M. (1987). Early Archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona Group, Australia. *Science*, *237*, 70–73.

Schrodinger, LLC. (2015). *The PyMOL Molecular Graphics System, Version 1.8*.

Segura Broncano, L., Pukacz, K. R., Reichel-Deland, V., Schluter, U., Triesch, S., & Weber, A. P. M. (2023). Photorespiration is the solution, not the problem. *J Plant Physiol*, *282*, 153928.

Shieh, Y. W., Minguez, P., Bork, P., Auburger, J. J., Guilbride, D. L., Kramer, G., & Bukau, B. (2015). Operon structure and cotranslational subunit association direct protein assembly in bacteria. *Science*, *350*(6261), 678–680.

Shin, Y., Berry, J., Pannucci, N., Haataja, M. P., Toettcher, J. E., & Brangwynne, C. P. (2017). Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell*, *168*(1-2), 159–171 e114.

Shin, Y., Chang, Y. C., Lee, D. S. W., Berry, J., Sanders, D. W., Ronceray, P., Wingreen, N. S., Haataja, M., & Brangwynne, C. P. (2018). Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. *Cell*, *175*(6), 1481–1491 e1413.

Sinetova, M. A., Kupriyanova, E. V., Markelova, A. G., Allakhverdiev, S. I., & Pronina, N. A. (2012). Identification and functional role of the carbonic anhydrase Cah3 in thylakoid membranes of pyrenoid of Chlamydomonas reinhardtii. *Biochim Biophys Acta*, *1817*(8), 1248–1255.

So, A. K., Espie, G. S., Williams, E. B., Shively, J. M., Heinhorst, S., & Cannon, G. C. (2004). A novel evolutionary lineage of carbonic anhydrase (epsilon class) is a component of the carboxysome shell. *J Bacteriol*, *186*(3), 623–630.

So, A. K., John-McKay, M., & Espie, G. S. (2002). Characterization of a mutant lacking carboxysomal carbonic anhydrase from the cyanobacterium Synechocystis PCC6803. *Planta*, *214*(3), 456–467.

South, P. F., Cavanagh, A. P., Liu, H. W., & Ort, D. R. (2019). Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science*, *363*(6422), eaat9077.

Spreitzer, R. J. (2003). Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Arch Biochem Biophys*, *414*(2), 141–149.

Stotz, M., Mueller-Cajar, O., Ciniawsky, S., Wendler, P., Hartl, F. U., Bracher, A., & Hayer-Hartl, M. (2011). Structure of green-type Rubisco activase from tobacco. *Nat Struct Mol Biol*, *18*(12), 1366–1370.

Sugumar, T., Shen, G., Smith, J., & Zhang, H. (2024). Creating Climate-Resilient Crops by Increasing Drought, Heat, and Salt Tolerance. *Plants (Basel)*, *13*(9).

Sun, H., Cui, N., Han, S. J., Chen, Z. P., Xia, L. Y., Chen, Y., Jiang, Y. L., & Zhou, C. Z. (2021). Complex structure reveals CcmM and CcmN form a heterotrimeric adaptor in beta-carboxysome. *Protein Sci*, *30*(8), 1566–1576.

Sun, Y., Casella, S., Fang, Y., Huang, F., Faulkner, M., Barrett, S., & Liu, L. N. (2016). Light Modulates the Biosynthesis and Organization of Cyanobacterial Carbon Fixation Machinery through Photosynthetic Electron Flow. *Plant Physiol*, *171*(1), 530–541.

Sun, Y., Chen, T., Ge, X., Ni, T., Dykes, G. F., Zhang, P., Huang, F., & Liu, L. N. (2025). Engineering CO(2)-fixing modules in Escherichia coli via efficient assembly of cyanobacterial Rubisco and carboxysomes. *Plant Commun*, *6*(3), 101217.

Sun, Y., Harman, V. M., Johnson, J. R., Brownridge, P. J., Chen, T., Dykes, G. F., Lin, Y., Beynon, R. J., & Liu, L. N. (2022). Decoding the Absolute Stoichiometric Composition and Structural Plasticity of alpha-Carboxysomes. *mBio*, *13*(2), e0362921.

Sun, Y., Sheng, Y., Ni, T., Ge, X., Sarsby, J., Brownridge, P. J., Li, K., Hardenbrook, N., Dykes, G. F., Rockliffe, N., Eyers, C. E., Zhang, P., & Liu, L. N. (2024). Rubisco packaging and stoichiometric composition of the native beta-carboxysome in Synechococcus elongatus PCC7942. *Plant Physiol*, *197*(1).

Sun, Y., Wollman, A. J. M., Huang, F., Leake, M. C., & Liu, L. N. (2019). Single-Organelle Quantification Reveals Stoichiometric and Structural Variability of Carboxysomes Dependent on the Environment. *Plant Cell*, *31*(7), 1648–1664.

Sutter, M., Boehringer, D., Gutmann, S., Gunther, S., Prangishvili, D., Loessner, M. J., Stetter, K. O., Weber-Ban, E., & Ban, N. (2008). Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat Struct Mol Biol*, *15*(9), 939–947.

Sutter, M., Greber, B., Aussignargues, C., & Kerfeld, C. A. (2017). Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science*, *356*(6344), 1293–1297.

Sutter, M., Laughlin, T. G., Sloan, N. B., Serwas, D., Davies, K. M., & Kerfeld, C. A. (2019). Structure of a Synthetic beta-Carboxysome Shell. *Plant Physiol*, *181*(3), 1050–1058.

Sutter, M., Melnicki, M. R., Schulz, F., Woyke, T., & Kerfeld, C. A. (2021). A catalog of the diversity and ubiquity of bacterial microcompartments. *Nat Commun*, *12*(1), 3809.

Suzuki, E., Fukuzawa, H., & Miyachi, S. (1991). Identification of a genomic region that complements a temperature-sensitive, high CO2-requiring mutant of the cyanobacterium, Synechococcus sp. PCC7942. *Mol Gen Genet*, *226*(3), 401–408.

Tabita, F. R. (1999). Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective. *Photosynthesis Research*, *60*(1), 1–28.

Tabita, F. R., Satagopan, S., Hanson, T. E., Kreel, N. E., & 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*(7), 1515–1524.

Tan, Y. Q., Ali, S., Xue, B., Teo, W. Z., Ling, L. H., Go, M. K., Lv, H., Robinson, R. C., Narita, A., & Yew, W. S. (2021). Structure of a Minimal alpha-Carboxysome-Derived Shell and Its Utility in Enzyme Stabilization. *Biomacromolecules*, *22*(10), 4095–4109.

Tan, Y. Z., Baldwin, P. R., Davis, J. H., Williamson, J. R., Potter, C. S., Carragher, B., & Lyumkis, D. (2017). Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods*, *14*(8), 793–796.

Tanaka, S., Kerfeld, C. A., Sawaya, M. R., Cai, F., Heinhorst, S., Cannon, G. C., & Yeates, T. O. (2008). Atomic-level models of the bacterial carboxysome shell. *Science*, *319*(5866), 1083–1086.

Tanaka, S., Sawaya, M. R., Phillips, M., & Yeates, T. O. (2009). Insights from multiple structures of the shell proteins from the beta-carboxysome. *Protein Sci*, *18*(1), 108–120.

Tanaka, S., Sawaya, M. R., & Yeates, T. O. (2010). Structure and mechanisms of a protein-based organelle in Escherichia coli. *Science*, *327*(5961), 81–84.

Topfer, N., Braam, T., Shameer, S., Ratcliffe, R. G., & Sweetlove, L. J. (2020). Alternative Crassulacean Acid Metabolism Modes Provide Environment-Specific Water-Saving Benefits in a Leaf Metabolic Model. *Plant Cell*, *32*(12), 3689–3705.

Toyokawa, C., Yamano, T., & Fukuzawa, H. (2020). Pyrenoid Starch Sheath Is Required for LCIB Localization and the CO(2)-Concentrating Mechanism in Green Algae. *Plant Physiol*, *182*(4), 1883–1893.

Trettel, D. S., Kerfeld, C. A., & Gonzalez-Esquer, C. R. (2024). Dynamic structural determinants in bacterial microcompartment shells. *Curr Opin Microbiol*, *80*, 102497.

Trosch, R., Muhlhaus, T., Schroda, M., & Willmund, F. (2015). ATP-dependent molecular chaperones in plastids--More complex than expected. *Biochim Biophys Acta*, *1847*(9), 872–888.

Tsai, Y., Sawaya, M. R., Cannon, G. C., Cai, F., Williams, E. B., Heinhorst, S., Kerfeld, C. A., & Yeates, T. O. (2007). Structural analysis of CsoS1A and the protein shell of the Halothiobacillus neapolitanus carboxysome. *PLoS Biol*, *5*(6), e144.

Tsai, Y., Sawaya, M. R., & 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*(Pt 9), 980–988.

Tsai, Y. C., Lapina, M. C., Bhushan, S., & Mueller-Cajar, O. (2015). Identification and characterization of multiple rubisco activases in chemoautotrophic bacteria. *Nat Commun*, *6*, 8883.

Tsuji, Y., Nakajima, K., & Matsuda, Y. (2017). Molecular aspects of the biophysical CO2-concentrating mechanism and its regulation in marine diatoms. *J Exp Bot*, *68*(14), 3763–3772.

Turmo, A., Gonzalez-Esquer, C. R., & Kerfeld, C. A. (2017). Carboxysomes: metabolic modules for CO2 fixation. *FEMS Microbiol Lett*, *364*(18).

Vekilov, P. G., & Vorontsova, M. A. (2014). Nucleation precursors in protein crystallization. *Acta Crystallogr F Struct Biol Commun*, *70*(Pt 3), 271–282.

von Caemmerer, S., & Furbank, R. T. (2016). Strategies for improving C4 photosynthesis. *Curr Opin Plant Biol*, *31*, 125–134.

Waheeda, K., Kitchel, H., Wang, Q., & Chiu, P. L. (2023). Molecular mechanism of Rubisco activase: Dynamic assembly and Rubisco remodeling. *Front Mol Biosci*, *10*, 1125922.

Walker, B. J., VanLoocke, A., Bernacchi, C. J., & Ort, D. R. (2016). The Costs of Photorespiration to Food Production Now and in the Future. *Annu Rev Plant Biol*, *67*, 107–129.

Wang, B., Zhang, L., Dai, T., Qin, Z., Lu, H., Zhang, L., & Zhou, F. (2021). Liquid-liquid phase separation in human health and diseases. *Signal Transduct Target Ther*, *6*(1), 290.

Wang, H., Yan, X., Aigner, H., Bracher, A., Nguyen, N. D., Hee, W. Y., Long, B. M., Price, G. D., Hartl, F. U., & Hayer-Hartl, M. (2019). Rubisco condensate formation by CcmM in beta-carboxysome biogenesis. *Nature*, *566*(7742), 131–135.

Wang, J., Choi, J. M., Holehouse, A. S., Lee, H. O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A., Drechsel, D., Poser, I., Pappu, R. V., Alberti, S., & Hyman, A. A. (2018). A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell*, *174*(3), 688–699 e616.

Wang, P., Li, J., Li, T., Li, K., Ng, P. C., Wang, S., Chriscoli, V., Basle, A., Marles-Wright, J., Zhang, Y. Z., & Liu, L. N. (2024). Molecular principles of the assembly and construction of a carboxysome shell. *Sci Adv*, *10*(48), eadr4227.

Wang, Y., Stessman, D. J., & Spalding, M. H. (2015). The CO2 concentrating mechanism and photosynthetic carbon assimilation in limiting CO2 : how Chlamydomonas works against the gradient. *Plant J*, *82*(3), 429–448.

Weerasooriya, H. N., DiMario, R. J., Rosati, V. C., Rai, A. K., LaPlace, L. M., Filloon, V. D., Longstreth, D. J., & Moroney, J. V. (2022). Arabidopsis plastid carbonic anhydrase betaCA5 is important for normal plant growth. *Plant Physiol*, *190*(4), 2173–2186.

Wheatley, N. M., Gidaniyan, S. D., Liu, Y., Cascio, D., & Yeates, T. O. (2013). Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. *Protein Sci*, *22*(5), 660–665.

Wheatley, N. M., Sundberg, C. D., Gidaniyan, S. D., Cascio, D., & Yeates, T. O. (2014). Structure and identification of a pterin dehydratase-like protein as a ribulose-bisphosphate carboxylase/oxygenase (RuBisCO) assembly factor in the alpha-carboxysome. *J Biol Chem*, *289*(11), 7973–7981.

Whitehead, L., Long, B. M., Price, G. D., & Badger, M. R. (2014). Comparing the in vivo function of alpha-carboxysomes and beta-carboxysomes in two model cyanobacteria. *Plant Physiol*, *165*(1), 398–411.

Whitney, S. M., Birch, R., Kelso, C., Beck, J. L., & Kapralov, M. V. (2015). Improving recombinant Rubisco biogenesis, plant photosynthesis and growth by coexpressing its ancillary RAF1 chaperone. *Proc Natl Acad Sci U S A*, *112*(11), 3564–3569.

Whitney, S. M., Houtz, R. L., & Alonso, H. (2011). Advancing our understanding and capacity to engineer nature's CO2-sequestering enzyme, Rubisco. *Plant Physiol*, *155*(1), 27–35.

Whitney, S. M., Sharwood, R. E., Orr, D., White, S. J., Alonso, H., & Galmes, J. (2011). Isoleucine 309 acts as a C4 catalytic switch that increases ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) carboxylation rate in Flaveria. *Proc Natl Acad Sci U S A*, *108*(35), 14688–14693.

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol*, *112*, 531–552.

Wilson, R. H., & Hayer-Hartl, M. (2018). Complex Chaperone Dependence of Rubisco Biogenesis. *Biochemistry*, *57*(23), 3210–3216.

Woodger, F. J., Badger, M. R., & Price, G. D. (2003). Inorganic carbon limitation induces transcripts encoding components of the CO(2)-concentrating mechanism in Synechococcus sp. PCC7942 through a redox-independent pathway. *Plant Physiol*, *133*(4), 2069–2080.

Wu, A., Brider, J., Busch, F. A., Chen, M., Chenu, K., Clarke, V. C., Collins, B., Ermakova, M., Evans, J. R., Farquhar, G. D., Forster, B., Furbank, R. T., Groszmann, M., Hernandez-Prieto, M. A., Long, B. M., McLean, G., Potgieter, A., Price, G. D., Sharwood, R. E.,…Hammer, G. L. (2023). A cross-scale analysis to understand and quantify the effects of photosynthetic enhancement on crop growth and yield across environments. *Plant Cell Environ*, *46*(1), 23–44.

Wunder, T., Cheng, S. L. H., Lai, S. K., Li, H. Y., & Mueller-Cajar, O. (2018). The phase separation underlying the pyrenoid-based microalgal Rubisco supercharger. *Nat Commun*, *9*(1), 5076.

Wunderlich, Z., Acton, T. B., Liu, J., Kornhaber, G., Everett, J., Carter, P., Lan, N., Echols, N., Gerstein, M., Rost, B., & Montelione, G. T. (2004). The protein target list of the Northeast Structural Genomics Consortium. *Proteins*, *56*(2), 181–187.

Wyatt, P. J. (1993). Light-Scattering and the Absolute Characterization of Macromolecules. *Analytica Chimica Acta*, *272*(1), 1–40.

Xu, J., & Zhang, Y. (2010). How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics*, *26*(7), 889–895.

Yamori, W., Takahashi, S., Makino, A., Price, G. D., Badger, M. R., & von Caemmerer, S. (2011). The roles of ATP synthase and the cytochrome b6/f complexes in limiting chloroplast electron transport and determining photosynthetic capacity. *Plant Physiol*, *155*(2), 956–962.

Yang, Y., Jones, H. B., Dao, T. P., & Castaneda, C. A. (2019). Single Amino Acid Substitutions in Stickers, but Not Spacers, Substantially Alter UBQLN2 Phase Transitions and Dense Phase Material Properties. *J Phys Chem B*, *123*(17), 3618–3629.

Yeates, T. O., Crowley, C. S., & Tanaka, S. (2010). Bacterial microcompartment organelles: protein shell structure and evolution. *Annu Rev Biophys*, *39*, 185–205.

Yusuf, A., Usman, A., Isah, M. B., Dang, M., & Zhang, X. (2025). Liquid-liquid phase separation in microorganisms: Insights into existence, functions, and applications. *Microbiol Res*, *292*, 128026.

Zallot, R., Oberg, N., & Gerlt, J. A. (2019). The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry*, *58*(41), 4169–4182.

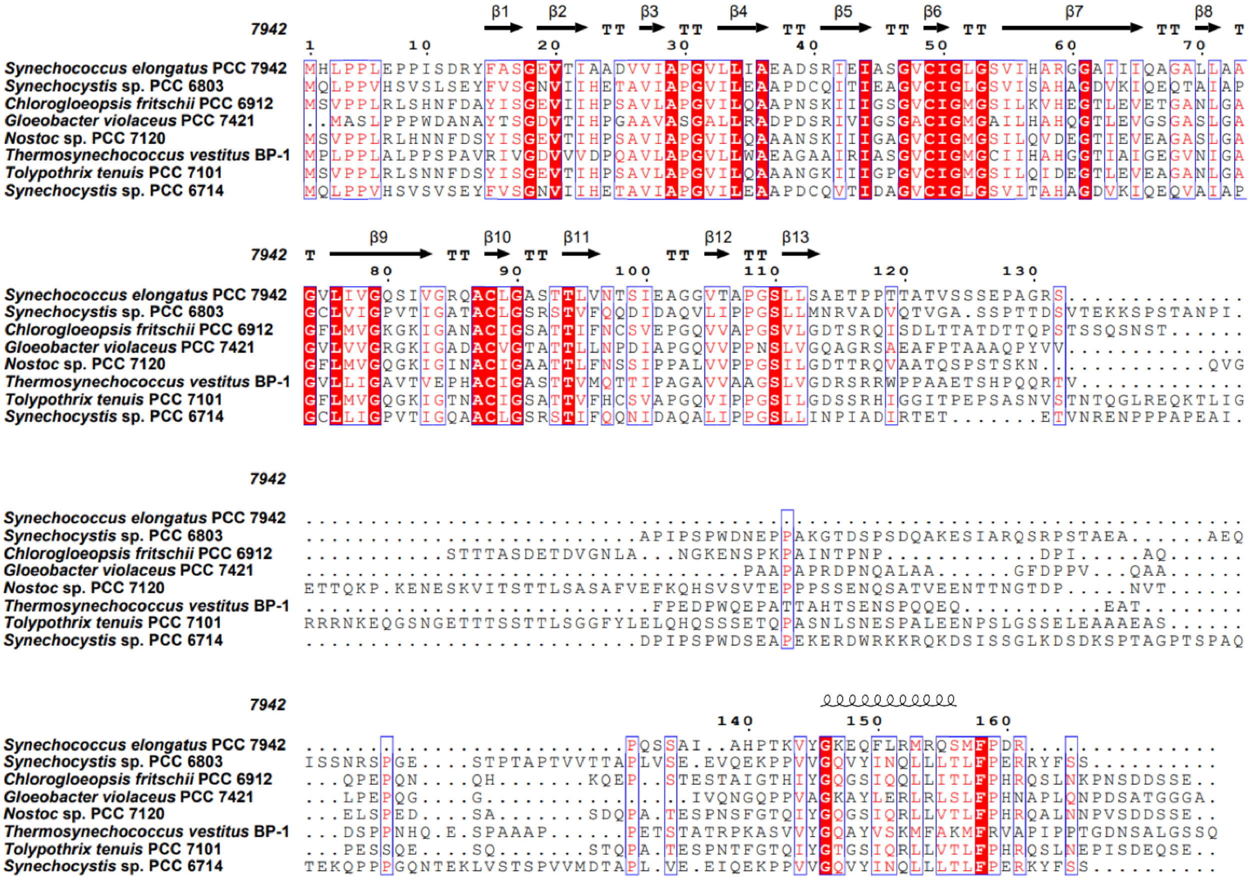
Zang, K., Wang, H., Hartl, F. U., & Hayer-Hartl, M. (2021). Scaffolding protein CcmM directs multiprotein phase separation in beta-carboxysome biogenesis. *Nat Struct Mol Biol*, *28*(11), 909–922.

Zhou, R. Q., Jiang, Y. L., Li, H., Hou, P., Kong, W. W., Deng, J. X., Chen, Y., Zhou, C. Z., & Zeng, Q. (2024). Structure and assembly of the alpha-carboxysome in the marine cyanobacterium Prochlorococcus. *Nat Plants*, *10*(4), 661–672.

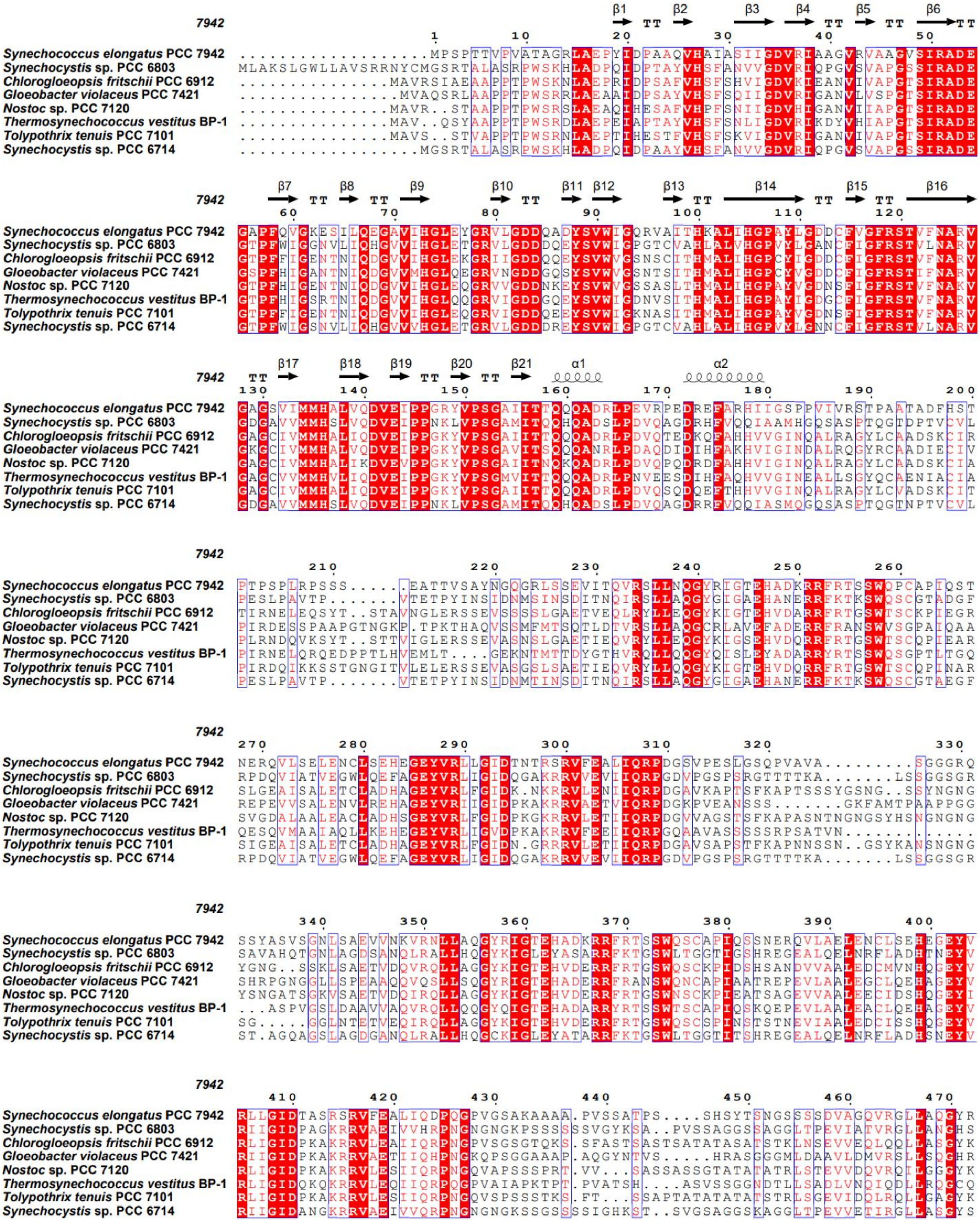
A. Appendix

A.1. Multiple-sequence alignment of ApN and CM proteins

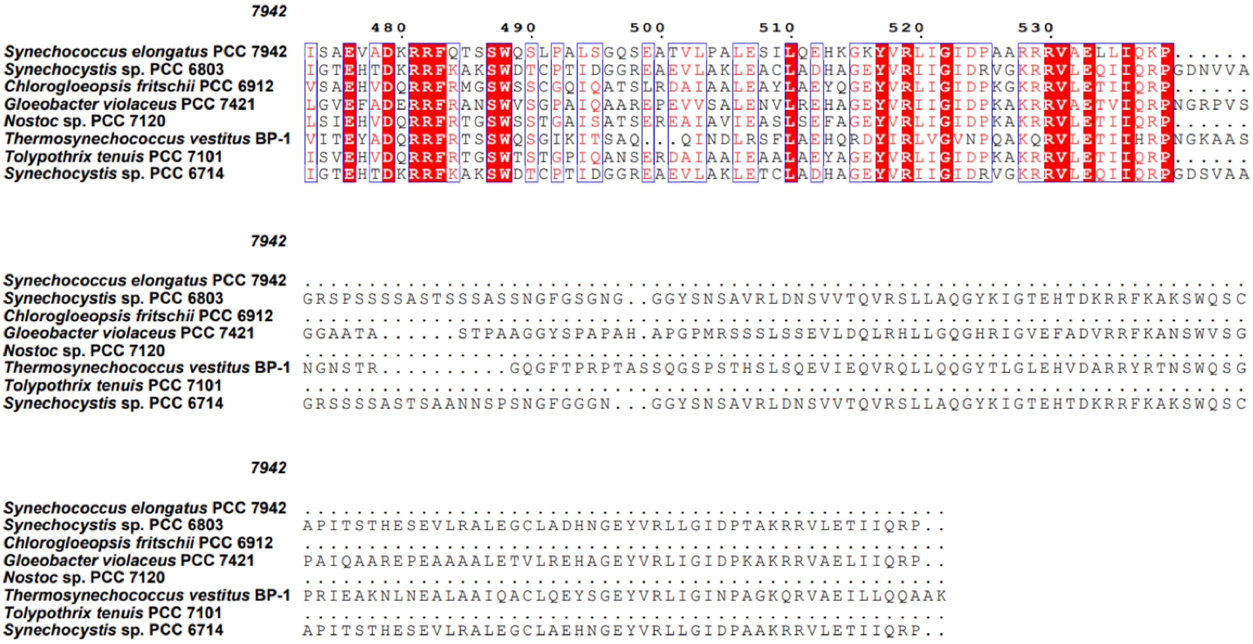
The following are the multiple-sequence alignments of the full-length ApN and CM proteins, respectively. These refer to Chapter **3.1**.



A.1 Multiple-sequence alignment of full-length ApN.
Multiple-sequence alignment of full-length ApN (residues 1-161) homologs were analyzed using ESPript 3.0. The secondary structural elements are shown above the alignment. Identical residues are shown in white text on a red background, and similar residues are shown in red. Homologous regions are outlined in blue. Species are as indicated in Chapter **2.2** and the figure.



```
                                               1         10        20        30        40        50
Synechococcus elongatus PCC 7942     ...................MPSPTTVPVATAGRLAEPYIDPAAQVHAIASIIGDVRIAAGVRVAAGVSIRADE
Synechocystis sp. PCC 6803           MLAKSLGWLLAVSRRNYCMGSRTALASRPWSKHLADPQIDPTAYVHSFANVVGDVRIQPGVSVAPGSSIRADE
Chlorogloeopsis fritschii PCC 6912   ...............MAVRSIAEAAPPTPWSKNLAEPRIDPSAFVHSFSHVIGDVKIEANVIVAPGTSIRADE
Gloeobacter violaceus PCC 7421       ................MVAQSRLAAPPTPWSRRLAEAAIDPSAYVHSPSQIIGDVRIGANVLVSPGTSIRADE
Nostoc sp. PCC 7120                  ...............MAVR..STAAPPTPWSRSLAEAQIHESAFVHPFSNIIGDVHIGANVIIAPGTSIRADE
Thermosynechococcus vestitus BP-1    ...............MAV..QSYAAPPTPWSRDLAEPEIAPTAYVHSFSNLIGDVRIKDYVHIAPGTSIRADE
Tolypothrix tenuis PCC 7101          ...............MAVS..STVAPPTPWSRNLAEPTIHESTFVHSFSKVIGDVRIGANVIVAPGTSIRADE
Synechocystis sp. PCC 6714           ..................MGSRTALASRPWSKHLADPQIDPAAYVHSFANVVGDVRIQPGVSVAPGSSIRADE

                                         60        70        80        90       100       110       120
Synechococcus elongatus PCC 7942     GAPFQVGKESILQEGAVIHGLEYGRVLGDDQADYSVWIGQRVAITHKALIHGPAYLGDDCFVGFRSTVFNARV
Synechocystis sp. PCC 6803           GTPFWIGGNVLIQHGVVIHGLETGRVLGDDDQEYSVWIGPGTCVAHLALVHGPVYLGANCFIGFRSTVLNARV
Chlorogloeopsis fritschii PCC 6912   GTPFFIGENTNIQDGVVIHGLEKGRIIGDDQQEYSVWVGSNSCITHMALIHGPCYIGDDCFIGFRSTVFNARV
Gloeobacter violaceus PCC 7421       GSPFHIGANTNIQDGVVMHGLQEGRVNGDDGQSYSVWVGSNTSITHMALIHGPCYVGDDCFIGFRSTIFNARV
Nostoc sp. PCC 7120                  GTPFHIGENTNIQDGVVIHGLEQGRVVGDDNKEYSVWVGSSASLTHMALIHGPAYVGDNSFIGFRSTVFNAKV
Thermosynechococcus vestitus BP-1    GTPFHIGSRTNIQDGVVIHGLQQGRVIGDDGQEYSVWIGDNVSITHMALIHGPAYIGDGCFIGFRSTVFNARV
Tolypothrix tenuis PCC 7101          GTPFFIGENTNIQDGVVIHGLEQGRVIGDDQEEYSVWIGKNASITHMALIHGPAYVGDNSFIGFRSTVFNARV
Synechocystis sp. PCC 6714           GTPFWIGSNVLIQHGVVVHGLETGRVLGDDDREYSVWIGPGTCVAHLALIHGPVYLGNNCFIGFRSTVLNARV

                                     130       140       150       160       170       180       190       200
Synechococcus elongatus PCC 7942     GAGSVIMMHALVQDVEIPPGRYVPSGAIITTQQQADRLPEVRPEDREFARHIIGSPPVIVRSTPAATADFHST
Synechocystis sp. PCC 6803           GDGAVVMMHSLVQDVEIPPNKLVPSGAMITQQHQADSLPDVQAGDRHFVQQIAAMHGQSASPTQGTDPTVCVL
Chlorogloeopsis fritschii PCC 6912   GAGCIVMMHALIQDVEIPPGKYVPSGAIITTQQQADRLPDVQTEDKQFAHHVVGINQALRAGYLCAADSKCIR
Gloeobacter violaceus PCC 7421       GKGCIVMMHALVQDVEIPPGKYVPSGAVITSQQQANRLPDAQDIDIHFAKHVIGINDALRQGYRCAADIECIV
Nostoc sp. PCC 7120                  GAGCIVMMHALIKDVEVPPGKYVPSGAIITNQKQADRLPDVQPQDRDFAHHVIGINQALRAGYLCAADSKCIA
Thermosynechococcus vestitus BP-1    GAGCVVMMHVLIQDVEIPPGKYVPSGMVITTQQQADRLPNVEESDIHFAQHVVGINEALLSGYQCAENIACIA
Tolypothrix tenuis PCC 7101          GAGCIVMMHALIQDVEIPPGKYVPSGAIITTQQQADRLPDVQSQDQEFTHHVVGINQALRAGYLCVADSKCIT
Synechocystis sp. PCC 6714           GDGAVVMMHSLVQDVEIPPNKLVPSGAMITQQHQADSLPDVQAGDRRFVQQIASMQGQSASPTQGTNPTVCVL

                                             210       220       230       240       250       260
Synechococcus elongatus PCC 7942     PTPSPLRPSSS......EATTVSAYNGQGRLSSEVITQVRSLLNQGYRIGTEHADKRRFRTSSWQPCAPIQST
Synechocystis sp. PCC 6803           PESLPAVTP.......VTETPYINSIDNMSINSDITNQIRSLLAQGYGIGAEHANERRFKTKSWQSCGTADGF
Chlorogloeopsis fritschii PCC 6912   TIRNELEQSYT..STAVNGLERSSEVSSSSLGAETVEQLRYLLEQGYKIGTEHVDARRFRTGSWTSCKPIEGR
Gloeobacter violaceus PCC 7421       PIRDESSPAAPGTNGKP.TPKTHAQVSSMFMTSQTLDTVRSLLAQGCRLAVEFADERRFRANSWVSGPAIQAA
Nostoc sp. PCC 7120                  PLRNDQVKSYT..STTVIGLERSSEVASNSLGAETIEQVRYLLEQGYKIGSEHVDQRRFRTGSWTSCQPIEAR
Thermosynechococcus vestitus BP-1    PIRNELQRQEDPPTLHVEMLT....GEKNTMTTDYGTHVRQLLQQGYQISLEYADARRYRTSSWQSGPTLTGQ
Tolypothrix tenuis PCC 7101          PIRDQIKKSSTGNGITVLELERSSEVASGSLSAETIEQVRYLLQQGYKIGTEHVDQRRFRTGSWTSCQPINAR
Synechocystis sp. PCC 6714           PESLPAVTP.......VTETPYINSIDNMTINSDITNQIRSLLAQGYGIGAEHANERRFKTKSWQSCGTAEGF

                                     270       280       290       300       310       320            330
Synechococcus elongatus PCC 7942     NERQVLSELENCLSEHEGEYVRLLGIDTNTRSRVFEALIQRPDGSVPESLGSQPVAVA.........SGGGRQ
Synechocystis sp. PCC 6803           RPDQVIATVEGWLQEFAGEYVRLIGIDQGAKRRVVEVIIQRPGDVPGSPSRGTTTTKA.......LSSGGSGR
Chlorogloeopsis fritschii PCC 6912   SLGEAISALETCLADHAGEYVRLFGIDK.NKRRVLENIIQRPDGAVKAPTSFKAPTSSSYGSNG..SSYNGNG
Gloeobacter violaceus PCC 7421       REPEVVSALENVLREHAGEYVRIIGIDPKAKRRVAETVIQRPDGKPVEANSSS.......GKFAMTPAAPPGG
Nostoc sp. PCC 7120                  SVGDALAALEACLADHSGEYVRLFGIDPKGKRRVLETIIQRPDGVVAGSTSFKAPASNTNGNGSYHSNGNGNG
Thermosynechococcus vestitus BP-1    QESQVMAAIAQLLKEHEGEYVRLIGVDPKAKRRVFEEIIQRPGQAAVASSSSSRPSATVN.............
Tolypothrix tenuis PCC 7101          SIGEAISALETCLADHAGEYVRLFGIDN.GRRRVLETIIQRPDGAVSAPSTFKAPNNSSN..GSYKANSNGNG
Synechocystis sp. PCC 6714           RPDQVIATVEGWLQEFAGEYVRLIGIDQGAKRRVVEVIIQRPGDVPGSPSRGTTTTKA.......LSSGGSGR

                                             340       350       360       370       380       390       400
Synechococcus elongatus PCC 7942     SSYASVSGNLSAEVVNKVRNLLAQGYRIGTEHADKRRFRTSSWQSCAPIQSSNERQVLAELENCLSEHEGEYV
Synechocystis sp. PCC 6803           SAVAHQTGNLAGDSANQLRALLHQGYKIGLEYASARRFKTGSWLTGGTIGSHREGEALQELNRFLADHTNEYV
Chlorogloeopsis fritschii PCC 6912   YGNG..SSKLSAETVDQVRQLLAGGYKIGTEHVDERRFRTGSWQSCKPIDSHSANDVVAALEDCMVNHQGEYV
Gloeobacter violaceus PCC 7421       SHRPGNGGLLSPEAAQQVQSLLSQGYRIGTEHADERRFRANSWQNCAPIAATREPEVLAALEGCLQEHAGEYV
Nostoc sp. PCC 7120                  YSNGATSGKVSAETVDQIRQLLAGGYKIGTEHVDERRFRTGSWNSCKPIEATSAGEVVAALEECIDSHQGEYI
Thermosynechococcus vestitus BP-1    ...ASPVGSLDAAVVAQVRQLLQQGYQIGTEHADARRYRTSSWTSCAPIQSKQEPEVLAALEACLQEHAGEYV
Tolypothrix tenuis PCC 7101          SG.....GGLNTETVEQIRQLLAGGYKIGTEHVDERRFRTGSWQSCSPINSTSTNEVIAALEDCISSHQGEYV
Synechocystis sp. PCC 6714           ST.AGQAGSLAGDGANQLRALLHQGCKIGLEYATARRFKTGSWLTGGTITSHREGEALQELNRFLADHSNEYV

                                          410       420       430       440       450       460       470
Synechococcus elongatus PCC 7942     RLLGIDTASRSRVFEALIQDPQGPVGSAKAAAA.PVSSATPS....SHSYTSNGSSSSDVAGQVRGLLAQGYR
Synechocystis sp. PCC 6803           RIIGIDPAGKRRVAEIVVHRPNGNGNGKPSSSSSSVGYKSA..PVSSAGGSSAGGLTPEVIATVRGLLANGHS
Chlorogloeopsis fritschii PCC 6912   RLIGIDPKAKRRVLEAIIQRPNGPVSGSGTQKS.SFASTSASTSATATASATSTKLNSEVVEQLQQLLASGYK
Gloeobacter violaceus PCC 7421       RIIGIDPKAKRRVAETIIQHPNGKQPSGGAAAP.AQGYNTVS.....HRASGGGMLDAAVLDMVRSLLSQGHR
Nostoc sp. PCC 7120                  RLIGIDPKAKRRVLESIIQRPNGQVAPSSSPRT.VV...SASSASSGTATATATRLSTEVVDQVRQILGGGYK
Thermosynechococcus vestitus BP-1    RLIGIDQKQKRRVLEQIIQRPQGPVAIAPKTPT.PVATSH....ASVSSGGNDTLLSADLVNQIQDLLRQGCQ
Tolypothrix tenuis PCC 7101          RLIGIDPKAKRRVLESIIQRPNGQVSPSSSTKS.FT...SSAPTATATATATSTRLSGEVIDQLRQLLGAGYK
Synechocystis sp. PCC 6714           RIIGIDANGKRRVAEIVVQRPNGNGNGKSSGSSSSIGHKST..SVGSAGGSKAGGLTPEVVETIRGLLASGYS
```



```
                                 7942
                                          480       490       500       510       520       530
Synechococcus elongatus PCC 7942    ISAEVADKRRFQTSSWQSLPALSGQSEATVLPALESILQEHKGKYVRLIGIDPAARRRVAELLIQKP......
Synechocystis sp. PCC 6803          IGTEHTDKRRFKAKSWDTCPTIDGGREAEVLAKLEACLADHAGEYVRIIGIDRVGKRRVLEQIIQRPGDNVVA
Chlorogloeopsis fritschii PCC 6912  VSAEHVDQRRFRMGSWSSCGQIQATSLRDAIAALEAYLAEYQGEYVRLIGIDPKGKRRVLETIIQRP......
Gloeobacter violaceus PCC 7421      LGVEFADERRFRANSWVSGPAIQAAREPEVVSALENVLREHAGEYVRIIGIDPKAKRRVAETVIQRPNGRPVS
Nostoc sp. PCC 7120                 LSIEHVDQRRFRTGSWSSTGAISATSEREAIAVIEASLSEFAGEYVRLIGIDPKAKRRVLETIIQRP......
Thermosynechococcus vestitus BP-1   VITEYADQRRFRTSSWQSGIKITSAQ...QINDLRSFLAEHQRDYIRLVGVNPQAKQRVLETIIHRPNGKAAS
Tolypothrix tenuis PCC 7101         ISVEHVDQRRFRTGSWTSTGPIQANSERDAIAAIEAALAEYAGEYVRLIGIDPKAKRRVLETIIQRP......
Synechocystis sp. PCC 6714          IGTEHTDKRRFKAKSWDTCPTIDGGREAEVLAKLETCLADHAGEYVRIIGIDRVGKRRVLEQIIQRPGDSVAA

                                 7942

Synechococcus elongatus PCC 7942    .........................................................................
Synechocystis sp. PCC 6803          GRSPSSSSASTSSSASSNGFGSGNG..GGYSNSAVRLDNSVVTQVRSLLAQGYKIGTEHTDKRRFKAKSWQSC
Chlorogloeopsis fritschii PCC 6912  .........................................................................
Gloeobacter violaceus PCC 7421      GGAATA......STPAAGGYSPAPAH.APGPMRSSSLSSEVLDQLRHLLGQGHRIGVEFADVRRFKANSWVSG
Nostoc sp. PCC 7120                 .........................................................................
Thermosynechococcus vestitus BP-1   NGNSTR..........GQGFTPRPTASSQGSPSTHSLSQEVIEQVRQLLQQGYTLGLEHVDARRYRTNSWQSG
Tolypothrix tenuis PCC 7101         .........................................................................
Synechocystis sp. PCC 6714          GRSSSSASTSAANNSPSNGFGGGN...GGYSNSAVRLDNSVVTQVRSLLAQGYKIGTEHTDKRRFKAKSWQSC

                                 7942

Synechococcus elongatus PCC 7942    ..................................................
Synechocystis sp. PCC 6803          APITSTHESEVLRALEGCLADHNGEYVRLLGIDPTAKRRVLETIIQRP..
Chlorogloeopsis fritschii PCC 6912  ..................................................
Gloeobacter violaceus PCC 7421      PAIQAAREPEAAAALETVLREHAGEYVRLIGIDPKAKRRVAELIIQRP..
Nostoc sp. PCC 7120                 ..................................................
Thermosynechococcus vestitus BP-1   PRIEAKNLNEALAAIQACLQEYSGEYVRLIGINPAGKQRVAEILLQQAAK
Tolypothrix tenuis PCC 7101         ..................................................
Synechocystis sp. PCC 6714          APITSTHESEVLRALEGCLAEHNGEYVRLLGIDPAAKRRVLETIIQRP..
```

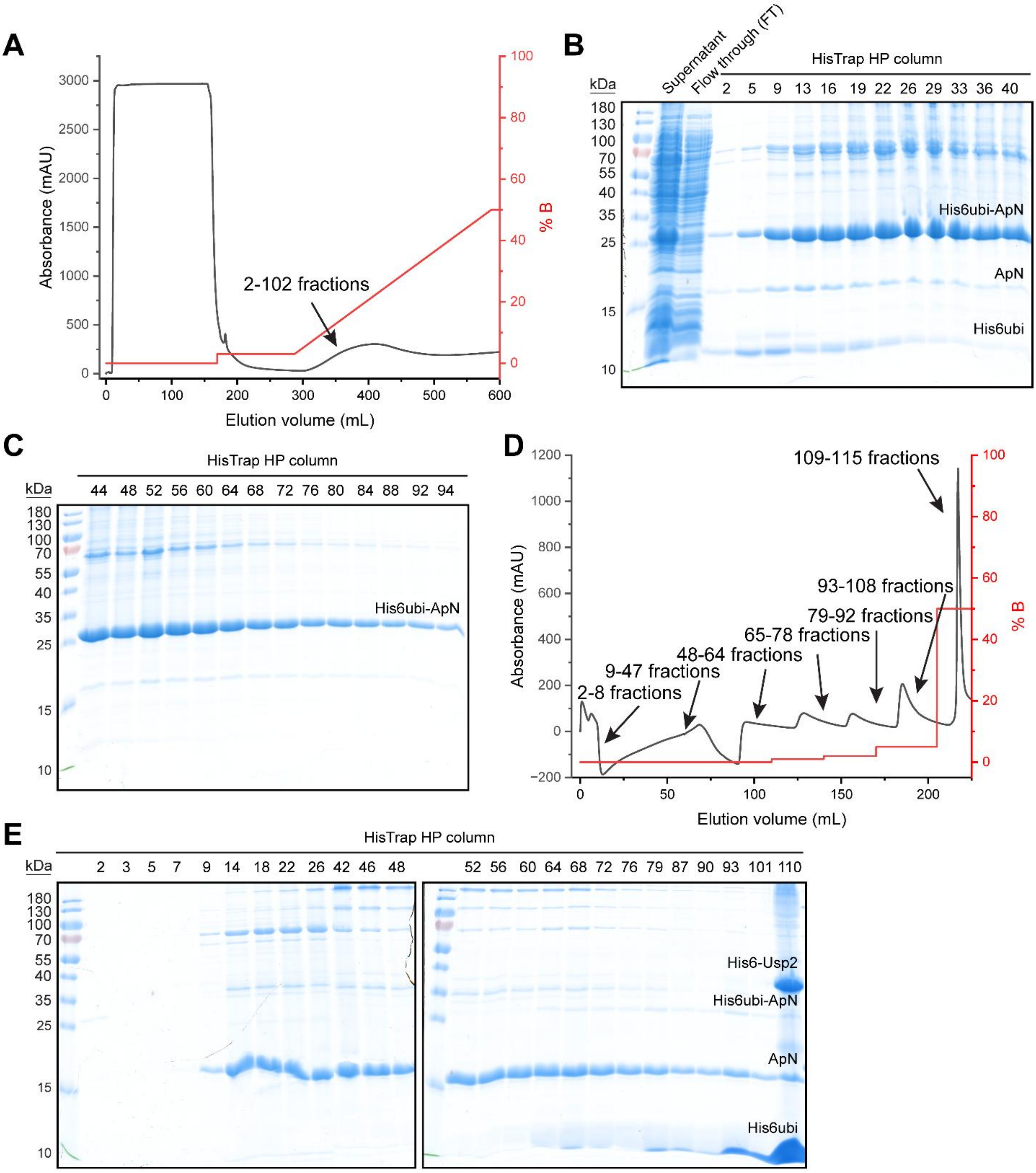
A.2 Multiple-sequence alignment of full-length CM.

Multiple-sequence alignment of full-length CM (residues 1-539) homologs were analyzed using ESPript 3.0. The secondary structural elements are shown above the alignment. Identical residues are shown in white text on a red background, and similar residues are shown in red. Homologous regions are outlined in blue. Species are as indicated in Chapter **2.2** and the figure.

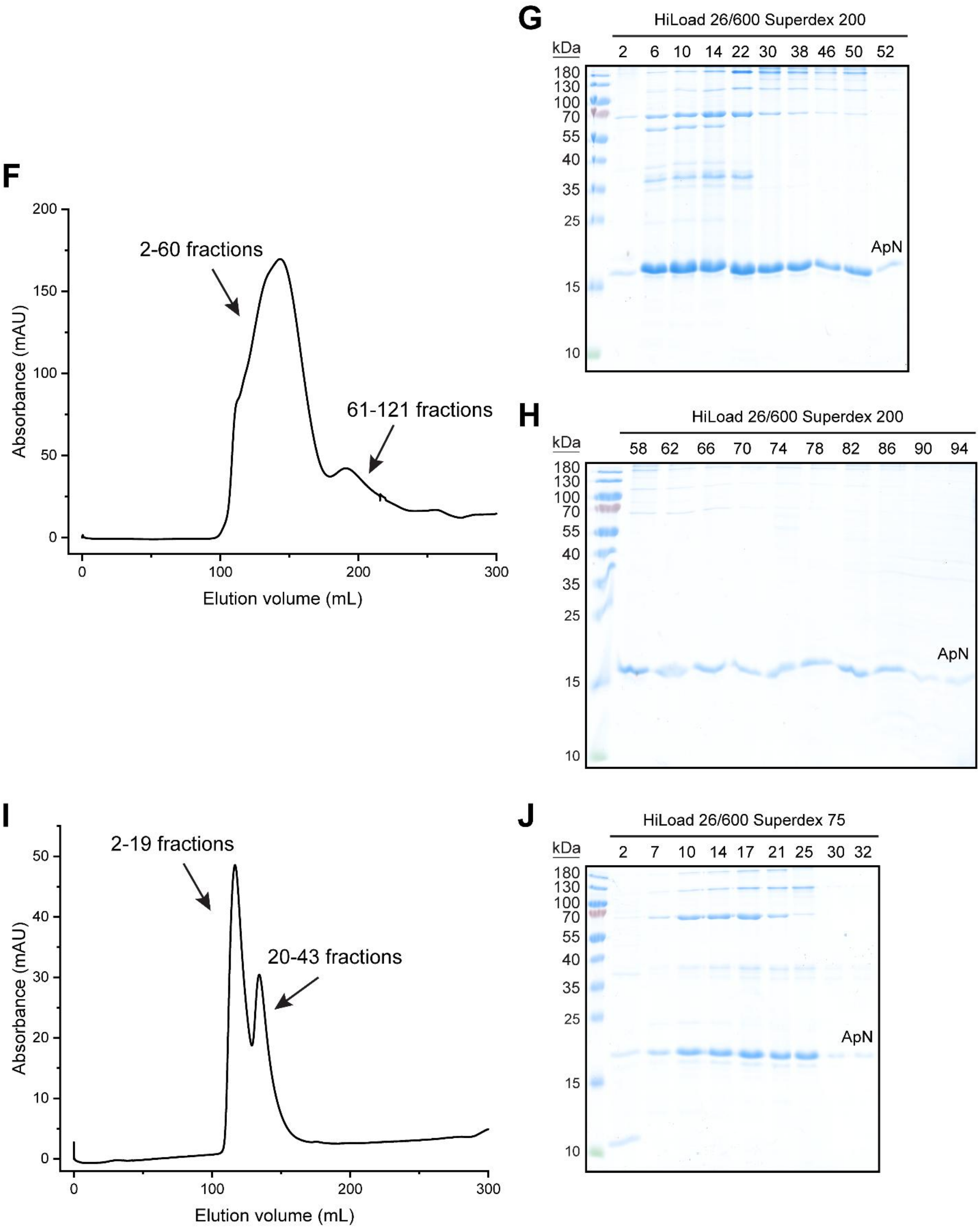
A.2. Protein purification

In this chapter, we describe the purification of the wild-type proteins, including ApN, SIIApN, the CM-ApN hetero-complex, the (CM-H_{10})-SIIApN hetero-complex, and the (CM-H_{10})-ApN hetero-complex. Their respective mutants follow the same procedure. These refer to Chapter **2.4.1.2**.

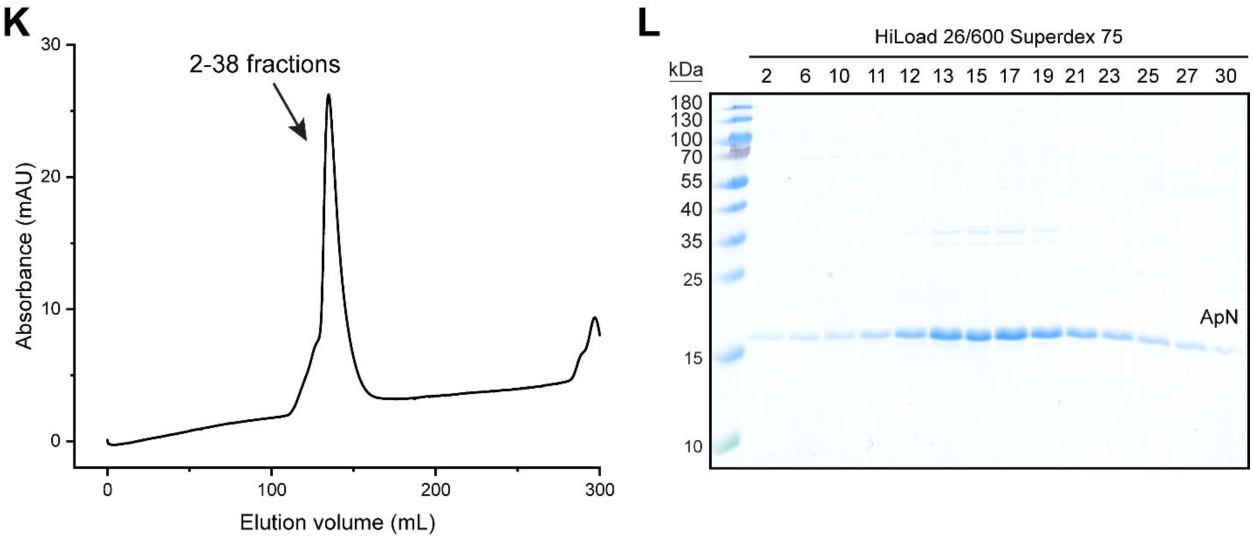
A.2.1. Purification of ApN



...continues on next page...

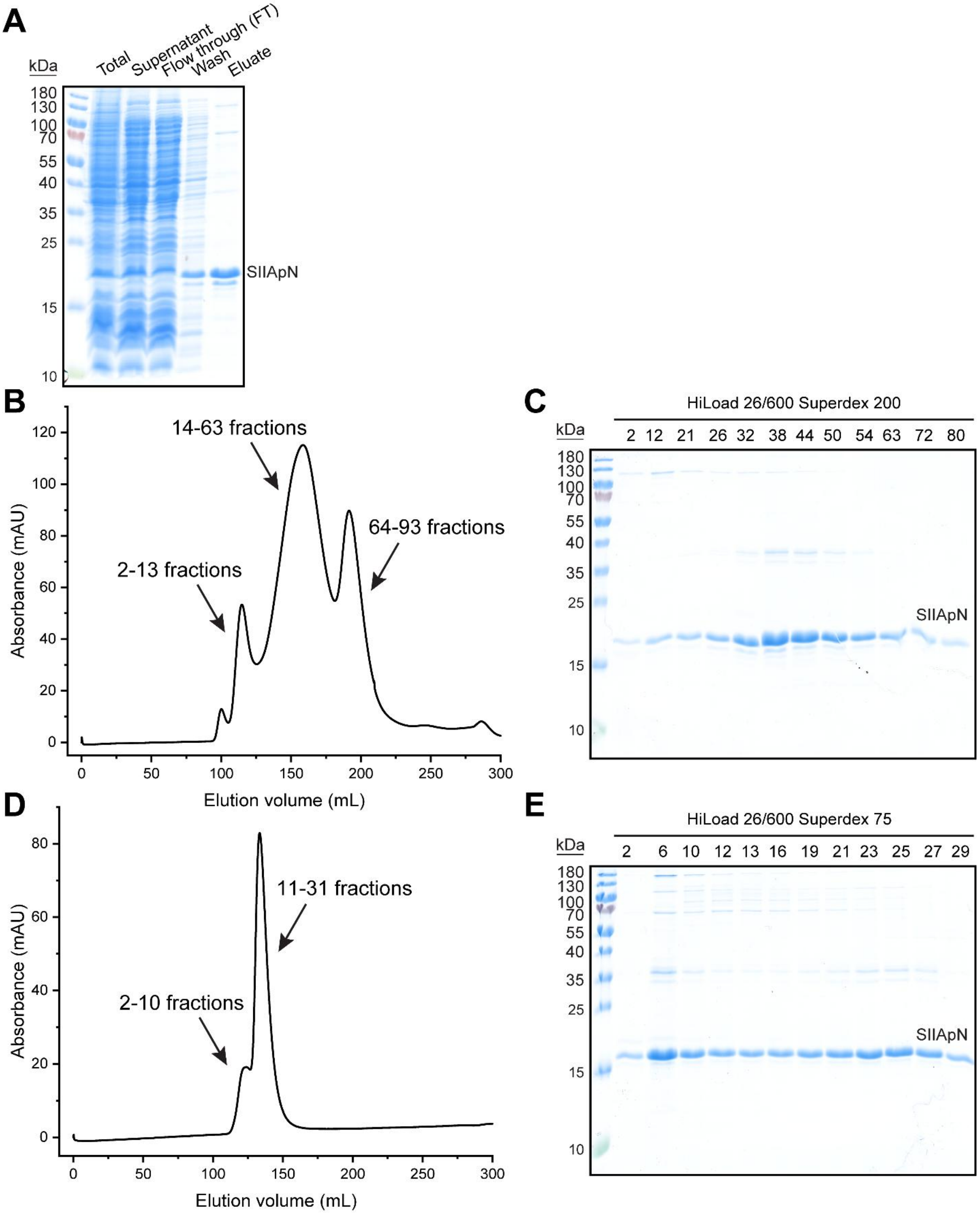


...continues on next page...



A.3 Representative purification of ApN.
(A-L) Purification of ApN. **(A-C)** Purification of His6ubi-ApN with a HisTrap column. Chromatogram of affinity purification with a HisTrap column **(A)**. Protein was loaded and washed prior to gradient elution with 20-500 mM imidazole. SDS-PAGE analysis of the protein during cell lysis and purification **(B-C)**. His6-Usp2 was added to 9-60 fractions overnight to cleave His6ubi. **(D-E)** Purification of ApN with a HisTrap column. Chromatogram of the affinity purification with a HisTrap column after His6ubi cleavage **(D)**. Protein was buffer exchanged to 0 mM imidazole before being loaded onto the column. SDS-PAGE analysis of protein **(E)**. The ApN protein is in the flow through and wash fractions (9-103 fractions). **(F-H)** Purification of ApN with a HiLoad 26/600 Superdex 200 column. Protein from **(D)** (50-90 fractions) was loaded onto a HiLoad 26/600 Superdex 200 column **(F)**. SDS-PAGE analysis of protein **(G-H)**. **(I-J)** Purification of ApN with a HiLoad 26/600 Superdex 75 column. Protein from the second peak of **(F)** (20-42 fractions) was loaded onto a HiLoad 26/600 Superdex 75 column **(I)**. SDS-PAGE analysis of protein **(J)**. **(K-L)** Purification of ApN with a HiLoad 26/600 Superdex 75 column. Protein from the second peak of **(I)** was loaded onto a HiLoad 26/600 Superdex 75 column **(K)**. SDS-PAGE analysis of protein **(L)**. 12-19 fractions were collected.

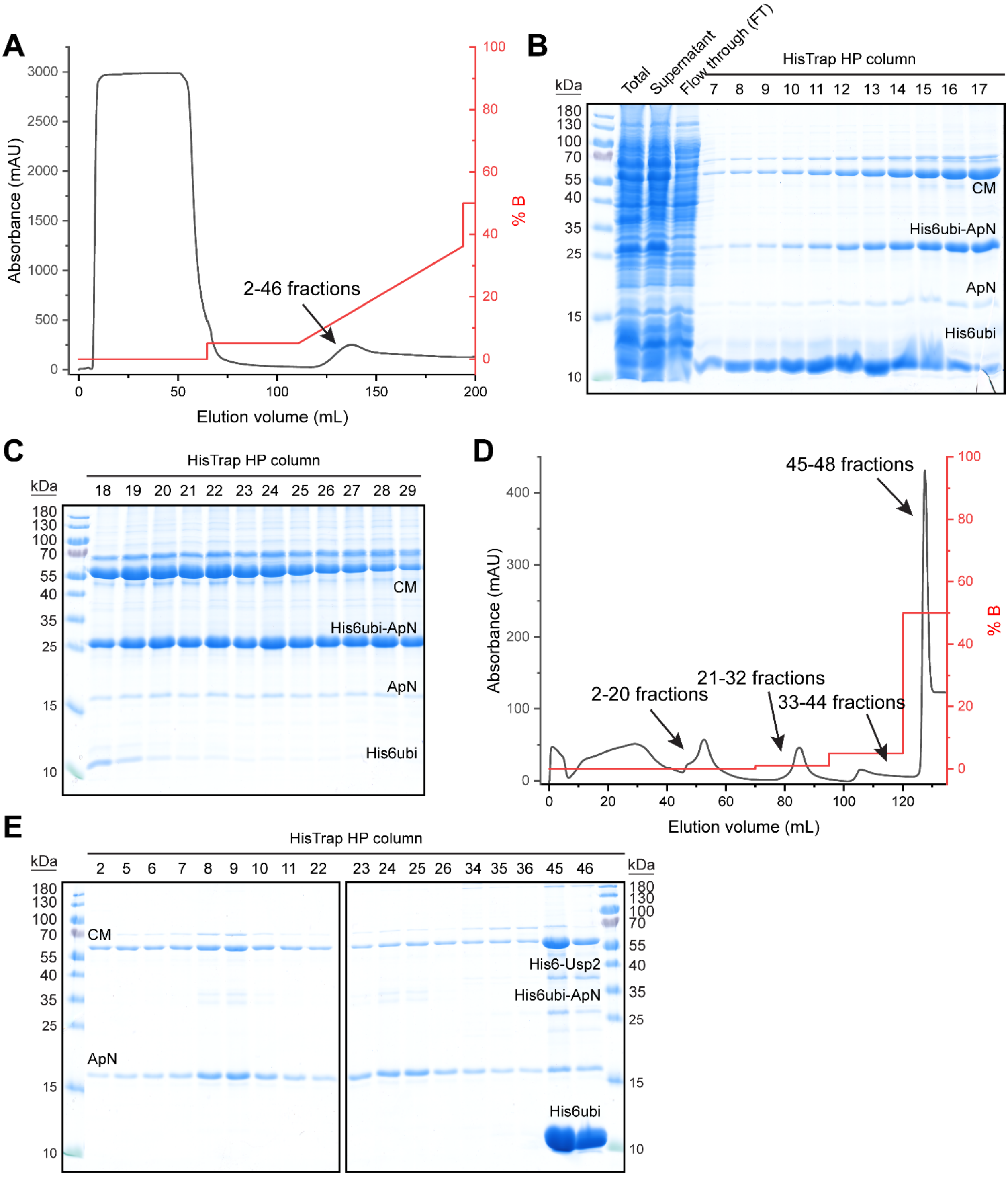
A.2.2. Purification of SIIApN



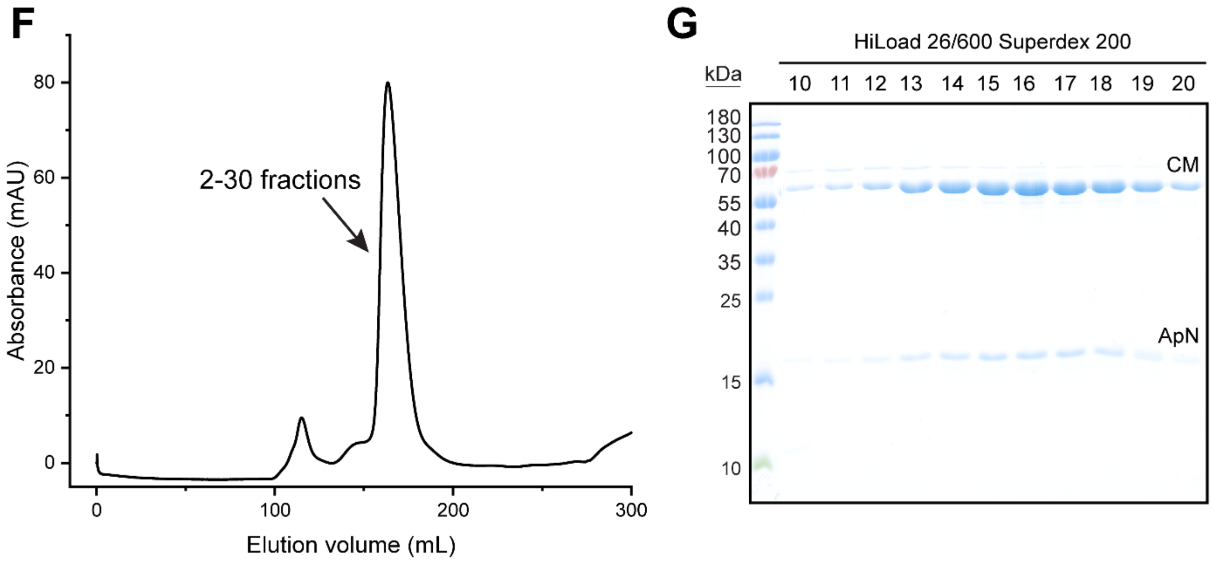
A.4 Representative purification of SIIApN.

(A-E) Purification of SIIApN. **(A)** Purification of SIIApN with a Strep-Tactin gravity column. **(B-C)** Purification of SIIApN with a HiLoad 26/600 Superdex 200 column. Chromatogram of the SEC purification with HiLoad 26/600 Superdex 200 column **(B)**. SDS-PAGE analysis of protein **(C)**. **(D-E)** Purification of SIIApN with a HiLoad 26/600 Superdex 75 column. Protein from the last peak of **(B)** (63-80 fractions) was loaded onto a HiLoad 26/600 Superdex 75 column **(D)**. SDS-PAGE analysis of protein **(E)**. 13-21 fractions were collected.

A.2.3. Purification of CM-ApN



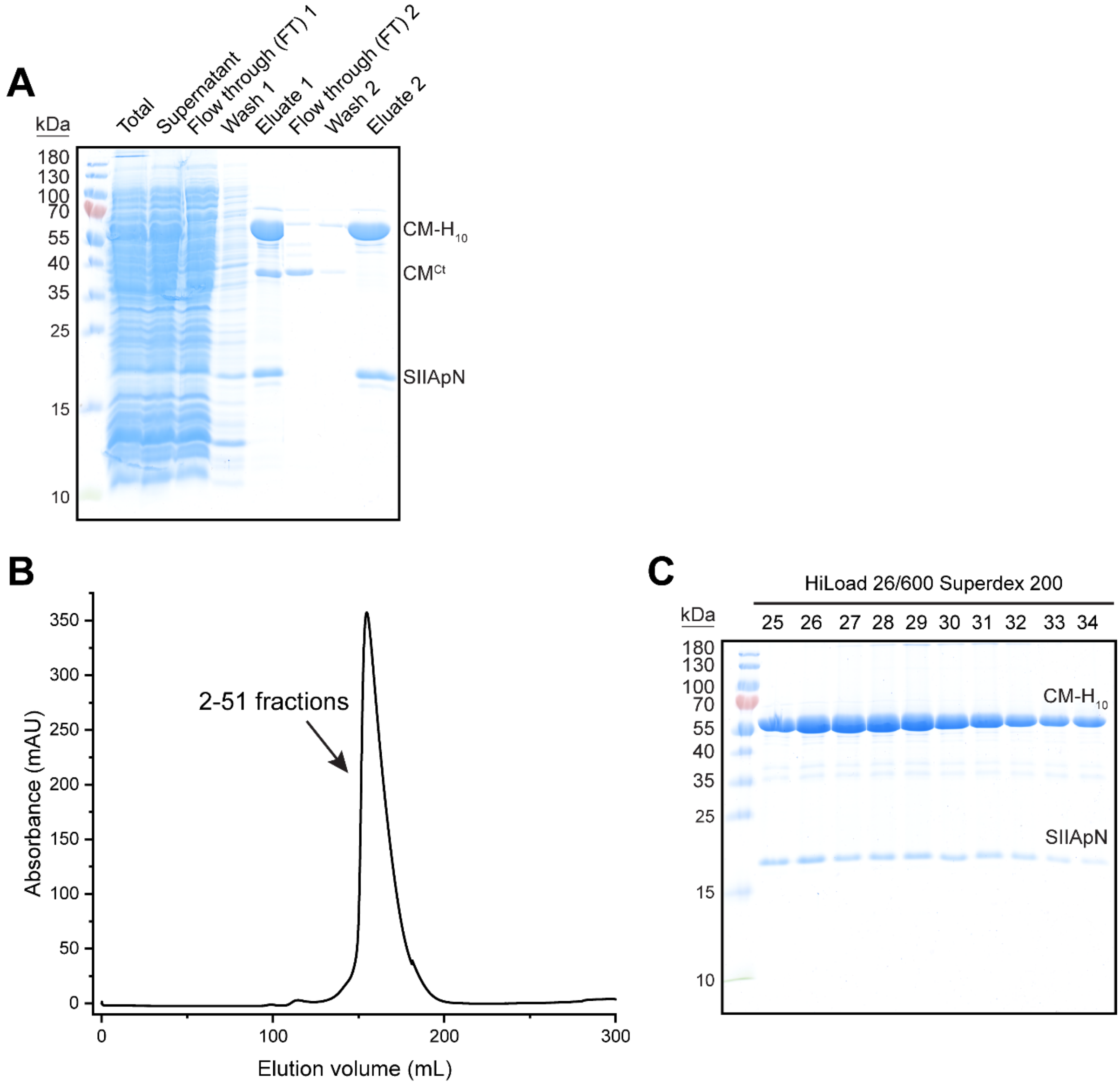
...continues on next page...



A.5 Representative purification of CM-ApN.

(A-G) Purification of CM-ApN. **(A-C)** Purification of CM-(His6ubi-ApN) with a HisTrap column. Chromatogram of affinity purification with a HisTrap column **(A)**. Protein was loaded and washed prior to gradient elution with 20-500 mM imidazole. SDS-PAGE analysis of the protein during cell lysis and purification **(B-C)**. His6-Usp2 was added to 15-22 fractions overnight to cleave His6ubi. **(D-E)** Purification of CM-ApN with a HisTrap column. Chromatogram of the affinity purification with a HisTrap column after His6ubi cleavage **(D)**. Protein was buffer exchanged to 0 mM imidazole before being loaded onto the column. SDS-PAGE analysis of protein **(E)**. The CM-ApN protein is in the flow through and wash fractions (2-11 and 22-26 fractions). **(F-G)** Purification of CM-ApN with a HiLoad 26/600 Superdex 200 column. Protein from **(D)** was loaded onto a HiLoad 26/600 Superdex 200 column **(F)**. SDS-PAGE analysis of protein **(G)**. 12-20 fractions were collected.

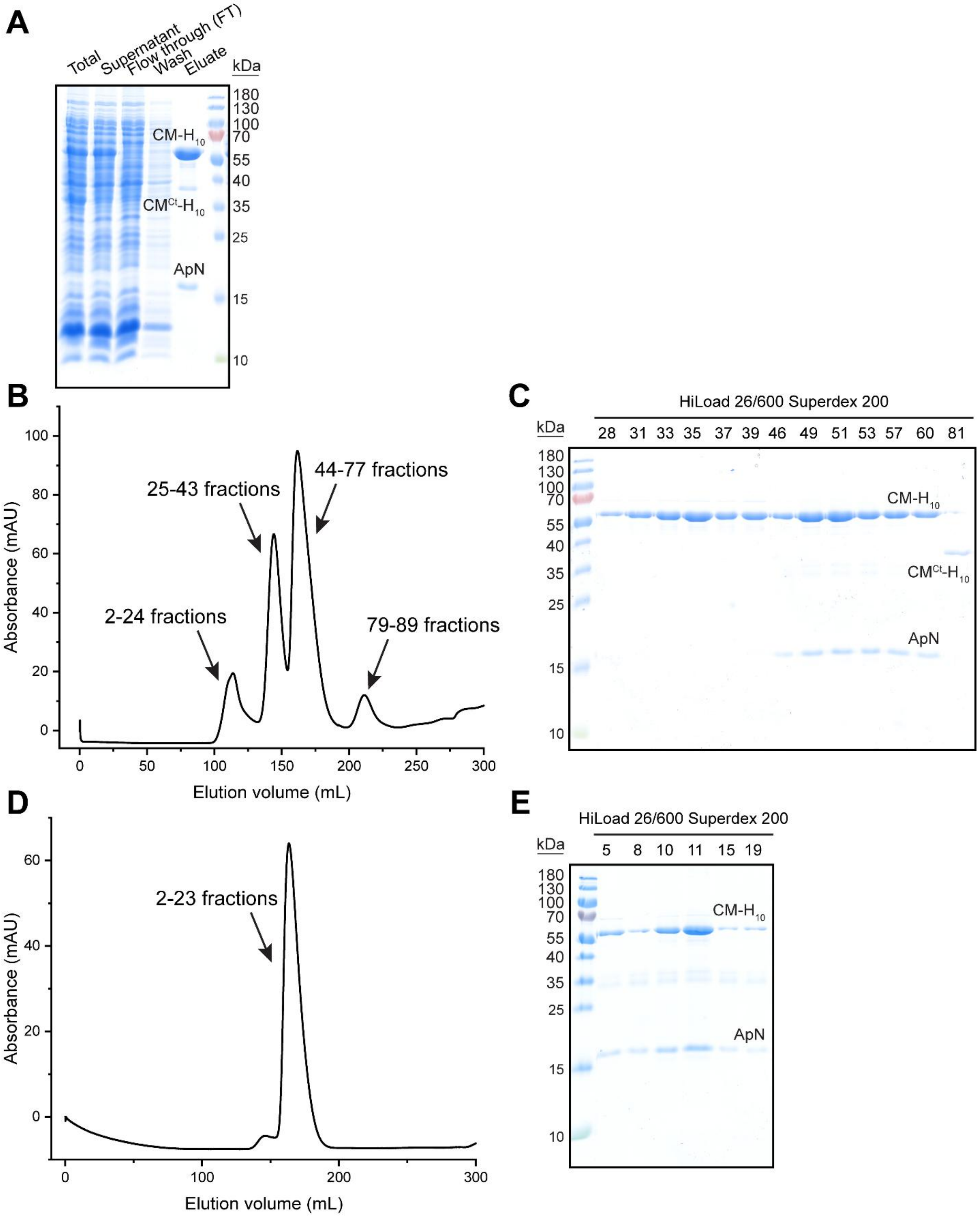
A.2.4. Purification of (CM-H_{10})-SIIApN



A.6 Representative purification of (CM-H_{10})-SIIApN.

(A-C) Purification of (CM-H_{10})-SIIApN. **(A)** Two-step affinity purification of (CM-H_{10})-SIIApN with a Ni-NTA gravity column followed by a Strep-Tactin gravity column. SDS-PAGE analysis of protein was shown. **(B-C)** Purification of (CM-H_{10})-SIIApN with a HiLoad 26/600 Superdex 200 column. Chromatogram of the SEC purification with HiLoad 26/600 Superdex 200 column **(B)**. SDS-PAGE analysis of protein **(C)**. 25-34 fractions were collected.

A.2.5. Purification of (CM-H$_{10}$)-ApN



A.7 Representative purification of (CM-H$_{10}$)-ApN.

(A-E) Purification of (CM-H$_{10}$)-ApN. **(A)** Purification of (CM-H$_{10}$)-ApN with a Ni-NTA gravity column. SDS-PAGE analysis of protein was shown. **(B-C)** Purification of (CM-H$_{10}$)-

ApN with a HiLoad 26/600 Superdex 200 column. Chromatogram of the SEC purification with a HiLoad 26/600 Superdex 200 column **(B)**. SDS-PAGE analysis of protein **(C)**. **(D-E)** Purification of (CM-H_{10})-ApN with a HiLoad 26/600 Superdex 200 column. (CM-H_{10})-ApN fractions from **(B)** (46-60 fractions) were loaded onto a HiLoad 26/600 Superdex 200 column for further purification **(D)**. SDS-PAGE analysis of protein **(E)**. 5-13 fractions were collected.

A.3. Molar mass and hydrodynamic radius of proteins determined by SEC-MALS

The following are the theoretical and measured molar mass of proteins used in this study. These refer to Chapters **3.1**, **3.3**, **3.4** and **3.5**.

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

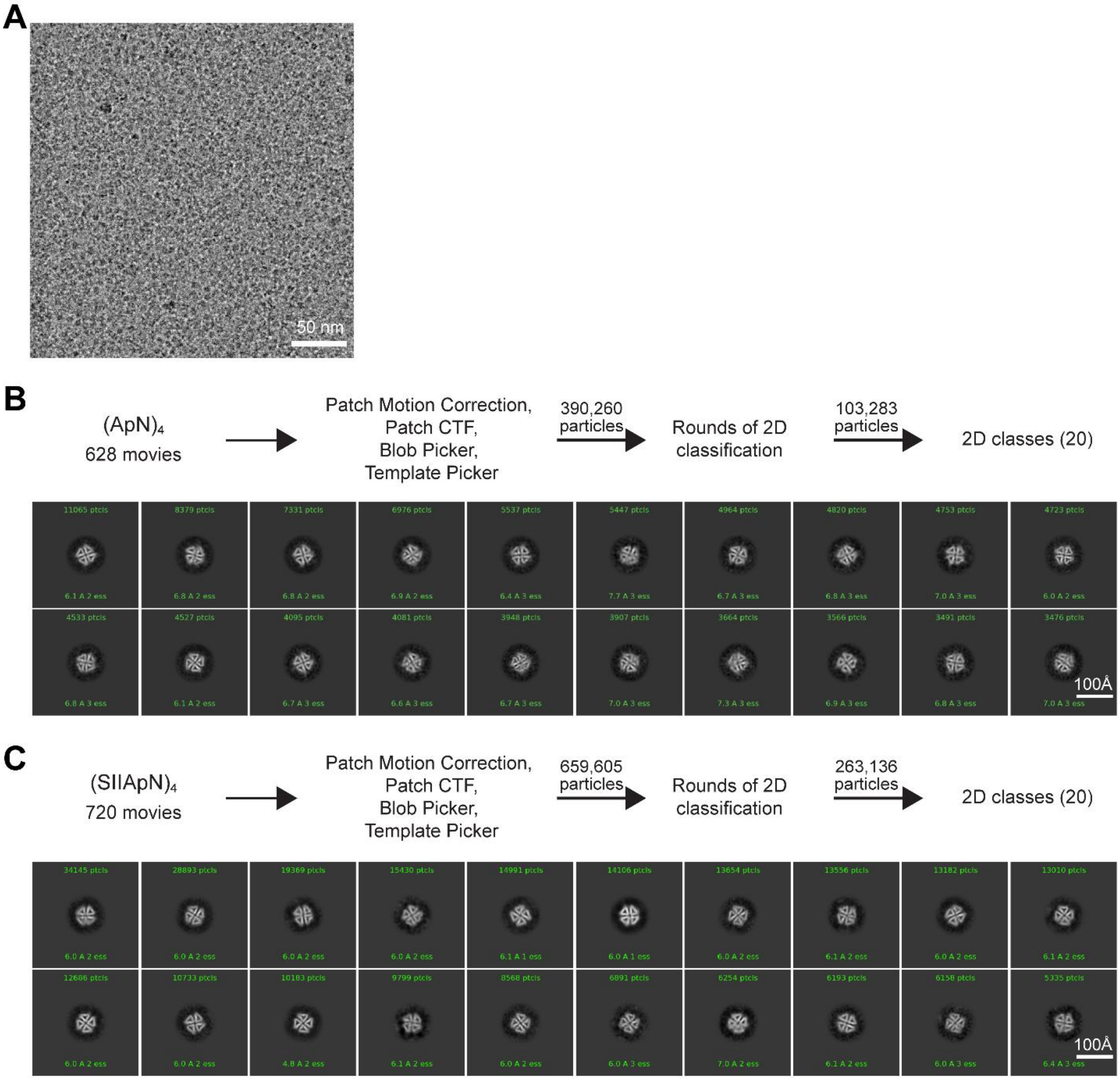
Sample*	**Molar mass Theoretical Protomer Da**	**1CM:1ApN Theoretical Da**	**1CM:2ApN Theoretical Da**	**1CM:3ApN Theoretical Da**	**SEC-MALS Expt. 01 kDa**	**SEC-MALS Expt. 02 kDa**	**SEC-MALS Avearge kDa**	**Oligomeric state**
ApN	16326.77	···	···	···	**63.07**	**64.02**	**~63.5**	homo-tetramer (theoretical 65.31 kDa)
SIIApN	17712.23	···	···	···	**69.54**	**69.07**	**~69.3**	homo-tetramer (theoretical 70.85 kDa)
CM	57832.84	···	···	···	**169.00**	**167.70**	**~168.4**	homo-trimer (theoretical 173.50 kDa)
CM-H$_{10}$	59549.57	···	···	···	**181.50**	**181.70**	**~181.6**	homo-trimer (theoretical 178.65 kDa)
CM^{Nt} (1-198)	20976.88	···	···	···	**64.55**	**64.09**	**~64.3**	homo-trimer (theoretical 62.93 kDa)
CM^{Nt}-H$_{10}$	22693.61	···	···	···	**68.97**	**69.42**	**~69.2**	homo-trimer (theoretical 68.08kDa)
CM + ApN		74159.61	90486.38	106813.15	**105.90**	**101.60**	**~103.8**	hetero-tetramer (ApN)$_3$:CM
CM-H$_{10}$ + SIIApN		77261.8	94974.03	112686.26	**112.00**	**111.10**	**~111.6**	hetero-tetramer (SIIApN)$_3$:CM-H$_{10}$
CM-H$_{10}$ + ApN		75876.34	92203.11	108529.88	**114.50**	**107.20**	**~110.9**	hetero-tetramer (ApN)$_2$:CM-H$_{10}$
CM-H$_{10}$ + ApNΔN		74795.06	90040.55	105286.04	**88.00**	**86.78**	**~87.4**	hetero-trimer (ApNΔN)$_2$:CM-H$_{10}$
CM^{Nt} + ApN		37303.65	53630.42	69957.19	**68.05**	**67.88**	**~68.0**	hetero-tetramer (ApN)$_3$: CM^{Nt}
CM^{Nt}-H$_{10}$ + SIIApN		40405.84	58118.07	75830.3	**77.84**	**77.98**	**~77.9**	hetero-tetramer (SIIApN)$_3$:CM^{Nt}-H$_{10}$
CM^{Nt}-H$_{10}$ + SIIApN (H_2O_2)		40405.84	58118.07	75830.3	**58.00**	**62.11**	**~60.1**	hetero-trimer (SIIApN)$_2$:CM^{Nt}-H$_{10}$
CM + ApN-2A		74095.49	90358.14	106620.8	**111.80**	**111.70**	**~111.8**	hetero-tetramer (ApN-2A)$_3$:CM
CM-H$_{10}$-SIIApN-2A		77197.68	94845.79	112493.9	**114.50**	**114.70**	**~114.6**	hetero-tetramer (SIIApN-2A)$_3$:CM-H$_{10}$
CM^{Nt}-H$_{10}$-SIIApN-2A		40341.72	57989.83	75637.94	**81.16**	**79.54**	**~80.4**	hetero-tetramer (SIIApN-2A)$_3$:CM^{Nt}-H$_{10}$
CM^{Nt}-H$_{10}$-SIIApN-2A (H_2O_2)		40341.72	57989.83	75637.94	**79.21**	**81.00**	**~80.11**	hetero-tetramer (SIIApN-2A)$_3$:CM^{Nt}-H$_{10}$
CM + ApN-2S (2023)		74127.49	90422.14	106716.8	**91.76**	**91.52**	**~91.6**	hetero-trimer (ApN-2S)$_2$:CM
CM + ApN-2S (2024)		74127.49	90422.14	106716.8	**88.85**	**89.80**	**~89.3**	hetero-trimer (ApN-2S)$_2$:CM
CM-H$_{10}$-SIIApN-2S		77229.68	94909.79	112589.9	**93.46**	**93.25**	**~93.4**	hetero-trimer (SIIApN-2S)$_2$:CM-H$_{10}$
CM^{Nt}-H$_{10}$-SIIApN-2S		40373.72	58053.83	75733.94	**57.97**	**56.50**	**~ 57.2**	hetero-trimer (SIIApN-2S)$_2$:CM^{Nt}-H$_{10}$

* SEC-MALS analysis was performed by Kun Zang.

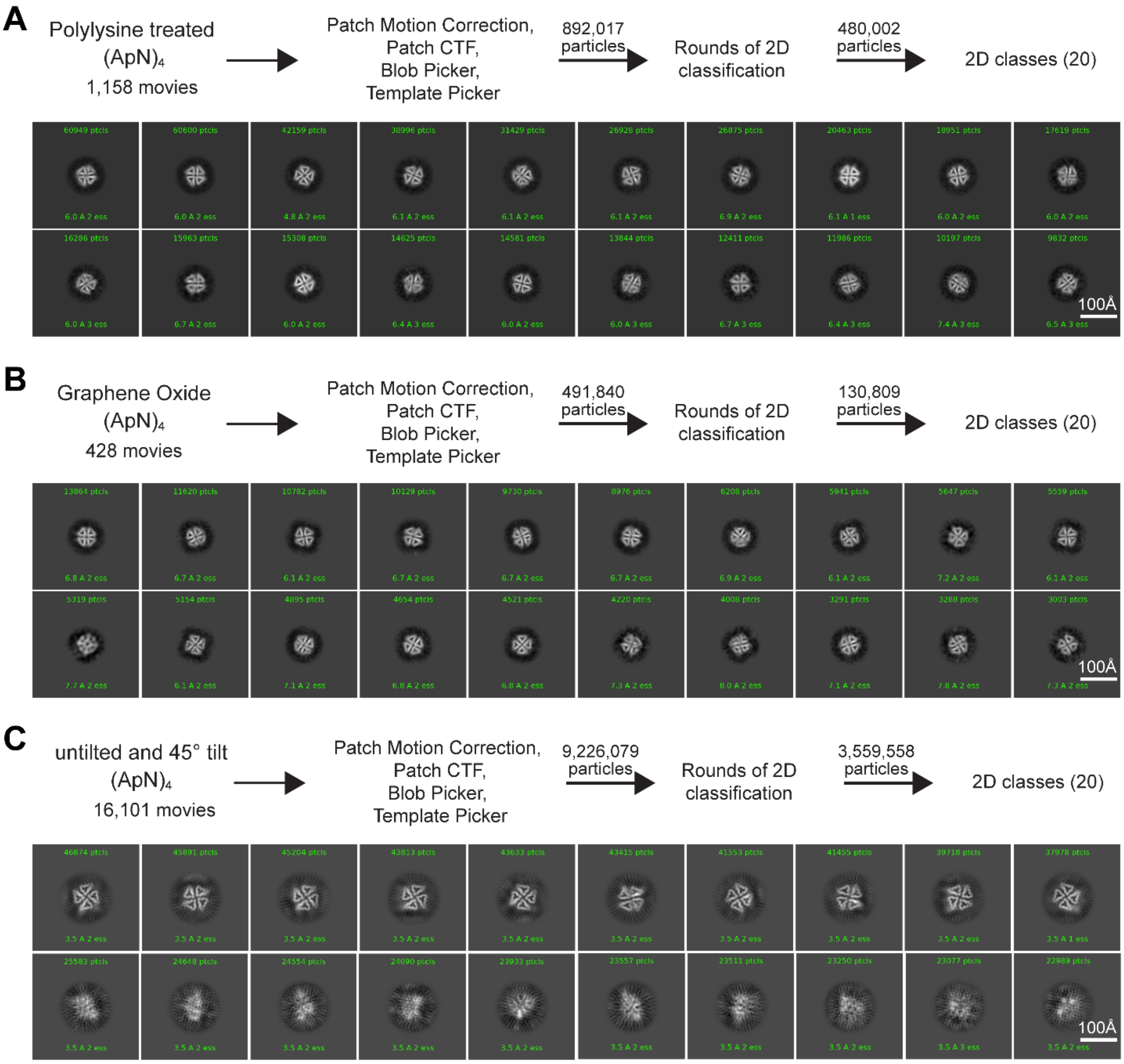
A.4. Cryo-EM single-particle analysis

In this chapter, I present the workflow of cryo-EM single-particle analysis, as well as the final top twenty classes of each protein. These include ApN, SIIApN, the CM-ApN hetero-complex, the (CM-H_{10})-SIIApN hetero-complex, the oxidized (CM-H_{10})-SIIApN hetero-complex, the CM-(ApN-2A) hetero-complex, the (CM-H_{10})-(SIIApN-2A) hetero-complex, the CM-(ApN-2S) hetero-complex, and the (CM-H_{10})-(SIIApN-2S) hetero-complex. I also discuss the attempts to obtain the structures of ApN and the (CM-H_{10})-SIIApN hetero-complex. None of these methods were successful. These refer to Chapters **3.1**, **3.4** and **3.5**.

A.4.1. Cryo-EM single-particle analysis of ApN and SIIApN

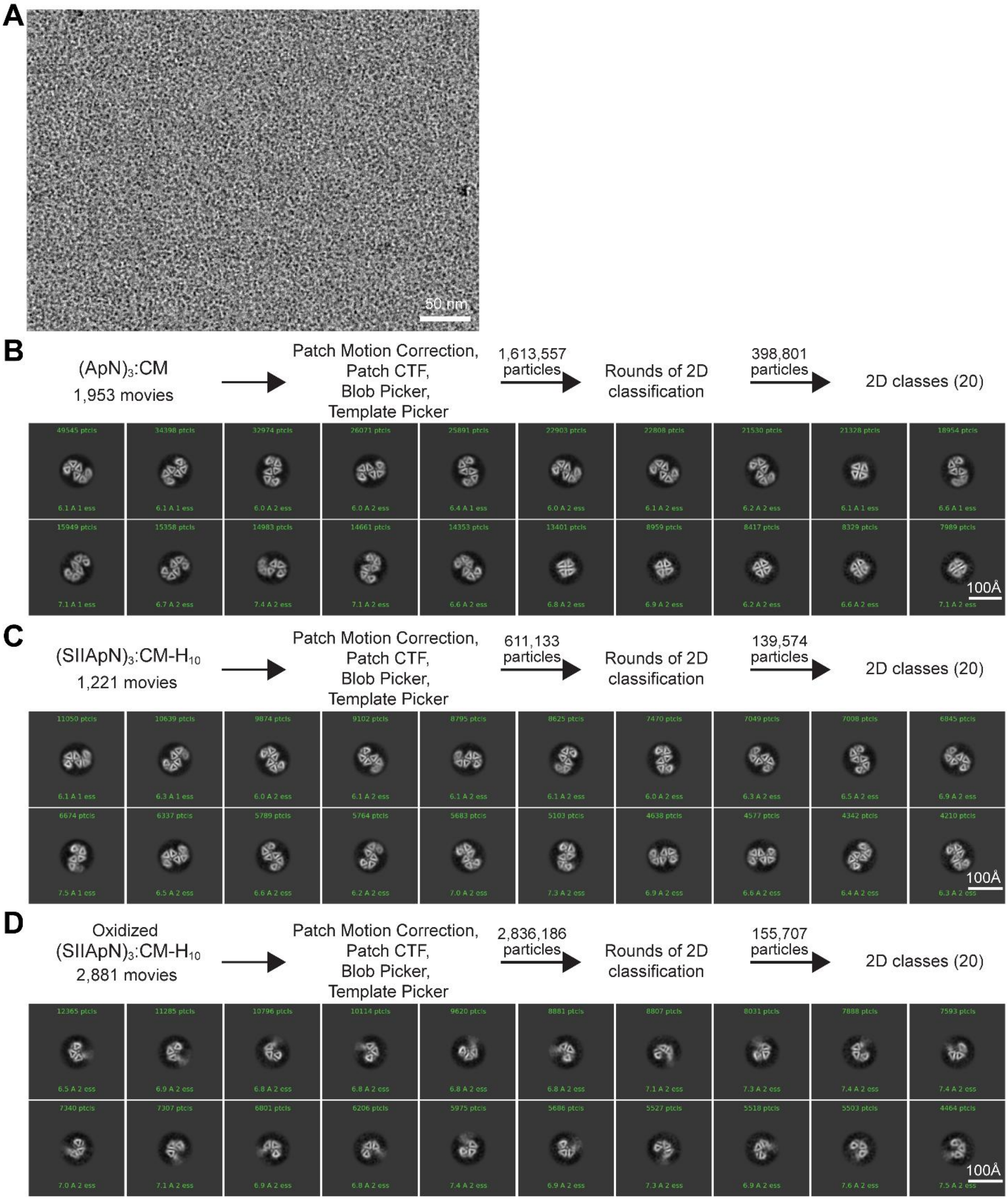


A.8 Cryo-EM single-particle analysis of ApN and SIIApN.
(A) A representative micrograph is shown. Scale bar, 50 nm. **(B-C)** Single-particle data processing workflow (top), and the top twenty 2D class averages (bottom) of ApN **(B)** and SIIApN **(C)** is shown. Particle numbers and resolution of each class are indicated. Scale bar, 100 Å.



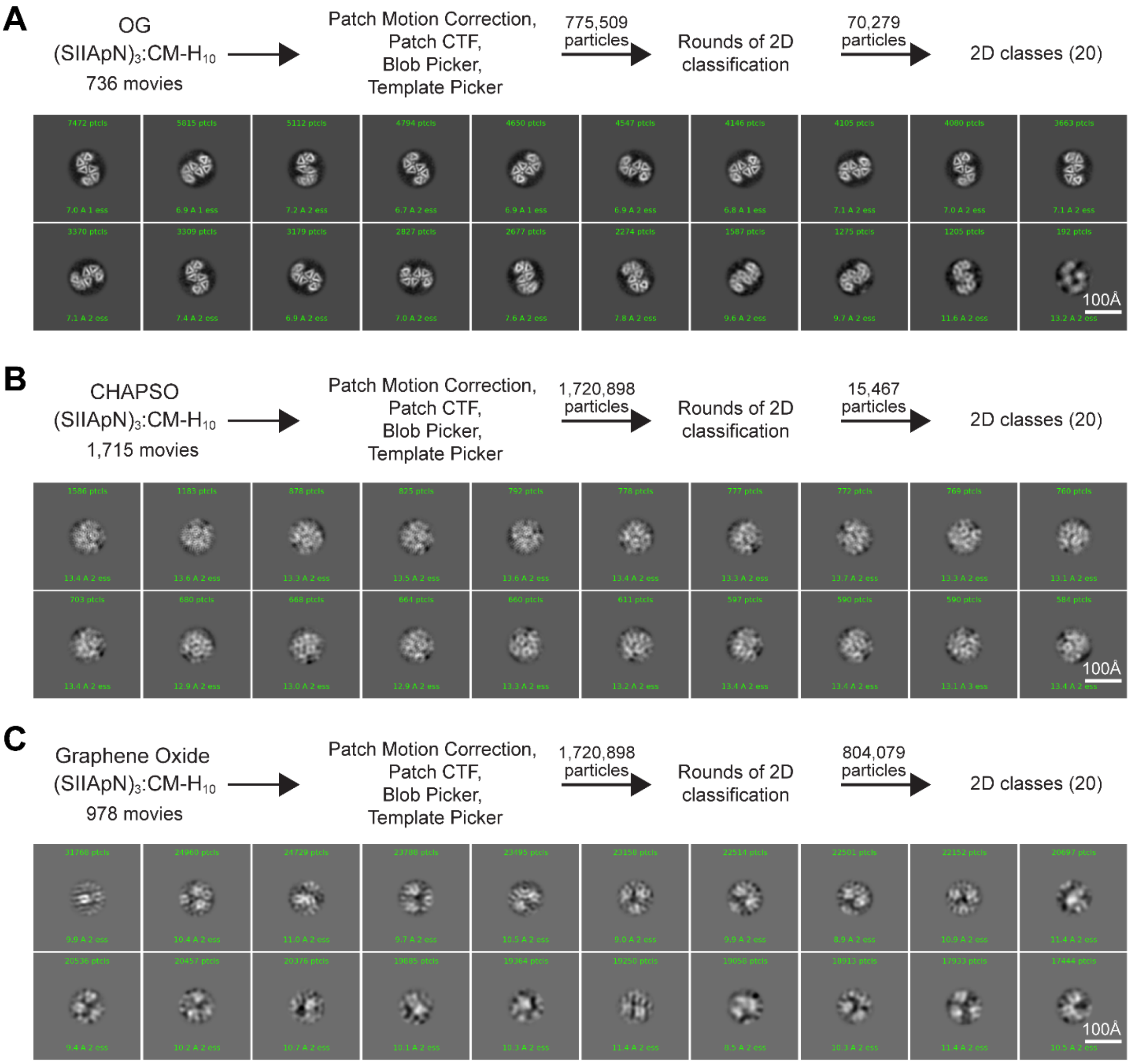
A.9 Attempts to obtain the structure of ApN.
(A-C) Single-particle data processing workflow (top), and the top twenty 2D class averages (bottom) of ApN on polylysine treated grids **(A)**, on graphene oxide-coated grids (B), and untilted and 45° tilt data collection **(C)** is shown. Particle numbers and resolution of each class are indicated. Scale bar, 100 Å.

A.4.2. Cryo-EM single-particle analysis of CM-ApN and (CM-H_{10})-SIIApN



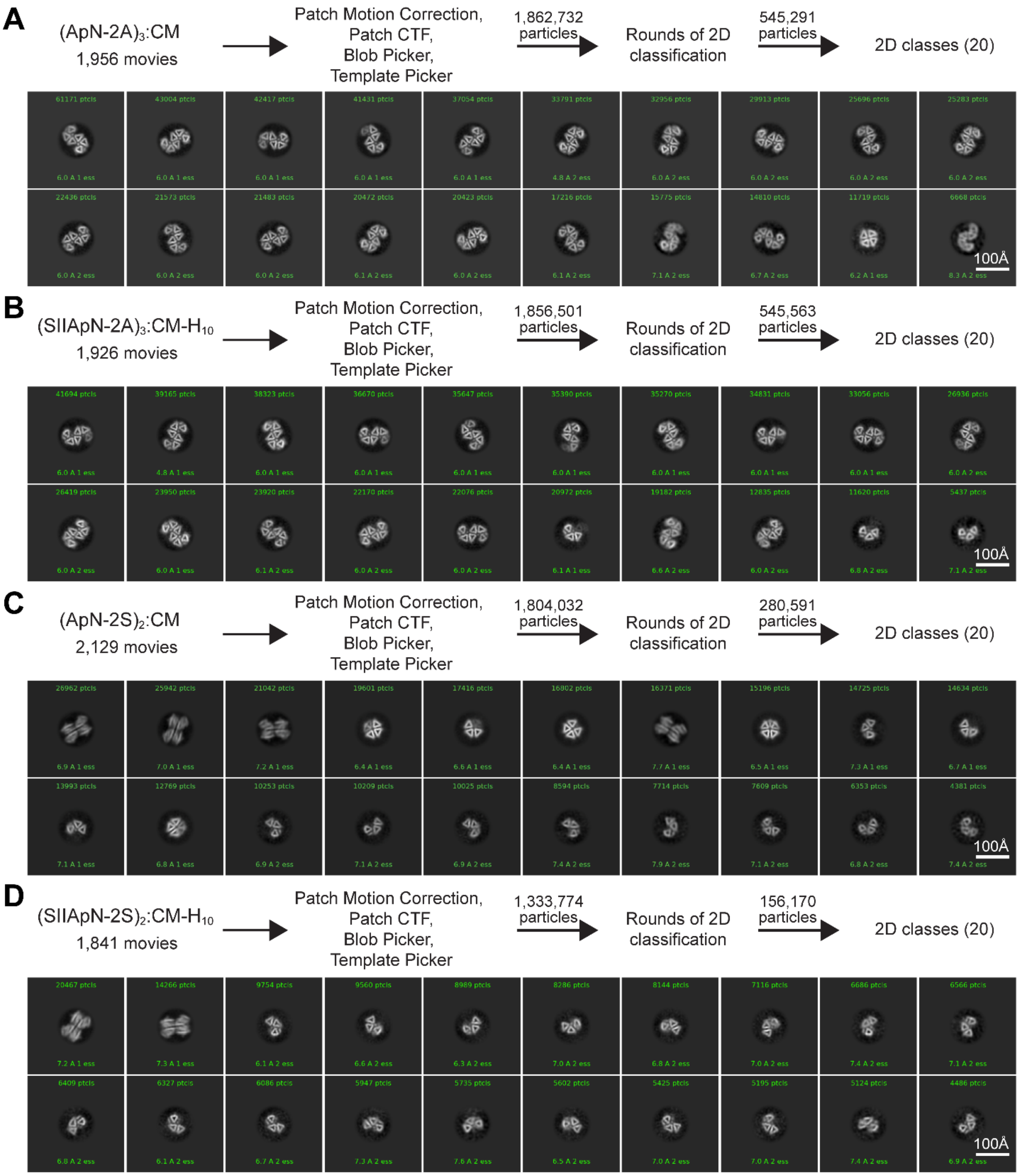
A.10 Cryo-EM single-particle analysis of CM-ApN and (CM-H_{10})-SIIApN.

(A) A representative micrograph is shown. Scale bar, 50 nm. **(B-D)** Single-particle data processing workflow (top), and the top twenty 2D class averages (bottom) of CM-ApN **(B)**, (CM-H_{10})-SIIApN **(C)**, and oxidized (CM-H_{10})-SIIApN **(D)** is shown. Particle numbers and resolution of each class are indicated. Scale bar, 100 Å.



A.11 Attempts to obtain the structure of (CM-H_{10})-SIIApN.
(A-C) Single-particle data processing workflow (top), and the top twenty 2D class averages (bottom) of (CM-H_{10})-SIIApN with the addition of OG **(A)**, the addition of CHAPSO **(B)**, and on graphene oxide-coated grids **(C)** is shown. Particle numbers and resolution of each class are indicated. Scale bar, 100 Å.

A.4.3. Cryo-EM single-particle analysis of CM-ApN and (CM-H$_{10}$)-SIIApN mutants



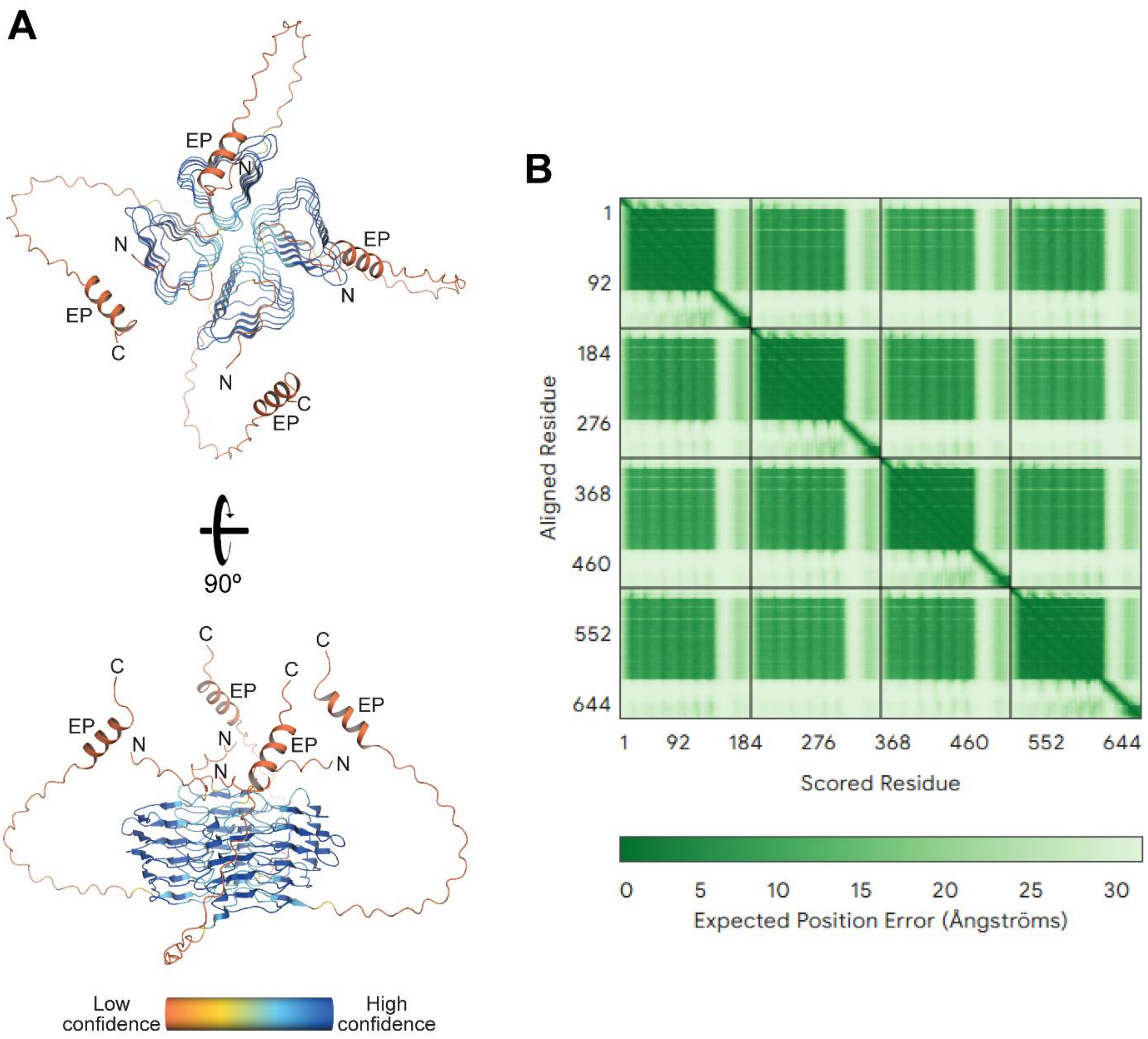
A.12 Cryo-EM single-particle analysis of CM-ApN and (CM-H$_{10}$)-SIIApN mutants. **(A-D)** Single-particle data processing workflow (top), and the top twenty 2D class averages (bottom) of CM-(ApN-2A) **(A)**, (CM-H$_{10}$)-(SIIApN-2A) **(B)**, CM-(ApN-2S) **(C)**, and (CM-

H_{10})-(SIIApN-2S) **(D)** is shown. Particle numbers and resolution of each class are indicated. Scale bar, 100 Å.

A.5. Protein structure prediction

The following are the structural models of (ApN)$_4$ (residues 1-161) and the ApN$_3$:CM (ApN, residues 1-161; CM, residues 1-198) hetero-tetrameric complex, respectively. These refer to Chapters **3.1** and **3.3**.

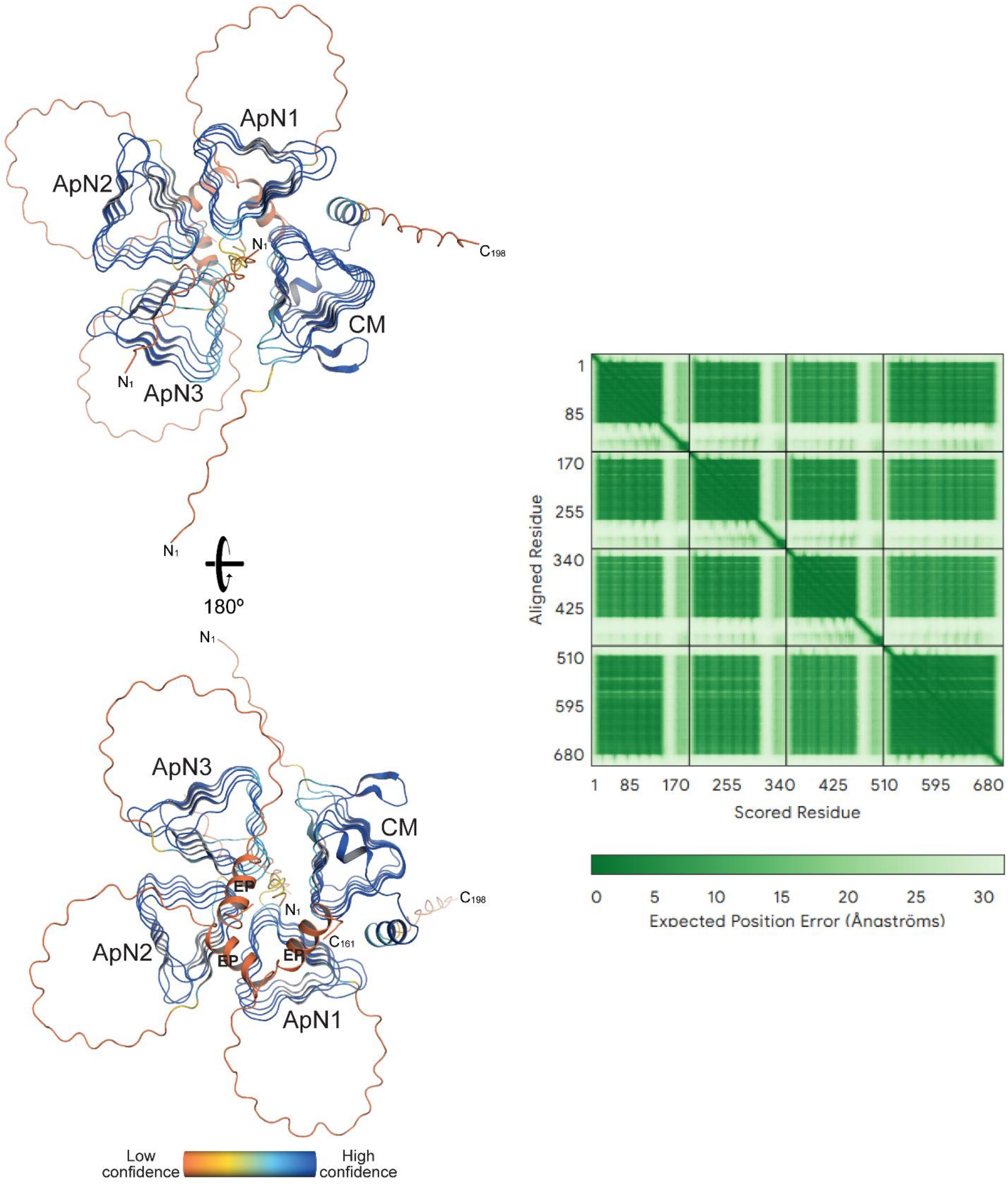
A.5.1. Predicted structural model of (ApN)$_4$



A.13 Predicted structure of full-length (ApN)$_4$ by AlphaFold 3.
(A) AlphaFold 3 structural model of (ApN)$_4$ (residues 1-161). End-on (left top) and side (left bottom) views of (ApN)$_4$ in ribbon representation with the N-termini on the top (clockwise orientation of wall A to wall C), highlighting the N- and C-termini and the encapsulation peptide (EP). The color bar indicates the confidence scores of the AlphaFold 3 model (pLDDT),

with high confidence (blue) and low confidence (orange) transitioning through the color spectrum. **(B)** 2D plot of the predicted aligned error (PAE) showing the confidence of the relative positions of different residues in the structure. X axis and Y axis present residue positions, and each cell (i, j) shows the expected position error (in Ångströms) at residue *i* when the predicted and true structures are aligned on residue *j*, with low error/high confidence (dark green) and high error/low confidence (light green) transitioning through the color spectrum.

A.5.2. Predicted structural model of (ApN)$_3$:CM



A.14 Predicted structure of (ApN)$_3$:CM hetero-complex by AlphaFold 3.
(A) AlphaFold 3 structural model of (ApN)$_3$:CM (ApN, residues 1-161; CM, residues 1-198). End-on views of (ApN)$_3$:CM in ribbon representation with the N-termini on the top (left top)

and C-termini on the top (left bottom), highlighting the N- and C-termini and the encapsulation peptide (EP) of ApN. The color bar indicates the confidence scores of the AlphaFold 3 model (pLDDT), with high confidence (blue) and low confidence (orange) transitioning through the color spectrum. **(B)** 2D plot of the predicted aligned error (PAE) showing the confidence of the relative positions of different residues in the structure. X axis and Y axis present residue positions, and each cell (i, j) shows the expected position error (in Ångströms) at residue *i* when the predicted and true structures are aligned on residue *j*, with low error/high confidence (dark green) and high error/low confidence (light green) transitioning through the color spectrum.