CONCURRENT AND CONVERGENT EVOLUTION OF A PROMISCUOUS MECHANISM TO PHASE SEPARATE EUKARYOTIC CARBON FIXATION

James Rowan Barrett

PhD

University of York
Biology
September 2023

ABSTRACT

Eukaryotic algae perform approximately one third of global CO_2 fixation in organelles called pyrenoids which supercharge fixation and are found across vast evolutionary taxa. Transplanting algal pyrenoids into pyrenoid-less crop plants to supercharge their CO_2 fixation offers a promising strategy to increase yields. It has long been known that algal pyrenoids are densely packed with the CO_2-fixing enzyme Rubisco. More recently it has emerged that Rubisco packaging and formation of the pyrenoid is underpinned by a dynamic process called liquid-liquid phase separation (LLPS) that is critical to pyrenoid function. Understanding of pyrenoid assembly and evolution outside of model species has been limited by the lack of sequence conservation of the proteins that drive LLPS of Rubisco across algal lineages, termed 'linker' proteins. To address this knowledge gap, a sequence-independent Fast Linker Identification Pipeline for Pyrenoids (FLIPPer) was developed to identify analogous linker proteins in unexplored pyrenoids. Assessment of the time-calibrated phylogenetic distribution of these linker proteins allowed identification of a shared timepoint for the convergent evolution of pyrenoids that coincided with plant-driven atmospheric oxygenation at the end of the Paleozoic Era. Detailed characterisation of a linker protein identified in the green microalga *Chlorella sorokiniana* (CsLinker) subsequently demonstrated its role as a *bona fide* pyrenoid linker protein. Structurally, it was shown that LLPS of *Chlorella* Rubisco is underpinned by multivalent binding of CsLinker to a highly conserved interface on the Rubisco large subunit which enables CsLinker to promiscuously phase separate other Rubiscos. This promiscuity enables CsLinker to functionally replace the native linker protein in the pyrenoid of *Chlamydomonas*, despite sharing no sequence homology with the native sequence. Excitingly, the promiscuity of CsLinker also extends to certain plant Rubiscos, where it demonstrated both *in vitro* and *in planta* Rubisco condensation. CsLinker therefore presents exciting future opportunities for engineering pyrenoids in crop plants.

DECLARATION

I hereby declare that this thesis is a presentation of original work. Unless otherwise stated at the relevant sections, I am the sole author. This work has not previously been presented for a degree or other qualification at this University of elsewhere. Some sections of this thesis are from accepted and submitted manuscripts, as stated within. All sources are acknowledged as references.

ACKNOWLEDGEMENTS

It has been a true honour to be afforded the opportunity to pursue a research direction which so deeply interests me; let alone with the unwavering support of my supervisors, colleagues, friends, and family. Above all, I am most grateful to my PhD supervisor Professor Luke Mackinder whose enthusiasm, patience, rigour, and ambition have inspired me throughout my studies – it is immeasurable the benefit these qualities bestowed upon me. The innumerable hours of mentorship and consideration Luke has provided since welcoming me as a student should not be overlooked. I extend my thanks also to Professor Mark Leake FInstP, FRMS, FRSB for his unwaning council and support despite changes in project direction.

The Mackinder lab, the Centre for Novel Agricultural Products (CNAP) and the University of York more generally have provided an ideal platform to complete my studies and foster my professional development. I am indebted to many great scientists and mentors that belong to each of these communities. I would like to extend particular thanks to Dr. Jamie Blaza for unquestionably supporting my endeavours into cryo-EM; a direction that was so central to the success of this study. I am also thankful to the great scientists and advisors that served on my advisory panel throughout my studies: Professor Tom McLeish FRS FRSC FInstP and Dr Michael Plevin; thank you for your invaluable perspectives on my studies. Within the lab, the professional, experimental, and personal support of Dr Philipp Girr and Dr Charlotte Walker have been vital to my progress, and I am forever grateful for their support. I am also grateful to have benefitted from the expertise of Dr Mihris Naduthodi, Dr Katrin Beßer, Dr Onyou Nam and Dr Liat Adler. I also thank and wish my fellow Mackinder lab PhD students Chun Sing Lau and Sabina Musiał well with the remainder of their studies. The support and technical staff at York are also owed my thanks for so generously providing their time and expertise: Dr Grant Calder, Dr Adam Dowle, Dr Andrew Leech, Dr Johan Turkenberg and Sam Hart.

Outside of York I also extend my thanks to René Inckemann for his advice on chloroplast expression in *Chlamydomonas*, Dr Oliver Caspari for his advice on expression of RbcSs, Professor Alistair McCormick and Yuwei Mao for their enthusiasm and willingness to collaborate on plant expression work. More generally, I thank the Combining Algal and Plant Photosynthesis (CAPP) consortium, and the wider CCM field for welcoming junior researchers such as myself so openly into their communities.

Finally, I extend my utmost gratitude to my family, partner, and friends. My parents Dr Sara Barrett and Philip Barrett have always inspired and supported me, and I would not be where I am without them. My mother's interest in everything from minute to global undoubtedly fostered my interest in this field. My brother Matthew and my close friends Liam, Conor and Christian have been unwaveringly supportive throughout; thank you. And to my dear partner Carys, thank you for everything.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

LIST OF TABLES

Chapter 2

Chapter 3

Chapter 4

Chapter 5

LIST OF ABBREVIATIONS

α3	Complete repeat fragment centred on Helix 3 of CsLinker
α3-α4	Complete repeat fragment centred on Helices 3 and 4 of CsLinker
2-PG:	2-phosphoglycolate ($C_2H_2O_6P^{3-}$)
3-PGA	3-phosphoglycerate ($C_3H_7O_7P$)
aadA	aminoglycoside-3"-adenylyltransferase gene
ATP	Adenosine triphosphate ($C_{10}H_{16}N_5O_{13}P_3$)
BLASTp	Basic Local Alignment Search Tool protein-protein
CAH	Carbonic anhydrase
CAM	Crassulacean Acid Metabolism
CBB	Calvin-Benson-Bassham
CBM20	Carbohydrate Binding Module 20
CCM	Carbon Concentrating Mechanism
CI_{95}	95% Confidence Interval
co-IP	co-Immunoprecipitation
CO_2	Carbon dioxide
cryo-EM	Cryogenic Electron Microscopy
cryo-ET	Cryo-Electron Tomography
CsLinker	*Chlorella sorokiniana* UTEX1230 Linker
CTF	Contrast Transfer Function
cTP	Chloroplast Transit Peptide
EPYC1	Essential PYrenoid Component 1
FLIPPer	Fast Linker Identification Pipeline for Pyrenoids
Fmol	Femtomoles
FRAP	Fluorescence Recovery After Photobleaching
FSC	Fourier Shell Correlation
G3P	Glyceraldehyde-3-Phosphate ($C_3H_7O_6P$)
Gya	Billion years ago
HCO_3^-	Bicarbonate ion
kDa	Kilodalton
LC-MS/MS	Liquid Chromatography with tandem Mass Spectrometry
LLPS	Liquid-Liquid Phase Separation
LPIA	Late Paleozoic Ice Age
MBP	Maltose Binding Protein
mEGFP	monomeric Enhanced Green Fluorescent Protein
MSA	Multiple Sequence Alignment
My	Million years
Mya	Million years ago
NADPH	Nicotinamide Adenine Dinucleotide PHosphate
NCBI	National Center for Biotechnology Information
NTD	N-Terminal Domain
O_2	Single oxygen
PAE	Predicted Aligned Error
PDB	Protein Data Bank
PEPC	Phosphoenolpyruvate carboxylase
pI	Isoelectric point
pLDDT	Predicted Local-Distance Difference Test
PMST	Pyrenoid Matrix Starch Tether
PMTT	Pyrenoid Matrix Thylakoid Tether
PYCO1	PYrenoid COmponent 1
QFDEEM	Quick Freeze Deep-Etch Electron Microscopy
RbcL	Rubisco Large subunit
RbcS	Rubisco Small subunit

RBM	Rubisco-Binding Motif
RBMP	Rubisco-Binding Membrane Protein
Rubisco	Ribulose-1,5-bisphosphate carboxylase/oxygenase
RuBP	Ribulose-1,5-bisphosphate
S.D.	Standard Deviation
S.E.M.	Standard Error of the Mean
SAGA1	StArch Granules Abnormal 1
SDM	Site-Directed Mutagenesis
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
SPR	Surface Plasmon Resonance
TEM	Transmission Electron Microscope
TEV	Tobacco Etch Virus Protease
TSAR	Telonemia, Stramenopila, Alveolata, Rhizaria
UTC	Ulvophyceae, Trebouxiophyceae, Chlorophyceae
YFP	Yellow Fluorescent Protein

1. INTRODUCTION

1.1. Declaration

The purpose of this chapter is to provide the reader with an overview of the literature base that was interpreted in the formation of hypotheses that provided the framework for Chapters 2-5. The majority of writing within this section is consistent with the review "Pyrenoids: CO_2-fixing phase separated liquid organelles" that was published in April 2021 in Biochimica et Biophysica Acta (BBA)- Molecular Cell Research as part of the special issue "State without borders: Membrane-less organelles and liquid–liquid phase transitions". Where appropriate, sections have been updated considering new literature.

1.2. Chapter overview

As this introductory chapter encompasses a wide and extensive literature base and in places does not follow a linear digestibility, a brief explanation of the overall structure of the chapter is provided here for clarity. In section 1.3 the basis of Rubisco-mediated carbon fixation is described, along with the drawbacks of Rubisco and how these are compensated in different photosynthetic organisms. From this section emerges the commonality of Rubisco condensation by liquid-liquid phase separation (LLPS) in one of these approaches – biophysical CCMs. As such, section 1.4 follows by describing the general theory, importance and understanding of LLPS in biological systems to provide basis for discussion of LLPS in the context of pyrenoids subsequently. Section 1.5 demonstrates the breadth of pyrenoids in eukaryotic algae and summarises the discovery of pyrenoids, and significant milestones in their characterisation. Given that the majority of our understanding of pyrenoid-based CCMs, pyrenoid assembly, Rubisco phase separation and the liquid-like properties of pyrenoids comes from the wealth of *Chlamydomonas* pyrenoid literature, a deep-dive on the current state of knowledge is provided in section 1.6. Section 1.7 follows the description of the key LLPS indicators in the *Chlamydomonas* pyrenoid, by extending the qualitative observations to those of pyrenoids in diverse algae to also rationale their formation by LLPS. The current state of understanding of pyrenoid evolution and structural diversity is outlined in section 1.8, and how this might relate to ongoing pyrenoid engineering efforts in plants in section 1.9 is also introduced.

1.3. Carbon fixation and concentration

1.3.1. Photosynthesis, photorespiration, and Rubisco's bottleneck

Photosynthesis is the gateway between inorganic carbon (*i.e.,* CO_2) and organic carbon in the global carbon cycle. It harnesses energy from sunlight to annually reduce ~400 gigatonnes of CO_2 (Net Primary Production; (Field *et al.,* 1998)) whilst simultaneously releasing O_2. Photosynthesis is broadly split into two reaction schemes: the light and dark reactions. During the light reactions, light energy is converted into biochemical energy in the form of adenosine triose phosphate (ATP) and nicotinamide adenine dinucleotide phosphate (NADPH) by the action of a series of light-dependent reactions at the thylakoid membranes within the plastids of photosynthetic organisms. The high-energy compounds NADPH and ATP that are products of the light reactions subsequently power the dark reactions which exist as part of the Calvin-Benson-Bassham (CBB) cycle and are the entry point for CO_2 in the organic carbon cycle (Fig. 1-1). The CBB cycle revolves around the five-carbon compound ribulose-1,5-bisphosphate (RuBP) and can be broadly divided into 3 phases: carboxylation, reduction and regeneration (Benson *et al.,* 1950). In the carboxylation reaction, CO_2 is covalently joined with RuBP by the enzyme Ribulose-1,5-bisphosphate carboxylase oxygenase (Rubisco) to produce a short-lived six-carbon intermediate which rapidly decomposes into two molecules of 3-phosphoglycerate (3-PGA). In the reductive phase, ATP and NADPH are utilised to reduce 3-PGA into glyceraldehyde-3-phosphate (G3P), which is the 3 carbon compound that gives rise to organic compounds and is the result of so-called 'C_3 photosynthesis'. G3P that is not used for organic compounds by the organism is regenerated back to RuBP through a ribulose-5-phosphate intermediate at the expense of further ATP. 3 completions of the CBB cycle are required for the net production of a G3P molecule at a cost of 9 ATP and 6 NADPH molecules per G3P produced. Virtually all global carboxylation of RuBP with CO_2 is completed by Rubisco, making it the most abundant and geochemically important enzyme on earth (Bar-On & Milo, 2019; Raven, 2013). However, carboxylation of RuBP is also typically the rate-limiting factor in the CBB cycle due to the relatively slow catalytic rate of Rubisco compared to other central metabolic enzymes (typically 8-10 times lower than the median value (Bar-Even *et al.,* 2011). The slow speed of Rubisco contributes one of its two primary drawbacks.

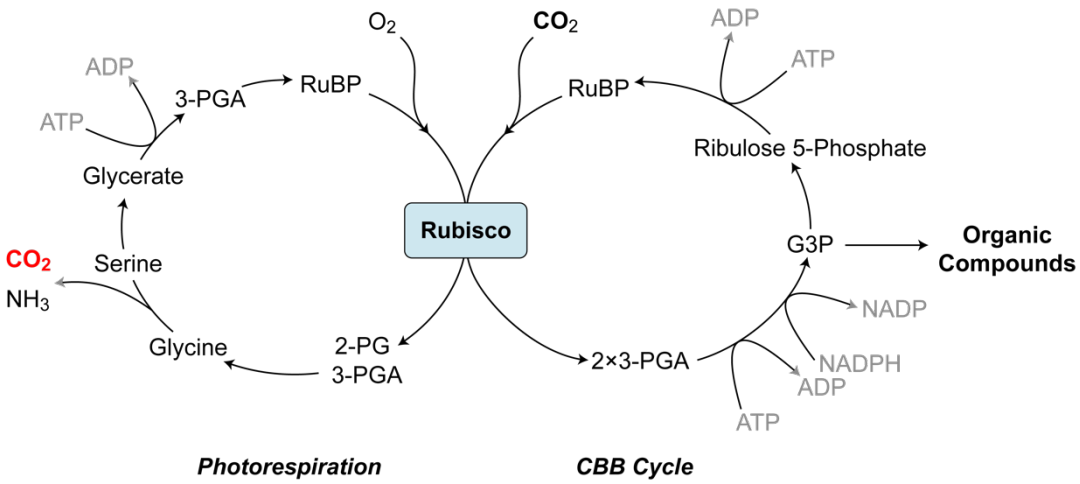


Figure 1-1 | The CBB and photorespiratory cycles.
Simplified schematic diagrams of the Calvin-Benson-Bassham (CBB) and photorespiratory cycles. Exogenous biochemical inputs are shown in grey. The schematic is based on an original image by Wikimedia Commons user Rachel Purdon and is adapted and distributed under the Creative Commons Attribution-Share Alike 3.0 Unported license.

The other primary drawback of Rubisco concerns the specificity for its substrate. In addition to the carboxylation of RuBP using CO_2, Rubisco can also catalyse the oxygenation of RuBP using O_2 as a substrate, for which Rubisco possesses a similar catalytic rate (k_{cat} ~2-3 s^{-1} in plants) (Flamholz *et al.,* 2019). The poor substrate specificity of Rubisco has been proposed to be due to difficulty in distinguishing between the similar size, shape and electrostatic potential of CO_2 and O_2 molecules in the active site (Tcherkez *et al.,* 2006). The oxygenation reaction of Rubisco results in the production of one 3-PGA molecule in addition to a 2-phosphoglycolate (2-PG) molecule, consistent with the decomposition of an oxygenated 5-carbon intermediate formed from the oxygenation of RuBP (Fig. 1-1). In contrast to 3-PGA, 2-PG cannot be used in the CBB cycle and must be scavenged back to 3-PGA through an energetically expensive process known as photorespiration (Ogren, 1984). Not only does the fixation of an O_2 molecule prevent fixation of a CO_2 at the same site, but the photorespiratory pathway wastes another molecule of CO_2 that had previously been fixed by the CBB cycle (Fig. 1-1), resulting in an effective loss of two CO_2 molecules per O_2 fixed. The respiratory short-fall of Rubisco has been accentuated by the vastly different composition of the present-day atmosphere, as compared to when Rubisco evolved in the Archean Eon (2,500-4,000 million years ago (Mya)) (Shih *et al.,* 2016). Estimates suggest CO_2 was historically in vast excess over O_2, which was almost non-existent when

Rubisco evolved (Berner, 2009; Berner & Kothavala, 2001). The atmospheric composition changes that occurred over the subsequent ~2,500 My to the present day not only reduced the absolute availability of CO_2 as Rubisco's primary substrate, but also flooded Rubisco with O_2, resulting in compounded reduced carboxylation due to both availability and specificity challenges.

Given that nearly all carbon fixation is performed by Rubisco and it is one the earliest evolved enzymes, it is puzzling that over its >2.5 billion year existence it has remained generally slow and has a relatively poor selectivity for CO_2 over O_2 under current atmospheric concentrations (Whitney *et al.,* 2011). Arguments have been made previously that these characteristics are due to an evolutionary trade-off between Rubisco's catalytic rate and its specificity for CO_2 over O_2 (Flamholz *et al.,* 2019; Studer *et al.,* 2014; Tcherkez *et al.,* 2006). This hypothesis effectively argues that Rubiscos have evolved to balance being either; i) slower and more specific (*e.g.,* in red algae) or ii) faster and less specific (*e.g.,* in cyanobacteria) (Fig. 1-2). More recent arguments have suggested this trade-off is driven more by the phylogenetic constraints of the sequences (Bouvier *et al.,* 2021) and that Rubisco itself is still evolving to optimise both speed and selectivity, at least in C_3 plants (Bouvier *et al.,* 2023). What is clear though is that Rubiscos across all phylogenetic taxa are amongst the slowest evolving proteins on earth, with one amino acid substitution occurring on average every 7.2 My in the catalytic domain (Bouvier *et al.,* 2023). This suggests that changes in atmospheric conditions have effectively occurred faster than Rubisco could evolutionarily respond, and that Rubisco is slowly catching up with the challenges presented by present day atmospheric conditions. The inefficiencies bestowed upon Rubisco by its slow evolution of speed and selectivity traits result in it contributing the bottleneck to photosynthesis in most cases and contributing to the sub-optimal operation of carbon fixation in many organisms.

1.3.2. Rubisco form and function

Plants, cyanobacteria, most proteobacteria and most algae possess form I Rubisco, which consists of a hexadecameric complex comprised of an octameric core of four Rubisco large subunit dimers ($RbcL_8$) capped by two sets of four Rubisco small subunits ($RbcS_8$) that form the holoenzyme (L_8S_8) (Fig. 1-2). Within each RbcL subunit lies an active site that catalyses the carboxylation/oxygenation of RuBP. Form I Rubiscos can be further subdivided into Form IA and IB, that comprise the 'green' Rubiscos of all plants and green algae and some proteobacteria and cyanobacteria, and form IC and ID that comprise the 'red' Rubiscos of proteobacteria and non-green algae. The overall structure of form I Rubiscos is highly conserved (Fig. 1-2), though sub-form differences do exist. Other Rubisco forms (I', II, III and IV) exist (Tabita *et al.,* 2008), but are vastly less represented in phototrophs and not a focus of this study.

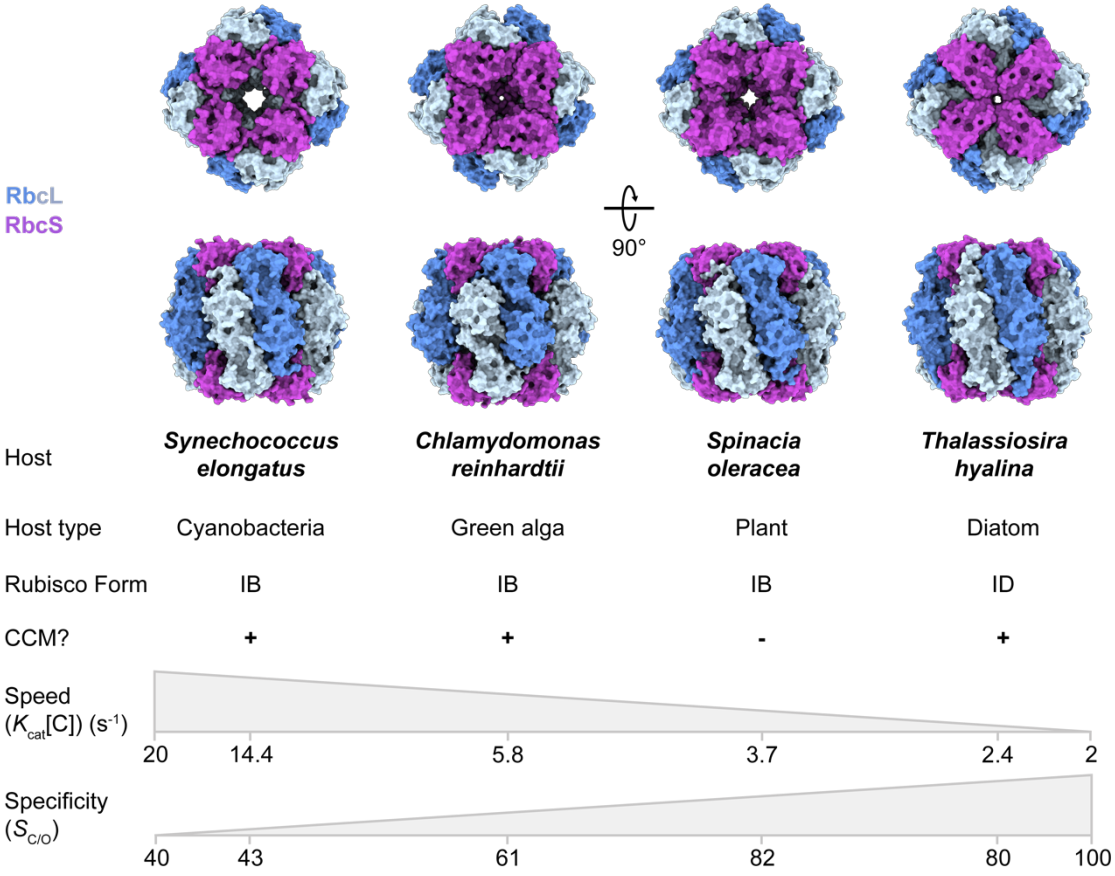


Figure 1-2 | Form I Rubisco variation.
Surface representations of Form I Rubisco structures from *Synechococcus elongatus* (PDB: 1RBL; (Newman *et al.,* 1993)), *Chlamydomonas reinhardtii* (PDB: 7JFO; (He *et al.,* 2020)), *Spinacea oleracea* (PDB: 8RUC; (Andersson, 1996)) and *Thalassiosira hyalina* (PDB: 5N9Z; (Valegård *et al.,* 2018)). Subunits of the structures are coloured as indicated. The trend in general properties of the enzyme are indicated below each structure, though the positioning is not exact. Reported values are taken from (Ang *et al.,* 2022) and (Young *et al.,* 2016).

1.3.3. Carbon concentrating mechanisms accelerate photosynthesis

Photosynthetic organisms have evolved diverse strategies to overcome Rubisco's inefficiencies and maximise carbon fixation. Many terrestrial plants that operate C_3 photosynthesis attain high CO_2 fixation and reduce photorespiration by investing large amounts of resources into Rubisco that is catalytically slow but has a relatively high specificity for CO_2 over O_2 (*e.g.,* Spinach (*Spinacia oleracea*) – Fig. 1-2) (Flamholz *et al.,* 2019). This results in Rubisco typically accounting for approximately 25% of soluble protein in plant leaves (Galmés *et al.,* 2014), accounting for its title as most abundant protein on earth (Ellis, 1979; Raven, 2013). C_3 plants pay huge energy, nitrogen, and water usage costs to produce such vast amounts of Rubisco, yet are still mostly limited by the bottleneck of Rubisco in the CBB cycle in agricultural environments (Sharkey, 1988), which translates to real-world reductions in crop yields that threaten future food security (Walker *et al.,* 2016).

An alternative strategy to increase carbon fixation evolved by some plants (non-C_3) and nearly all aquatic photosynthetic organisms is to operate carbon concentrating mechanisms (CCMs). CCMs function to actively increase the concentration of free CO_2, and in turn the CO_2:O_2 ratio, at the active site of Rubisco, thus enabling high CO_2 fixation rates by reducing the amount of photorespiratory wastage. Generally, Rubiscos that operate within CCMs are faster and less specific than their non-CCM counterparts, though exceptions to this trend do exist (*e.g.,* diatom CCMs) (Fig. 1-2). The operation of fast non-specific Rubisco in a high CO_2:O_2 environment affords the host organisms significant advantages in carbon fixation, meaning less energy, nitrogen, and water is invested to produce Rubisco, which consequently increases the growth rate of organisms operating CCMs. CCM is a broad term that covers mechanisms that can be broadly grouped into two types; those with a biochemical basis and those that are so-called 'biophysical'. The focus of this thesis is on biophysical CCMs, the dominant CCM type found in aquatic photosynthetic organisms, though biochemical CCMs are described here in brief for completeness and comparison.

1.3.3.1. Biochemical CCMs

Biochemical CCMs are most abundant in non-C_3 land plants, where they evolved convergently in independent lineages and operate on similar principles (Edwards & Ogburn, 2012). In most characterised biochemical CCMs, inorganic carbon in the form of bicarbonate (HCO_3^-) is fixed into a four carbon compound by the enzyme phosphoenolpyruvate carboxylase (PEPC), constituting the 'biochemical' phase of the concentrating mechanism. The resulting four carbon compound is shuttled to and subsequently decarboxylated in the vicinity of Rubisco to allow fixation of the released high concentration CO_2 in the CBB cycle (Ku *et al.,* 1996). Two main categories of biochemical CCMs exist: C_4 and Crassulacean Acid Metabolism (CAM). The more anciently evolved of the two, CAM (~30 Mya) (Sage *et al.,* 2023), relies on the temporal separation of carbon capture and Rubisco carboxylation and is found in plants such as cacti, aloe and pineapple (Nobel, 1991). In CAM plants, PEPC is active at night and accumulates a large pool of four carbon compounds that are stored in the vacuole. During the day the four carbon compounds are released, transported to the vicinity of Rubisco and decarboxylated to release CO_2 for fixation in the CBB cycle, powered by the light reactions (Ming *et al.,* 2015; Wai *et al.,* 2017). C_4 photosynthesis, which evolved more recently (~20 Mya) in maize, sugar cane and grasses (Kadereit *et al.,* 2003; Sage, 2004), operates on the principle of spatial rather than temporal separation of carbon capture. Here, specialized mesophyll cells within the leaf structure are the predominant site of carboxylation by PEPC, after which the resulting four carbon compounds are modified and shuttled to bundle sheath cells where Rubisco is localised. As with CAM, the four carbon compounds are then decarboxylated to release CO_2 for fixation by Rubisco in the CBB cycle (Hatch, 1987; Kanai & Edwards, 1999). Given the advantages afforded by C_4 photosynthesis in host plants, there is great interest in engineering C_4 photosynthesis in keystone C_3 crops such as rice (in the C_4 Rice Project), though these approaches require huge engineering efforts to alter cell type and metabolism and are consequently progressing slowly.

1.3.3.2. Biophysical CCMs

Biophysical CCMs operate similarly to C_4 photosynthesis in that carbon fixation is spatially segregated. However, instead of concentrating four carbon compounds that have been biochemically accumulated, in biophysical CCMs, inorganic carbon is directly concentrated, and CO_2 is released in the locale of Rubisco that is itself spatially segregated. Biophysical CCMs typically function via the concentration of inorganic carbon in the form of bicarbonate (HCO_3^-), given its abundance in aqueous environments that are commonly home to organisms operating biophysical CCMs. Generally, the functionality of biophysical CCMs is achieved through out-of-equilibrium carbonate chemistry, pH changes across membranes and the heterogeneous distribution of carbonic anhydrases (Fig. 1-3c) (Flamholz & Shih, 2020; Karlsson *et al.,* 1998; Mangan *et al.,* 2016; Price & Badger, 1989). pH determines the HCO_3^-:CO_2 ratio, with HCO_3^- ~100 times more abundant than CO_2 at pH 8. That being said, the equilibrium between HCO_3^- and CO_2 is exceptionally slow in the absence of catalytic carbonic anhydrase (Schulz *et al.,* 2006; Soli & Byrne, 2002), which enables the accumulation of HCO_3^- in conditions where it is out-of-equilibrium (low pH). Charged HCO_3^- ions also have much lower membrane permeability than CO_2, which further enables the accumulation of HCO_3^- in membrane-bound compartments (*e.g.,* the cytosol of cyanobacteria or the thylakoid lumen of algae). The spatial segregation of carbonic anhydrases in a pH environment that favours CO_2 in the equilibrium (pH < 7) completes the mechanism, as HCO_3^- can be rapidly converted to CO_2, diffuse across membranes where necessary, and be fixed by Rubisco (Fig. 1-3a,b).

Biophysical CCMs found in oxygenic phototrophs can be generally split into two types: carboxysome-based CCMs found in prokaryotic cyanobacteria, and pyrenoid-based CCMs found in eukaryotic algae and some non-vascular plants (most hornwort species). Cyanobacterial carboxysomes are icosahedral 100+ megadalton protein assemblies where densely packaged Rubisco is encapsulated in a protein shell with a typical diameter of 150-200 nm (Rae *et al.,* 2013) (Fig. 1-3a). Once formed, the carboxysome structure is largely stable in the cell. HCO_3^- concentrated in the cyanobacterial cytosol diffuses into the carboxysome through hexameric shell proteins where it is rapidly dehydrated to CO_2 via carbonic anhydrases for fixation by Rubisco (Fig. 1-3a). In contrast to the stroma of chloroplasts which are

universally pH ~8 (Fig. 1-3b), the cytosol of cyanobacteria is closer to pH 7 (Belkin *et al.,* 1987). Crucially then, considering the CO_2:HCO_3^- equilibrium is shifted towards CO_2 at pH 7, the cyanobacterial cytosol is devoid of carbonic anhydrases to allow the accumulation of out-of-equilibrium HCO_3^- (Price & Badger, 1989).This arrangement also allows the specific release of CO_2 in the carboxysome, which is the only region of the cell populated with carbonic anhydrases (Mangan & Brenner, 2014). As with C_4 photosynthesis, there is great interest in engineering these mechanisms in C_3 plants (Hanson *et al.,* 2016). Though these engineering efforts have been somewhat successful in the construction of carboxysomes that are able to support photosynthetic growth (Chen, Hojka, *et al.,* 2023), significant challenges face the future engineering approaches. Previous work has shown that the presence of a carbonic anhydrase outside of the carboxysome can short circuit the cyanobacterial CCM (Price & Badger, 1989), and it is therefore assumed that the presence of carbonic anhydrases in the chloroplast of host plants will need to ablated. Clearly this represents a huge engineering effort for each of the host plants in which carboxysome engineering is attempted, and recent results demonstrate a hugely detrimental effect on plant development from doing so in C_3 plants (Hines *et al.,* 2021)

Like carboxysomes, pyrenoid-based biophysical CCMs are also centred around spatially segregated Rubisco assemblies, but they are much larger (~1-2 µm diameter), lack a proteinaceous shell, and are typically traversed by membranes that are continuous with the thylakoid network that operates the light reactions of photosynthesis (Fig. 1-3b) (Meyer, Goudet, *et al.,* 2020). Their non-encapsulated form means they are also dynamic in size by growing and shrinking in response to CO_2 and light (see section 1.6). Algal pyrenoid-based CCMs operate on largely similar principles to those of carboxysomes, with some key differences. Instead of concentrating HCO_3^- in the cytosol and allowing diffusion into the spatially segregated Rubisco, HCO_3^- is first shuttled from the extracellular space through a series of channels into the acidic thylakoid lumen (pH ~5.5 in light) that is continuous with thylakoid material that traverses the pyrenoid (Fig. 1-3b). There, the specific localisation of a carbonic anhydrase in the region of thylakoid that traverses the pyrenoid allows rapid dehydration and release of CO_2 within the pyrenoid, creating a “point source” for delivery to Rubisco (Fig. 1-3b,c). Although in green plastids the

pyrenoid is not encapsulated in a protein shell to prevent CO_2 escape, most pyrenoids are surrounded by some kind of material that is proposed to function as a diffusion barrier (see section 1.8).

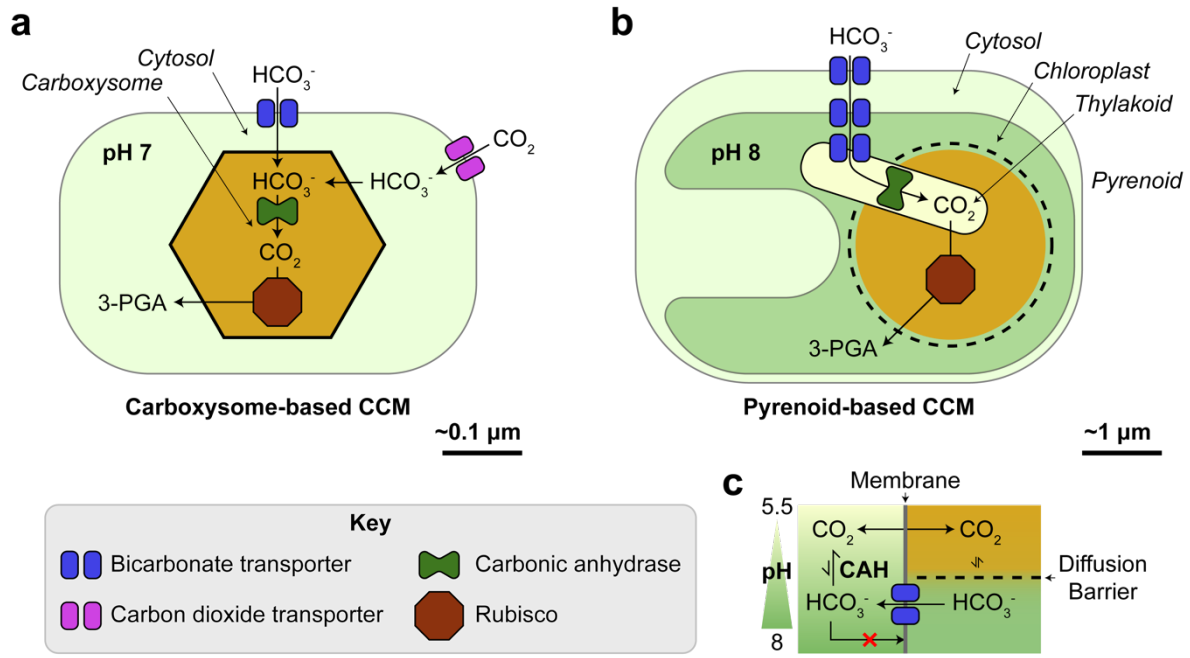


Figure 1-3 | Biophysical CCMs.
(a) Schematic representation of the general principles of a carboxysome-based CCM. The carboxysome is encapsulated by a protein shell (black line) that aids in the operation of the CCM by preventing efflux of CO_2. (b) Schematic representation of the general principles of a pyrenoid-based CCM. Most pyrenoids are encapsulated with some kind of diffusion barrier, represented by the dotted black line. (c) Carbonate chemistry across membranes is dependent on pH and carbonic anhydrase (CAH) presence. The specific scenario presented is reflective of the current understanding of pyrenoid-based CCMs.

Though the architecture of the mechanisms supplying CO_2 to Rubisco differs between carboxysomes and pyrenoids, common and essential to both mechanisms is the dense packing of Rubisco in a spatially segregated region that can be supplied with high concentration CO_2 (orange-brown in Fig. 1-3). The measured concentration of the packed Rubisco in carboxysomes (~800 µM; (Metskas *et al.,* 2022)) and pyrenoids (~650 µM; (Freeman Rosenzweig *et al.,* 2017)), far exceeds that of the calculated protein concentration of the bacterial (Zimmerman & Trach, 1991) and eukaryotic (Zeskind *et al.,* 2007) cytosolic spaces, which indicated an active mechanism for the concentration of Rubisco in these spatially segregated region.

1.3.4. Rubisco condensation is driven by disordered, multivalent Rubisco binding proteins

In recent years, it has become clear that the dense packing of Rubisco in pyrenoids and carboxysomes is underpinned by interaction with largely disordered, multivalent, Rubisco-binding proteins, termed 'linkers', in a process termed 'biomolecular condensation'. Although there is no sequence homology between the so-called linkers, structural analogy exists between them as discussed here. Expanding on the above description, there are two types of carboxysome, α and β, that appear to have evolved independently (Rae *et al.,* 2013). Nearly all carboxysome components have counterparts across α and β types, including linkers that display structural analogy (Fig. 1-3). In the α-carboxysome, the N-terminal domain (NTD) of CsoS2 multivalently binds Rubisco using 4 repeats of a helical NTD motif that are spaced by regions of disorder (CsoS2-NTD – Fig. 1-3) (Oltrogge *et al.,* 2020). Similarly, in β-carboxysomes the CcmM35 region of the CcmM protein performs an analogous role, binding Rubisco multivalently using 3 spaced repeats of an RbcS-like domain (Wang *et al.,* 2019). *In vitro,* spontaneous demixing into droplets densely packed with Rubisco occurs when the regions of CsoS2 or CcmM containing the multivalent Rubisco interacting domains (as displayed in Fig. 1-3), are mixed with their cognate Rubiscos (Oltrogge *et al.,* 2020; Wang *et al.,* 2019), indicating condensation is mediated by liquid-liquid phase separation (LLPS). In both cases, deletion of the full length linker proteins *in vivo* abrogates carboxysome assembly, leading to a high-CO_2 requiring phenotype – the characteristic signature of a non-functional CCM (Cai *et al.,* 2015; Ludwig *et al.,* 2000). Accordingly, it is postulated that Rubisco condensation by carboxysomal linker proteins plays a key role in assembly of both carboxysome types, and that this process is likely underpinned by LLPS.

In contrast to the two evolutionary origins of carboxysomes, and the clear conservation of carboxysome components across species (Kerfeld & Melnicki, 2016), pyrenoid evolution appears to be far more complex, owing to the complex phylogeny of algae (see sections 1.5, 1.7 and 1.8). Accordingly, there is a distinct absence of conserved structural components across diverse algal lineages (Mackinder *et al.,* 2016; Meyer, Goudet, *et al.,* 2020). Nearly all of our data on Rubisco condensation in pyrenoids is based on the model green alga *Chlamydomonas reinhardtii* (Chlamydomonas hereafter), where the interaction of the 5 repeated helical regions of the linker protein EPYC1 (Essential PYrenoid Component 1,

formerly LCI5) with Rubisco causes condensation by LLPS (Freeman Rosenzweig *et al.,* 2017; Mackinder *et al.,* 2016; Wunder *et al.,* 2018). Similar to the linkers of carboxysomes, knock-out of EPYC1 *in vivo* results in the failure to condense Rubisco in the pyrenoid and a high CO_2-requiring phenotype typical of a non-functioning CCM (Mackinder *et al.,* 2016). More recent evidence exists for the role of an analogous multivalent disordered protein in the condensation of Rubisco in the pyrenoid of the diatom *Phaeodactylum tricornutum*, where *in vitro* experiments demonstrate the ability of PYCO1 to demix Rubisco (Oh *et al.,* 2023). Like EPYC1 and CsoS2-NTD, the interaction between PYCO1 and Rubisco is mediated by repeated helices spaced by regions of disorder (Fig. 1-3). Common to all linkers is the quasi-even spacing of Rubisco-binding domains by regions of disordered protein sequence of ~30-80 amino acids in length, in line with the role of linkers in cross-linking Rubisco proteins to underpin LLPS. Importantly, the hexadecameric structure of Rubisco provides an equal opportunity for multivalent interaction with the respective linker proteins, containing 8 repeated asymmetrical units for binding (Fig.1-2).

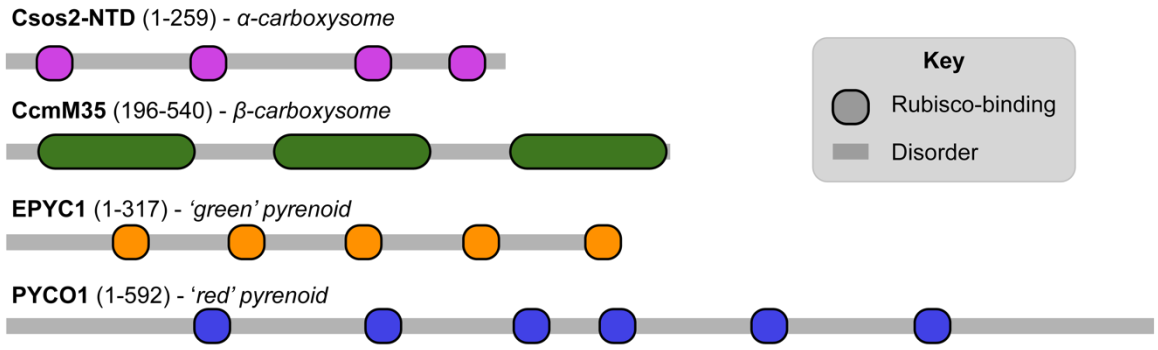


Figure 1-4 | Linker proteins form carboxysomes and pyrenoids.
Scaled schematic representation of the phase-separating domains of linkers from carboxysomes and the full length phase-separating proteins from pyrenoids.

1.4. General properties of liquid-liquid phase separated (LLPS) systems

Considering liquid-liquid phase separation (LLPS) is assumed to play an important role in the condensation of Rubisco in both carboxysomes and pyrenoids, at this point it is useful to describe the general properties of LLPS systems. The theoretical basis of general LLPS biology forms the basis for a large degree of subsequent discussion surrounding pyrenoid form and function in both Chlamydomonas and other pyrenoid-containing species in the remainder of this thesis.

Liquid-liquid phase separation (LLPS) is the process by which a homogeneous solution reversibly demixes to form a dense phase that is distinguished from a coexisting dilute phase. The solution composition, concentration, and conditions (pH, temperature etc.) define a phase diagram for demixing that is bounded by a coexistence line (also known as the binodal) that determines the one- and two-phase states (Fig. 1-5).

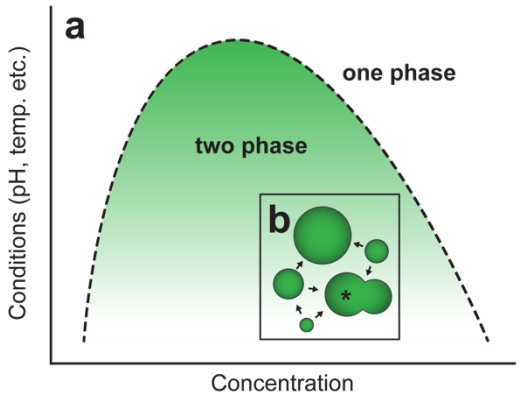


Figure 1-5 | Phase diagrams explain LLPS.
(a) A phase diagram determines a one- and two-phase state of the system, where the two-phase state consists of droplets distinguished from the coexisting dilute phase. Changes in solution conditions that affect interaction strength (y) and concentration of components (x) alter the phase state of the system relative to the coexistence line (black dashed). (b) Schematic of droplet growth where Ostwald ripening is shown with arrows that indicate the trafficking of solute from smaller to larger droplets. Asterisk indicates coalescence of two adjacent demixed droplets.

In biology, LLPS is prescribed to give rise to an array of membraneless bodies (Alberti, 2017) that provide spatiotemporal control over diverse cellular processes (Boeynaems *et al.,* 2018) by concentrating particular protein and nucleic acid species relative to the bulk phase (Li *et al.,* 2012), whilst permitting a rapidly diffusing biochemical environment (Banani *et al.,* 2017). Although membraneless bodies were described as early as 1803 (Vaucher, 1803), their liquid-like properties were

demonstrated much more recently. Brangwynne *et al.* (2009) reported fusion, dripping, fission and internal/external rearrangement of spherical P granules over second timescales in 2009. These observations hold true for many biomolecular condensates, where growth can occur by Ostwald ripening (growth of larger droplets at the expense of smaller droplets to reduce surface tension energetic penalty; Fig. 1-5b) and elastic ripening (transport of solute down a stiffness gradient) in addition to fusion by coalescence (Fig. 1-5b, asterisk) (Hyman *et al.,* 2014; Rosowski *et al.,* 2020). It should be highlighted that many systems have been classified as LLPS based on these qualitative descriptions, though other mechanisms of biomolecular condensate assembly are possible, and quantitative descriptors are required for their distinction (McSwiggen *et al.,* 2019).

Given their liquid-like nature, it is postulated that membraneless bodies can respond more rapidly to environmental cues than membrane-bound compartments (Cuevas-Velazquez & Dinneny, 2018), and as such are implicated in many transitory processes across a vast array of cellular contexts. Their composition is often accordingly vast (Mitrea & Kriwacki, 2016), and accompanied by an array of underpinning interactions, including electrostatic, π-π, cation-π, hydrophobic associations and hydrogen bonds between subsumed components (Dignon *et al.,* 2020). These interactions are often weak in nature and high in valency (number of binding sites on a binding partner) to facilitate formation of a network of interactions, required for phase separation (Peran & Mittag, 2020). This network forms homotypically, in simple coacervation, or between multiple protein species in complex coacervation (Alberti, 2017). Across these coacervation mechanisms, multivalency is provided by a range of associating sequence and structural features, comprising folded and/or unfolded domains, that can be loosely termed ‘stickers’. Variegating these stickers, are regions of structure or sequence that are termed ‘spacers’ (Choi & Pappu, 2020). Although often not directly involved in coacervation interactions, spacer presence and composition has marked effects on condensate properties, dependent on their solvation properties, but little effect on phase separation driving forces has been demonstrated (Choi *et al.,* 2020). Changes in valency, concentration and affinity of stickers sharply determines phase separation thresholds (Li *et al.,* 2012) and coacervate composition (Banani *et al.,* 2016). These changes often occur rapidly, through sharp transitions that can be influenced by a host of cellular factors, both

globally (pH, temperature and ionic strength - see Dignon *et al.* (2020) for review) and targeted (including methylation, phosphorylation, acetylation, SUMOylation - see Owen and Shewmaker (2019)), that account for condensate transiency.

Despite experiencing rapid transitions in response to relatively minute changes, biomolecular condensates can be stable throughout generational lifetimes (*e.g.,* the *Chlamydomonas reinhardtii* pyrenoid that is inherited and maintained through multiple cell division events (Freeman Rosenzweig *et al.,* 2017)), whilst retaining their liquid-like properties. Owing to their liquid nature, condensate morphology can be reversibly deformed, by wetting (adherence to solid surfaces due to intermolecular interactions) (Brangwynne *et al.,* 2009), disruption (Feric & Brangwynne, 2013) or compositional effects (Elbaum-Garfinkle *et al.,* 2015), commonly observed in the cellular environment. Surface tension underpins this behaviour (Han *et al.,* 2018), and is affected by coacervate component interaction strength and valency (Elbaum-Garfinkle *et al.,* 2015). A range of viscosity is also observed across coacervates and their lifetimes, and has been implicated in their functionality (Wei *et al.,* 2017). The maturation of condensates to more solid states has been proposed to occur *in vivo* (Banani *et al.,* 2017), mirroring the effect of gelation (transition towards a less dynamic structure underpinned by interaction strength increase) that influences droplet dynamics *in vitro* (Harmon *et al.,* 2017). The mechanistic implications of these macroscopic properties are relatively unexplored (Dignon *et al.,* 2020; Peran & Mittag, 2020), but are likely central to condensate activity regulation and physical resilience. Accordingly, microscopic perturbations that alter macroscopic properties are functionally intertwined. The movement of species within biomolecular condensates is influenced by both macroscopic and microscopic properties. The porosity of the primary scaffold components that constitute condensates determines the relative mobility of their subsumed components in a size-dependent manner (Banani *et al.,* 2017), referred to as the mesh-size, that is dependent on the extent of physical cross links (Peran & Mittag, 2020). Microscopically, the interaction of diffusing species with the biomolecular scaffold will also influence their mobility.

1.5. A brief summary of pyrenoids in algae

The majority of this thesis focusses on the characterisation of pyrenoids and their components, and it is therefore useful to provide a general overview prior to providing an in-depth account of the current state of knowledge of pyrenoid research. The breadth of pyrenoid presence in algae is also discussed here such that the context of the research within this thesis relative to the extent of uncharacterised pyrenoids can be considered.

1.5.1. Pyrenoids across algae

Pyrenoids are present in all algal lineages and most hornwort plants (Fig. 1-6) (Villarreal & Renner, 2012), but are not present in all other land plants (liverworts, mosses, ferns, and vascular plants) and presumably also not in non-photosynthetic organisms. Algae is a polyphyletic term for mostly aquatic photosynthetic eukaryotes, which includes over 70,000 described extant species (Guiry, 2012), though there are currently ~175,000 species records present in AlgaeBase (as of September 2023), suggesting this number could be significantly higher. Algal lineages are distributed widely across the eukaryotic tree of life (Fig. 1-6), with far greater diversity between algal species than between humans and other model eukaryotes. The complex phylogeny of algae is widely debated given their wide distribution (Leliaert *et al.,* 2012), though for the purposes of this study algae are grouped into seven clades according to their accepted chloroplast ancestry (Burki *et al.,* 2020) as follows. The first clade, Archaeplastida, which contains glaucophytes, rhodophytes (red algae) and green algae (chlorophytes, streptophytes, charophytes and prasinophytes) (and land plants), acquired their chloroplasts through separate primary endosymbiosis events of cyanobacteria >1 billion years ago (De Clerck *et al.,* 2012). Phylogeny of the Archaeplastida is perhaps most well understood relative to other clades (Cortona *et al.,* 2020; Jackson *et al.,* 2018; Yang, Ma, *et al.,* 2023), owing to the presence of several model organisms including *Chlamydomonas reinhardtii,* and the relationship with land plants within. All other algal clades inherited their chloroplasts through secondary or tertiary endosymbiotic events. The second clade, which contains only one photosynthetic group (euglenids), and the third clade, rhizaria, containing only photosynthetic chlorachniophytes, inherited their chloroplasts through separate secondary endosymbiosis events of green algae much more recently. The fourth clade, stramenopiles

(containing xanthophytes, chrysophytes, phaeophytes and bacilliarophytes/diatoms) inherited their chloroplasts through secondary endosymbiosis of a red alga. Stramenopiles contain likely the most geochemically important pyrenoids given the dominance of diatoms in the oceans (Mackinder *et al.,* 2016), and are in their infancy in terms of understanding of the molecular basis of pyrenoid assembly (Oh *et al.,* 2023). In the fifth clade, alveolates (containing dinoflagellates), chloroplast inheritance is complex with species having secondary or tertiary endosymbiosis-derived plastids originating from both red and green algal lineages. Algae belonging to clades 3, 4 and 5 are often summarised to the TSAR supergroup (Fig. 1-5). The sixth clade (containing haptophytes/coccolithophores) and the seventh clade (containing cryptophytes) both inherited their chloroplasts through separate secondary endosymbiosis events of red algae. Despite the breadth of algal species most of our molecular understanding of pyrenoids comes from two species *Chlamydomonas reinhardtii* and *Phaeodactylum tricornutum* (red asterisks – Fig. 1-6).

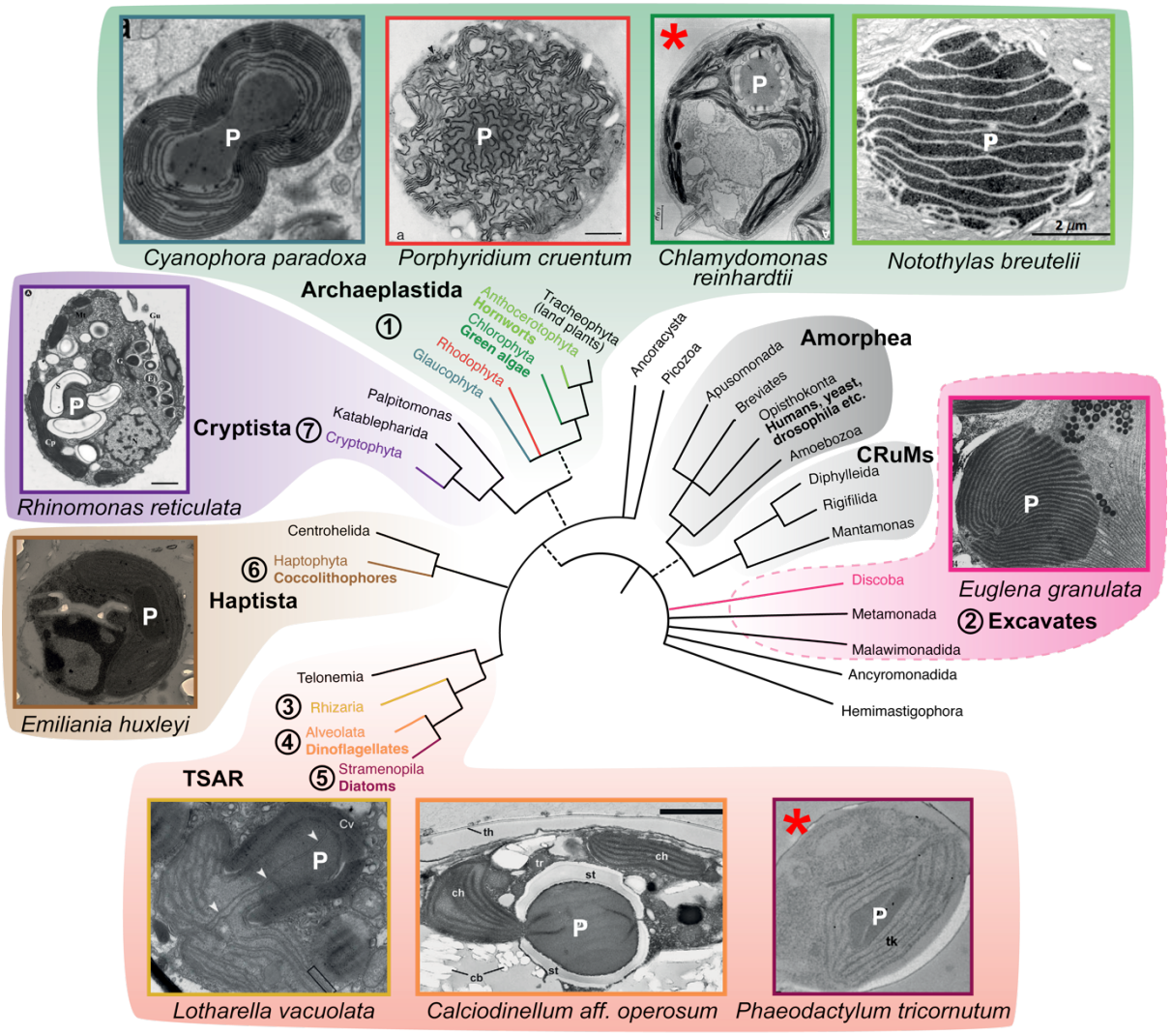


Figure 1-6 | Pyrenoid containing algae are polyphyletic and found across the eukaryotic tree of life. Tree is based on Burki *et al.* (2020). Dashed lines indicate uncertainty about the monophyletic nature of these groups. Transmission electron microscope image descriptions (clockwise from top left): A dividing pyrenoid (also known as central body) of the glaucophyte *Cyanophora paradoxa* (Sato *et al.,* 2009). *Porphyridium cruentum* with thylakoid membranes stained dark via photooxidation of 3,3'-diaminobenzidine.4HCl (DAB) that depicts Photosystem I activity (McKay & Gibbs, 1990). The model green alga *Chlamydomonas reinhardtii* (Ohad *et al.,* 1967). *Notothylas breutelii* (Li *et al.,* 2017). *Euglena granulata* (Walne & Arnott, 1967). The model diatom *Phaeodactylum tricornutum* (Tachibana *et al.,* 2011a). The dinoflagellate *Calciodinellum* aff. *operosum* (Zinssmeister *et al.,* 2013). *Lotharella vacuolata* (Ota *et al.,* 2005). The abundant biogeochemically important calcifying coccolithophore *Emiliania huxleyi* (author's collection). *Rhinomonas reticulata* var. *atrorosea* (Nam *et al.,* 2013). P indicates pyrenoid. The clades as mentioned in text are indicated. Red asterisks indicated species in which the most understanding of the pyrenoid is held.

1.5.2. A brief history of the pyrenoid

Many of the points in this section are expanded on in later sections, though this summary serves to demonstrate the acceleration of our understanding of pyrenoid biology over the last 220 years of study. The relatively large size and high density of pyrenoids make them easy to see via light microscopy, with descriptions in algae referring to the pyrenoid from 1803 (Fig. 1-7; (Vaucher, 1803)) and the first reports of pyrenoids in hornworts from 1885 (Schimper, 1885). As a result, pyrenoids may be the first LLPS organelles to be described, with the nucleolus described later in 1835 (Wagner, 1835). The term pyrenoid was coined in 1882 (Schmitz, 1882) and its presence and variation across evolutionarily-diverse algae was described throughout the mid-1900s by the increased use of transmission electron microscopy (TEM) imaging (Fig. 1-6), which allowed characterization of the pyrenoid ultrastructure. From early TEM images, it was assumed that the pyrenoid matrix, depending on the species, was either crystalline or amorphous; a view that was held until the development and implementation of high resolution imaging techniques in algae (Freeman Rosenzweig *et al.,* 2017). In the 1970s Rubisco was shown to be a major constituent of pyrenoids by enzymatic characterization of purified pyrenoids and analysis of Rubisco knock-out lines (Goodenough & Levine, 1970; Holdsworth, 1971; Kerby & Evans, 1978; Salisbury & Floyd, 1978), which was later confirmed by immunocytochemistry (Lacoste-Royal, Ginette & Gibbs, 1987; Vladimirova *et al.,* 1982). The association between pyrenoid presence and efficient CCM function was first made in the 1990s. Experimental observations showed that pyrenoid containing algae have an efficient CCM, with CCM induction concurrent with biochemical and structural changes to the pyrenoid, whereas algae lacking pyrenoids either lack a CCM or have a reduced ability to concentrate CO_2 (Badger *et al.,* 1993; Kuchitsu *et al.,* 1991; Palmqvist, 1993; Ramazanov *et al.,* 1994). The golden-age of understanding of the Chlamydomonas pyrenoid began in 2015, when a detailed understanding of the pyrenoid structure and Rubisco arrangement in the pyrenoid of *Chlamydomonas reinhardtii* was provided by high resolution cryogenic electron tomography (cryo-ET) (Engel *et al.,* 2015). The discovery that EPYC1 linked Rubisco to form the pyrenoid rapidly followed in 2016 (Mackinder *et al.,* 2016), with its liquid-like nature identified in 2017 (Freeman Rosenzweig *et al.,* 2017). The LLPS nature of the Rubisco-EPYC1 interaction was characterised *in vitro* in 2018 (Wunder *et al.,* 2018), and the structural basis of the interaction was determined by single particle

cryogenic electron microscopy in 2020 (He *et al.,* 2020). Later the same year, the characterisation of several proteins containing EPYC1-like Rubisco-binding motifs (RBMs) led to the formation of an RBM-guided assembly hypothesis for the pyrenoid (Meyer, Itakura, *et al.,* 2020). In 2023 the linker protein PYCO1 from the diatom *Phaeodactylum tricornutum* was characterised *in vitro* and represented the first step towards understanding the assembly of a pyrenoid outside of Chlamydomonas (Oh *et al.,* 2023). This brief history by no means represents a complete history of pyrenoid research, as many of the featured advances are built on years of fundamental research, though the more recent advances represent the most transformative research pieces with respect to our current understanding of the pyrenoid and its assembly. As outlined in this timeline, the majority of our detailed understanding of pyrenoid form and function comes from the model green alga *Chlamydomonas reinhardtii*.

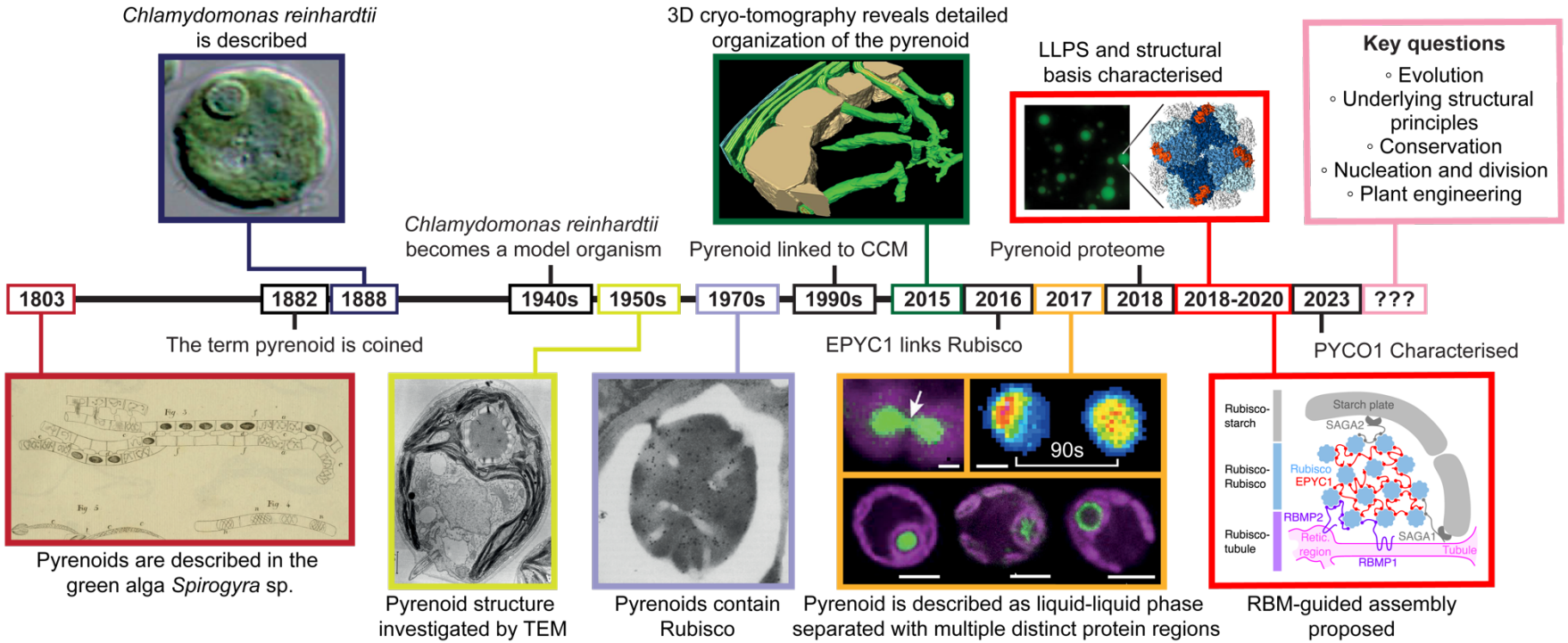

Chlamydomonas reinhardtii is described
3D cryo-tomography reveals detailed organization of the pyrenoid
LLPS and structural basis characterised
Key questions
◦ Evolution
◦ Underlying structural principles
◦ Conservation
◦ Nucleation and division
◦ Plant engineering
Chlamydomonas reinhardtii becomes a model organism
Pyrenoid linked to CCM
Pyrenoid proteome
1803
1882
1888
1940s
1950s
1970s
1990s
2015
2016
2017
2018
2018-2020
2023
???
The term pyrenoid is coined
EPYC1 links Rubisco
PYCO1 Characterised
90s
Starch plate
SAGA2
Rubisco-starch
Rubisco-Rubisco
Rubisco
EPYC1
RBMP2
Rubisco-tubule
Retic. region
SAGA1
Tubule
RBMP1
Pyrenoids are described in the green alga *Spirogyra* sp.
Pyrenoid structure investigated by TEM
Pyrenoids contain Rubisco
Pyrenoid is described as liquid-liquid phase separated with multiple distinct protein regions
RBM-guided assembly proposed

Figure 1-7 | A brief history of the pyrenoid
Pyrenoids were first described in the green alga *Spirogyra* by Jean-Pierre Vaucher in 1803. The original drawings by Vaucher display one ribbon-like chloroplast per cell that contains multiple spherical pyrenoids (Vaucher, 1803). In 1882, the term pyrenoid (Greek *pyrene*, stone kernel-like) was coined by Friedrich Schmitz, who observed pyrenoids in several algae species (Schmitz, 1882). Six years later, in 1888 the model alga *Chlamydomonas reinhardtii* was first described by Pierre Augustin Dangeard (Dangeard, 1888). *Chlamydomonas*, which is the central model for pyrenoid research, has one pyrenoid per cell that is visible in light microscopy (light microscopy image by Moritz Meyer (Meyer *et al.,* 2017)). In the 1940s, *Chlamydomonas* entered into research labs and over time became an essential model system (Salomé & Merchant, 2019). The use of TEM to image algae from the early 1950s onward made details of the pyrenoid ultrastructure with matrix, traversing thylakoids and starch sheath visible (TEM image of *Chlamydomonas* by Ohad *et al.* (Ohad *et al.,* 1967)). From the 1970s onward, it became clear that the pyrenoid contains most of the cell's Rubisco, which was later incontrovertibly proved by immunogold labelling (TEM image of immunogold-labelled Rubisco (black dots) in *Chlamydomonas* by Lacoste-Royal *et al.* (1987). The first associations of the pyrenoid with the CCM were made in the 1990s. In 2015, the 3D structure of the *Chlamydomonas* pyrenoid was resolved by cryo-electron tomography reconstruction of the pyrenoid (thylakoid tubules in green, starch sheath in beige, matrix is not displayed) by Engel *et al.* (2015). EPYC1 and its function as a Rubisco linker in the *Chlamydomonas* pyrenoid was discovered in 2016 (Mackinder *et al.,* 2016). In 2017 it was shown that the pyrenoid is formed by liquid-liquid phase separation (top left, the pyrenoid divides via fission; top right, fluorescent recovery after photobleaching shows that Rubisco undergoes internal mixing over second timescales in the pyrenoid) (Freeman Rosenzweig *et al.,* 2017) and multiple distinct protein regions were described, including the pyrenoid matrix (left, Rubisco), pyrenoid tubules (middle, PSAH) and starch sheath (right, LCI9) (Mackinder *et al.,* 2017) (fluorescence microscopy images (magenta: chlorophyll, green: labelled protein)). Zhan *et al.* (2018) reported a pyrenoid proteome in 2018. In 2018, it was demonstrated that Rubisco and EPYC1 demix by LLPS *in vitro* (Wunder *et al.,* 2018), and the structural basis for the interaction between EPYC1 and Rubisco that underpins this behaviour was later reported (He *et al.,* 2020). In the same year the RBM-guided assembly of the *Chlamydomonas* pyrenoid was proposed based on the identification of EPYC1-like Rubisco-binding motifs in a subset of *Chlamydomonas* pyrenoid proteins (Meyer, Itakura, *et al.,* 2020). In 2023, a linker protein (PYCO1) from the diatom *Phaeodactylum tricornutum* was reported (Oh *et al.,* 2023).

1.6. The Chlamydomonas pyrenoid

Given that the pyrenoid of Chlamydomonas is the most well-characterised, it is useful to here summarise the current state of knowledge of its function, assembly, properties, and other knowledge that are relevant to comparisons made in Chapters 2-5. Many of the studies outlined here were foundational to hypotheses made in subsequent research chapters. As before, this text is mainly based on the review completed in 2021 but updated where appropriate.

1.6.1. Structure and function of the Chlamydomonas pyrenoid

The functionality of the Chlamydomonas pyrenoid to concentrate CO_2 at Rubisco's active site requires structural features in addition to the pyrenoid matrix formed through Rubisco condensation (Fig. 1-3b). Central to CO_2 concentration and pyrenoid function is the shuttling of inorganic carbon through the subcellular environment to Rubisco within the pyrenoid. The spatial segregation of a carbonic anhydrase within the pyrenoid is essential for catalysing the subsequent dehydration of inorganic carbon (in the form of HCO_3^-) to CO_2, allowing release for carboxylation by Rubisco in the matrix (Fig. 1-3c). As discussed above, these characteristics are considered basal for the function of biophysical CCMs and are thus expected to be conserved across pyrenoid-based CCMs.

Although ultrastructural information is available for a wealth of species across diverse lineages, our most complete insights relating to pyrenoid form and function are from Chlamydomonas. Detailed three-dimensional structural information of the pyrenoid obtained by ion beam milling cryo-Electron Tomography (cryo-ET; (Engel *et al.,* 2015; Freeman Rosenzweig *et al.,* 2017; He *et al.,* 2020)), Quick Freeze Deep-Etch Electron Microscopy (QFDEEM; (Mackinder *et al.,* 2016)) and high throughput fluorescent protein localization (Mackinder *et al.,* 2017; Wang *et al.,* 2023), have significantly enhanced our understanding of pyrenoid architecture (Fig. 1-7). Further, proteomics of pyrenoid-enriched fractions have revealed the complex composition of the pyrenoid, containing at least 190 different proteins, many of which remain uncharacterized (Mackinder *et al.,* 2016; Zhan *et al.,* 2018). An integrated localization and interaction study also indicated a large number of pyrenoid components (89), many of which have been localized at sub-pyrenoid resolution (Mackinder *et al.,* 2017). A more recent

proximity labelling approach indicated a high confidence 'proxiome' specific to the pyrenoid matrix which contained 30 proteins, and revealed further novel components (Lau *et al.,* 2023). Although many different proteins have been localized to the pyrenoid, proteomic analysis shows the pyrenoid matrix consists mainly of Rubisco molecules (Freeman Rosenzweig *et al.,* 2017)) and the linker protein EPYC1, that is essential for condensation of Rubisco to form the pyrenoid matrix (Fig. 1-8b) (Freeman Rosenzweig *et al.,* 2017; Mackinder *et al.,* 2016). Here, Rubisco functions within the CBB cycle, with strong evidence supporting that it is the only CBB enzyme partitioned within the pyrenoid (Fig. 1-8c) (Küken *et al.,* 2018).

1.6.2. CO_2 delivery in the pyrenoid

In addition to the pyrenoid matrix, TEM has previously revealed the thylakoid membranes that traverse the pyrenoid form a characteristic star-shaped network within (Fig. 1-8a). More recent *in situ* cryo-ET revealed the intriguing complexity of this thylakoid membrane organization and structural changes as it enters the pyrenoid matrix. Thylakoid membranes outside of the pyrenoid that mediate the photosynthetic light reactions are organized in multiple parallel stacked membrane layers (Wietrzynski *et al.,* 2020), which drastically change as they enter the pyrenoid matrix through fenestrations in the transient stromal starch sheath. The membrane layers merge into cylindrical structures, termed pyrenoid tubules, that advance through the pyrenoid matrix and converge in the centre of the pyrenoid, forming an interconnected network of smaller, shorter tubules (Engel *et al.,* 2015). Within pyrenoid tubules, minitubules form luminal conduits between the chloroplast stroma and the pyrenoid matrix and based on their diameter (~3-4 nm at matrix opening) have been proposed to facilitate exchange of ATP and CBB metabolites (incoming RuBP and outgoing 3-PGA) but not proteins (Fig. 1-8c) (Engel *et al.,* 2015; Küken *et al.,* 2018; Mackinder *et al.,* 2016). The wider lumen of the pyrenoid tubules is continuous with the thylakoid lumen and is postulated to transport HCO_3^- towards the centre of the pyrenoid as outlined in the general scheme of a biophysical CCM (Fig. 1-3b). HCO_3^- is proposed to be channelled from the chloroplast stroma into the thylakoid lumen through bestrophin-like channels (Mukherjee *et al.,* 2019). Central to CO_2 delivery is the carbonic anhydrase CAH3, that catalyses the dehydration of HCO_3^- to CO_2 in the acidic lumen of the pyrenoid tubules, enabling the release of CO_2 across the tubule

membrane into the pyrenoid matrix (Fig. 1-8d). CAH3 has been localized to the pyrenoid tubules and a *cah3* mutant has a defective CCM, despite accumulating inorganic carbon at higher concentrations than wild-type (Duanmu *et al.,* 2009; Karlsson *et al.,* 1998; Mitra *et al.,* 2004; Sinetova *et al.,* 2012). There is strong evidence that the pyrenoid tubules also differ in their protein composition from the rest of the thylakoid membrane. Immunogold labelling and photosystem (PS) I and PSII activity assays suggested that the pyrenoid tubules contain active PSI but not PSII (McKay & Gibbs, 1991). However, more recent fluorescent protein tagging of several photosystem proteins revealed that subunits of both photosystems are present in the tubules, indicating that partially-assembled or inactive PSII may be present (Mackinder *et al.,* 2017). Strikingly, some PSI subunits (*e.g.*, PSAH) are even enriched in the tubules. The functional implications of these depletion/enrichment patterns are yet to be experimentally shown but could be related to reducing photorespiration by minimizing O_2 release within the pyrenoid through the absence of photosynthetic H_2O splitting at PSII reaction centres. Collectively, these observations highlight the importance of membrane traversions in the metabolic fluxes of the pyrenoid and suggest an important role in its photosynthetic operation. It should be noted here that the stellate thylakoid traversions and minitubule organisation observed in Chlamydomonas appear to be an outlying case, rather than the norm for membrane traversion in pyrenoids more generally (Fig. 1-6 and Fig. 1-9).

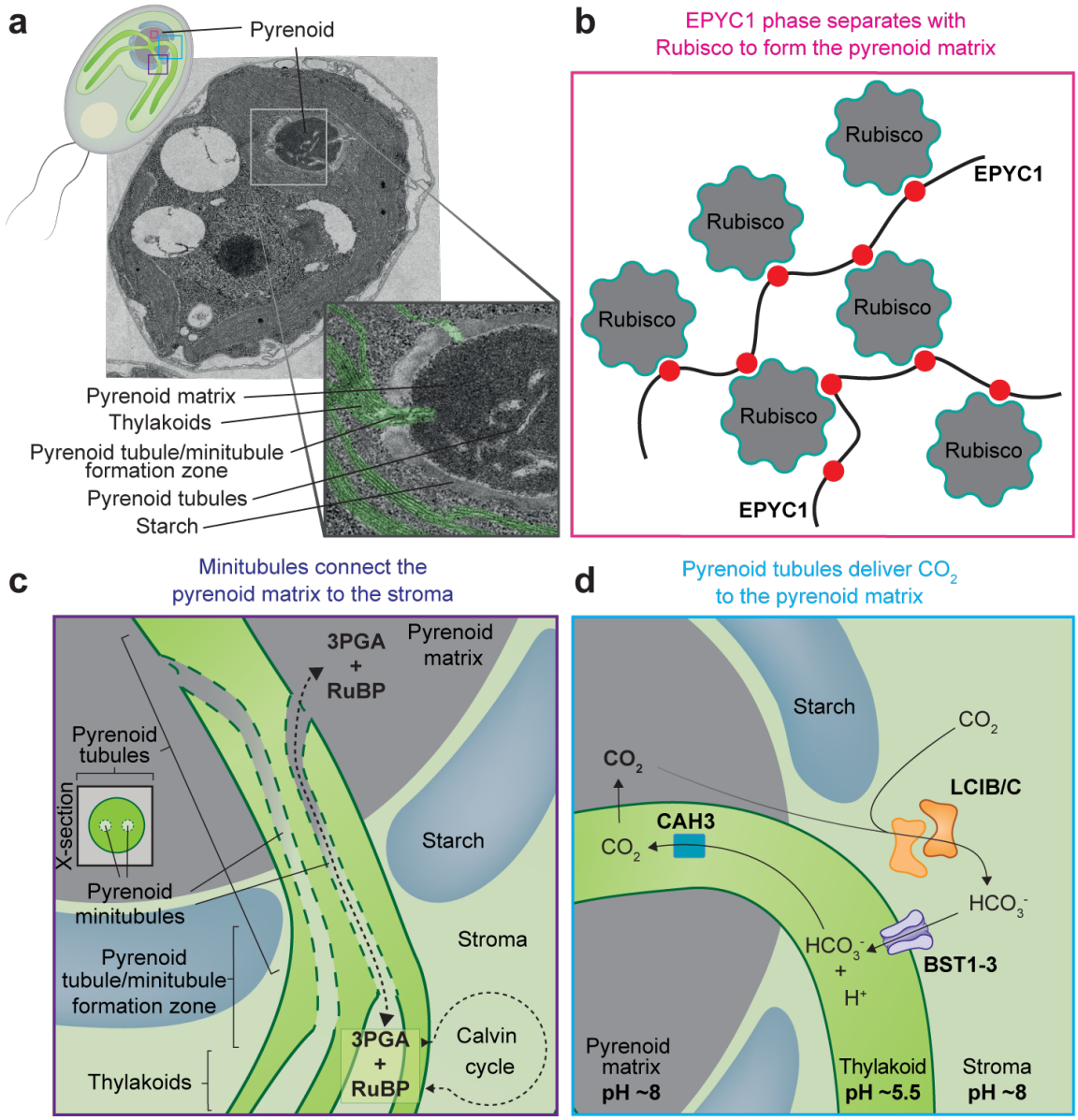


Figure 1-8 | The Chlamydomonas pyrenoid is at the heart of the CO_2 concentrating mechanism and enables efficient CO_2 fixation.
(a) TEM image of *Chlamydomonas reinhardtii* grown in light and under air levels of CO_2 where a complete pyrenoid is assembled. Zoom highlights key structural parts of the pyrenoid. Thylakoids false coloured green for clarity. Top left diagram is for orientation of panels B-D. (b) The pyrenoid matrix is predominantly composed of Rubisco-EPYC1 condensate. Multiple Rubisco binding regions on EPYC1 enable complex coacervation with the Rubisco holoenzyme which is a hexadecameric assembly of 8 large and 8 small subunits. (c) As thylakoids enter the pyrenoid they form pyrenoid tubules. Minitubules (dashed lines) form within the pyrenoid tubules and connect the pyrenoid matrix to the stroma. They are postulated to enable the large flux of metabolites in and out of the pyrenoid. Inset: cross-section (X-section) of minitubules within a pyrenoid tubule. (d) Pyrenoid tubules are proposed to deliver CO_2 to Rubisco in the pyrenoid matrix. Current data supports that HCO_3^- enters from the stroma into the thylakoid lumen via bestrophin-like channels. In the acidic lumen, HCO_3^- is converted to CO_2 via CAH3 and subsequently diffuses into the pyrenoid matrix. LCIB/LCIC is proposed to convert stromal CO_2 to HCO_3^- via active CO_2 uptake and CO_2 recapture from the pyrenoid. Minitubules are not shown for clarity.

1.6.3. Mechanisms to reduce CO_2 loss

The Chlamydomonas pyrenoid is surrounded by a sheath that is composed of several starch plates. The starch sheath develops rapidly under limiting CO_2 concentrations (Kuchitsu *et al.,* 1988) and has been proposed to act as a diffusion barrier to reduce the loss of CO_2, that diffuses readily between the stroma and matrix (Ramazanov *et al.,* 1994). Although it has been suggested that the absence of the starch sheath does not affect photosynthetic productivity (Villarejo *et al.,* 1996), recent studies indicate a correctly formed starch sheath is required for normal pyrenoid formation and the operation of an efficient CCM (Itakura *et al.,* 2019; Shimamura *et al.,* 2023; Toyokawa *et al.,* 2020). Modelling approaches have also demonstrated the benefit of the starch sheath as a diffusion barrier in the pyrenoid (Fei *et al.,* 2022). In addition to starch, the sheath also contains several proteins. These proteins appear to be distributed uniformly over the starch plates, or localized in distinct puncta or meshes in close proximity to the starch plates (Mackinder *et al.,* 2017). The functional implications of these different distribution patterns remain unclear, but their positioning appears to be important for CCM function (Shimamura *et al.,* 2023; Toyokawa *et al.,* 2020). A subset of proteins that localize in a plate-like pattern are predicted to function as starch-branching enzymes, whereas the mesh distributed proteins appear to fill the gaps between the starch plates, indicating a potential structural function (Mackinder *et al.,* 2017). A predicted starch-binding Rubisco-interacting protein, SAGA1 (StArch Granules Abnormal 1), that localizes to distinct puncta at the pyrenoid matrix/starch interface was shown to affect pyrenoid number and sheath morphology (Itakura *et al.,* 2019), and has also been shown to influence regulation of CCM components (Shimamura *et al.,* 2023). Interestingly, two carbonic anhydrase homologs, LCIB and LCIC, that are recruited to the pyrenoid in very low CO_2 concentrations (<0.02% CO_2) are also localized in distinct puncta but on the external surface of the starch sheath (Mackinder *et al.,* 2017; Yamano *et al.,* 2010). Here they are expected to minimize the loss of CO_2 from the pyrenoid by converting emanating CO_2 back to HCO_3^- such that it can be readily concentrated again (Fig. 1-8d) (Wang *et al.,* 2015). LCIB homologs show *in vitro* carbonic anhydrase activity, however this could not be demonstrated for the Chlamydomonas LCIB/LCIC proteins (Jin *et al.,* 2016), potentially indicating the absence of critical regulatory subunits or that activity requires specific cellular conditions.

1.6.4. EPYC1 links Rubisco to form the pyrenoid matrix

In Chlamydomonas, the abundant, low CO_2-induced linker protein, EPYC1, underpins the functional phase separation of Rubisco to form the pyrenoid (Mackinder *et al.,* 2016; Wunder *et al.,* 2018). In mutants depleted of EPYC1, Rubisco fails to aggregate and is dispersed in the chloroplast (Mackinder *et al.,* 2016), resulting in a deficient CCM (Mackinder *et al.,* 2016). EPYC1 is a low complexity, largely disordered ~35 kDa protein, consisting of five near-identical repeats (Fig. 1-4) (He *et al.,* 2020; Mackinder *et al.,* 2016; Wunder *et al.,* 2018). Each ~60 amino acid repeat contains a predicted α-helix, and significant charge patterning (Mackinder *et al.,* 2016; Wunder *et al.,* 2018). The high isoelectric point (pI) of EPYC1 (11.7) establishes a net positive charge of the unmodified protein in the slightly basic pyrenoid matrix (pH ~7-8.5), in both photosynthetic and non-photosynthetic conditions (Antal *et al.,* 2013; Freeman Rosenzweig *et al.,* 2017). *In vitro* demixing assays have demonstrated that LLPS of Rubisco by EPYC1 occurs via complex coacervation, in which both components are required (Wunder *et al.,* 2018). In line with general LLPS principles (Mitrea & Kriwacki, 2016), demixing was also demonstrated to require multivalent interactions between the Rubisco holoenzyme and EPYC1.

Prior to EPYC1 discovery and functional characterization, Meyer *et al.* (2012) demonstrated that the sequence composition of the surface-exposed α-helices of the Rubisco small subunit (RbcS) was conditional for Chlamydomonas pyrenoid formation. Wunder *et al.* (2018) confirmed the importance of this interface for *in vitro* demixing and suggested the association could be dominated by charge interactions between negative patches of the RbcS α-helices and regions of patterned positive charge in EPYC1. Yeast two-hybrid (Y2H) data confirmed the RbcS α-helices are necessary for interaction with EPYC1, and that other RbcS features enhance this interaction (Atkinson *et al.,* 2019). In line with previous predictions (Mackinder *et al.,* 2016), it was later demonstrated that EPYC1's interaction with Rubisco is enhanced by its repeating helical regions (Atkinson *et al.,* 2019). More recently, single particle cryogenic Electron Microscopy (cryo-EM) of a complex of Rubisco and a 24 amino acid EPYC1 peptide containing the helical region has outlined the structural basis for this interaction, revealing a primarily electrostatic and hydrophobic interface (He *et al.,* 2020). The peptide was bound to each of the small subunits of Rubisco, indicating the holoenzyme can bind one of EPYC1's five

helical regions up to 8 times. In the same study, mutation of EPYC1 interface residues decreased demixing of Rubisco *in vitro* and Rubisco substitutions at the interface prevented pyrenoid formation *in vivo*, confirming the role of this low affinity interaction in condensation of Rubisco. It is proposed that consecutive binding regions of the full length EPYC1 protein can facilitate the low affinity, multivalent interactions with multiple Rubisco molecules required for condensation into the pyrenoid matrix. Key to this model is the ability for the unstructured region between two adjacent helical regions to span the distance between Rubisco holoenzymes in the pyrenoid. *In situ* cryo-ET data indicates a median distance of ~4 nm between EPYC1 binding sites on adjacent holoenzymes (Engel *et al.,* 2015; He *et al.,* 2020). The ~40 amino acid unstructured regions between the 5 binding regions of EPYC1 are proposed to facilitate the spanning of this distance, with wormlike chain models indicating a minimal energetic cost (< 3 k_bT) for stretching (He *et al.,* 2020). More recent modelling approaches have begun to provide a theoretical framework for the condensation of Rubisco by EPYC1 (GrandPre *et al.,* 2023; Schaefer, 2023).

As discussed in section 1.3.4, EPYC1 displays functional similarity to the linkers PYCO1 of *Phaeodactylum tricornutum* and CsoS2 and CcmM of cyanobacteria (Cai *et al.,* 2015; Oh *et al.,* 2023; Oltrogge *et al.,* 2020; Wang *et al.,* 2019). EPYC1, PYCO1 and CsoS2 all utilise helical regions to contact Rubisco, whereas CcmM utilises a RbcS-like globular domain. Although all four Rubisco condensation events appear to be underpinned by a similar multivalent mechanism, the sequences of the Rubisco-interacting regions of CsoS2 and CcmM bear no homology to each other nor EPYC1 or PYCO1, suggesting this mechanism evolved independently in each lineage. CsoS2 and CcmM concurrently contact both the large and small subunits of the morphologically similar form I Rubisco holoenzyme in the cyanobacterium *Synechococcus elongatus* and the chemoautotrophic proteobacterium *Halothiobacillus neapolitanus* respectively (Oltrogge *et al.,* 2020; Wang *et al.,* 2019). It is postulated that the concurrent binding of CsoS2 and CcmM to both the Rubisco large and small subunits results in only fully assembled and functional Rubisco holoenzymes being incorporated into the carboxysome (Oltrogge *et al.,* 2020; Wang *et al.,* 2019). Current data suggests that EPYC1 may exclusively contact the small subunit (He *et al.,* 2020), whereas PYCO1 contacts both the large and

small subunit of Rubisco using separate helical interaction domains (Oh *et al.,* 2023). Additionally, whereas CsoS2 and CcmM facilitate aggregation using only a portion of their full-length sequence, EPYC1 and PYCO1 appear to dedicate their full length to multivalent interactions with Rubisco (Oltrogge *et al.,* 2020; Wang *et al.,* 2019). Although it is expected that linker proteins facilitate phase separation of other pyrenoids, the lack of obvious EPYC1/PYCO1 homologs suggests that analogous linker proteins will display a range of sequence characteristics across pyrenoid lineages, especially outside of the Archaeplastida (form IB Rubisco), where Rubisco forms are variant (dinoflagellates/alveolata [form II], all other clades [form ID]). Predictions based on characterized linkers, suggest that analogous proteins will contain a) regions of disorder that are continuous and cover part, or all of the protein sequence; b) repeat motifs within this disordered region that will interact with Rubisco using localized structure; c) patterning of charged residues throughout the full-length protein; and d) low complexity amino acid sequences. Mackinder *et al.* (2016) predicted the presence of analogous proteins in four other species using a search framework based on some of these constraints, but these are yet to be experimentally validated. These observations, alongside data from green algae that pyrenoid presence is not determined by RbcS sequence (Goudet *et al.,* 2020), certainly suggest that the presence of analogous linker proteins is probably widely determinant of pyrenoid formation across lineages.

1.6.5. A Rubisco-binding motif targets proteins to the pyrenoid and may guide pyrenoid assembly

Recent work has proposed a framework for assembly of the Chlamydomonas pyrenoid with its ultrastructural features (Fig. 1-7), based on the observation that multiple pyrenoid localized proteins contain the same Rubisco-binding motif (RBM) that underpins EPYC1's interaction with Rubisco (Meyer, Itakura, *et al.,* 2020). The RBM is repeated two or more times, including at the C-terminal, in other confirmed pyrenoid localised proteins with diverse structural features including predicted transmembrane domains and predicted starch binding domains. Two proteins that contain both RBMs and transmembrane domains (termed Rubisco Binding Membrane Proteins 1 and 2 – RBMP1/2) specifically localise to the pyrenoid tubules. A further two proteins containing RBMs and starch binding domains, SAGA1/2, localized to the pyrenoid matrix starch interface. From these data, an elegant assembly mechanism was suggested, in which RBMP1/2 tether the Rubisco matrix to the pyrenoid tubules and SAGA1/2 tether the starch sheath to the matrix (Meyer, Itakura, *et al.,* 2020). Since this hypothesis was proposed, characterization of RBMP1, also known as Bestrophin-like protein 4 (BST4) has suggested no direct role in thylakoid traversion (Adler *et al.,* 2023). Characterisation of RBMP2 deletion mutants has yet to be completed though there is supporting evidence for the role of SAGA1 as a Rubisco matrix starch tether (Itakura *et al.,* 2019). Further to this hypothesis, fusion of the RBM to both non-pyrenoid localised stromal, and thylakoid, proteins was shown to relocalise those proteins to the pyrenoid matrix and pyrenoid tubules respectively (Meyer, Itakura, *et al.,* 2020), suggesting RBMs also play a role in targeting of proteins to the pyrenoid. This hypothesis was supported by recent evidence in which the RBMs of RBMP1 were mutated, which caused relocalisation from the pyrenoid to the whole thylakoid network (Adler *et al.,* 2023). Taken together these evidence support the role of RBMs akin to those of EPYC1 in the assembly and function of the Chlamydomonas pyrenoid. A similar mechanism has not been explored in *Phaeodactylum*.

1.6.6. Evidence for pyrenoid LLPS

Several lines of both qualitative and quantitative evidence exist to support that the pyrenoid in Chlamydomonas is underpinned by LLPS. Qualitatively, the classical *in vivo* observations that the pyrenoid adopts a largely spherical morphology that can be reversibly deformed and appears to be wetted to the surrounding starch sheath (Griffiths, 1970) suggest the condensate experiences surface tension effects characteristic of liquid-like condensates (Hyman *et al.,* 2014). Further qualitative *in vivo* evidence in line with the reports of Brangwynne *et al.* (2009) were reported by the pioneering work of Freeman Rosenzweig *et al.* (2017) who observed fluorescently labelled Chlamydomonas pyrenoids through consecutive mitotic cellular divisions. In the majority case (~66%), the Chlamydomonas pyrenoid is inherited by fission of the pyrenoid from the mother cell into the two resulting daughter cells. Time-resolved imaging of this process highlighted a "bridge" of matrix between the dividing pyrenoids prior to separation (visible in Fig. 1-4 – 2017 box, top left), akin to the "dripping" observed in P granules (Brangwynne *et al.,* 2009); a hallmark behaviour of a simple liquid. Consistent with surface tension effects, the "bridge" of matrix retracted following separation to reform the spherical morphology of the 'vegetative' pyrenoid matrix. In cases where the pyrenoid is not divided by fission and asymmetric inheritance occurs, the pyrenoid-less daughter cell can form a pyrenoid *de novo*, thus demonstrating two additional indicators of LLPS. During the formation of the *de novo* pyrenoid, several fluorescent puncta containing both Rubisco and EPYC1 emerge over minute timescales, indicating a quantitative threshold for condensate formation in the daughter cell – another key indicator of LLPS, also termed the 'critical concentration'. These puncta subsequently fused by coalescence or grow, presumably by Ostwald ripening, to form a single pyrenoid on the scale of the mother pyrenoid, consistent with the droplet growth mechanisms outlined in section 1.4 (Fig. 1-5b). Prior to division, it was observed that a significant portion (~35-50%) of Rubisco and EPYC1 dissolve from the pyrenoid into the chloroplast stroma, and subsequently relocalise back to the pyrenoid following division. The rapid timescale of this dynamic process (1-5 minutes) whilst maintaining phase separation of the condensate demonstrated another key property of liquid-like compartments observed in other systems (Brangwynne *et al.,* 2009; Li *et al.,* 2012). It is expected that this dispersion reduces pyrenoid viscosity and surface tension and accordingly reduces the mechanical force required for pyrenoid fission by the

centrally positioned cleavage furrow of the dividing chloroplast, in line with physical theory (Giustiniani *et al.,* 2017). In addition to the observation of pyrenoids during division, Freeman Rosenzweig *et al.* (2017) also made several quantitative measurements of the pyrenoid matrix in 'vegetative' pyrenoids that affirmed their liquid-like nature. *In situ* measurement of the distribution of individual Rubisco molecules using cryo-ET demonstrated short-range distribution patterns that are characteristic of liquid-like order. Additionally, Fluorescence Recovery After Photobleaching (FRAP) measurements of EPYC1 and Rubisco in the pyrenoid demonstrated internal rearrangement over second timescales.

The second primary lines of evidence for the LLPS nature of the Chlamydomonas pyrenoid were provided by the equally pioneering *in vitro* work of Wunder *et al.* (2018). Here it was demonstrated that functional Rubisco could be demixed by the linker EPYC1 under physiologically relevant conditions and concentrations in a valency-dependent manner. The resultant phase demonstrated similarities with the *in vivo* pyrenoid counterparts, with FRAP analysis indicating rearrangement over similar timescales. The dynamic fusion, ripening and dissolution of droplets were also demonstrated. Taken together, the *in vivo* and *in vitro* observations of the Chlamydomonas pyrenoid demonstrate its liquid-like nature and formation by liquid-liquid phase separation (LLPS).

1.7. Does LLPS underpin pyrenoid assembly across lineages?

Pyrenoids in all lineages consist of an electron-dense matrix that is believed to be a Rubisco condensate (Fig. 1-6, Fig. 1-9). This assumption is based on observations across pyrenoid containing lineages that the pyrenoid matrix contains most of the cell's Rubisco (Blank & Trench, 1988; Kajikawa *et al.,* 1988; Kiss *et al.,* 1986; Lacoste-Royal, Ginette & Gibbs, 1987; McKay & Gibbs, 1990, 1989, 1991; McKay *et al.,* 1991; Mustardy *et al.,* 1990; Osafune *et al.,* 1989, 1990; Vaughn *et al.,* 1990; Vladimirova *et al.,* 1982). To date there is only conclusive evidence in Chlamydomonas that the pyrenoid is a LLPS organelle (Freeman Rosenzweig *et al.,* 2017; Wunder *et al.,* 2018), with emerging evidence that the LLPS may play a role in the assembly of the *Phaeodactylum tricornutum* pyrenoid (Oh *et al.,* 2023). There are however several lines of evidence that suggest that less explored pyrenoids could be formed by LLPS across diverse lineages also.

Firstly, analysis of ultrastructural reports of pyrenoids across diverse taxa indicates that almost universally, pyrenoids have spherical/elliptical shapes as described for the Chlamydomonas pyrenoid (Fig. 1-6, Fig. 1-9), consistent with surface tension effects. The pyrenoid matrix in many cases also appears to be commonly deformed by, or wetted to, the wide range of arrangements of ultrastructural features observed across pyrenoids. Secondly as outlined in section 1.6, much of the evidence for the liquid-like nature of the Chlamydomonas pyrenoid comes from observations of the pyrenoid during division. Fission (Chan, 1974) and *de novo* (Hoffman, 1968b; Ohiwa, 1976; Retallack & Butler, 1970) pyrenoid inheritance are classically described in green algae, but have also been described in hornworts (McAllister, 1914) and non-green lineages (fission: (Gantt & Conti, 1965; Hori & Inouye, 1981; Lander, 1935; Schornstein & Scott, 1982); *de novo* assembly: (Cox, 1981; Evans, 1966; McAllister, 1927; Nagasato & Motomura, 2002; Osafune *et al.,* 1990; Sun, 1962)), where both mechanisms appear equally prevalent. Besides Chlamydomonas, multiple concurrent pyrenoid inheritance mechanisms (*i.e.,* fission and *de novo* mechanisms occurring in the same cell population) have only been reported in *Arachnochrysis demoulinii* sp. nov. (stramenopila; (Han *et al.,* 2018)) and *Chlorogonium elongatum* (chlorophyta; (Ueda, 1963)). It is unlikely the observation of multiple pyrenoid inheritance is unique to these species and Chlamydomonas, and is instead likely reported due to more intensive characterization

of those species. Anecdotally, *de novo* pyrenoid formation following division appears to be reported more commonly in species where vegetative cells possess multiple distinct pyrenoids. This observation appears to extend across lineages, but exceptions do exist in some chlorophyta (Smith, 1918; Vítová *et al.,* 2008) and the definition of 'multiple pyrenoids' becomes unclear, especially in multicellular hornworts (Vaughn *et al.,* 1990). Pyrenoid fission in most cases is typically induced by plastid constriction, or less commonly, starch sheath invagination (Chan, 1974), with pyrenoid fission driven by plastid constriction being documented across lineages: green algae (chlorophyta) (Goodenough, 1970), haptophytes (haptophyta) (Manton, 1967), hornworts (bryophyta) (Lander, 1935; Schuette & Renzaglia, 2010), diatoms (stramenopila) (Mann, 1985; Subrahmanyan, 1945), brown algae (stramenopila) (Evans, 1966; Tanaka *et al.,* 2007), red algae (rhodophyta) (Patrone *et al.,* 1991; Schornstein & Scott, 1982) and dinoflagellates (alveolata) (Jenks & Gibbs, 2000). Thirdly, dissolution of the pyrenoid and dynamic Rubisco relocalization has been reported across diverse algae including the dinoflagellate *Gonyaulax* (Nassoury *et al.,* 2001) and the green alga *Dunaliella tertiolecta* (Lin & Carpenter, 1997) as well as *Euglena gracilis* (Osafune *et al.,* 1990) and in hornworts (Vaughn *et al.,* 1990). Though far from definitive, these qualitative indicators suggest that the dynamic processes associated with LLPS of the Chlamydomonas pyrenoid are present in a wide range of pyrenoid-containing species.

Lastly, along with observational evidence, bioinformatic analysis revealed the occurrence of proteins in a broad range of algae that show similarities to the Chlamydomonas Rubisco linker protein EPYC1 (Mackinder *et al.,* 2016). These proteins have a similar repeat number, length, isoelectric point, and disorder profile to EPYC1, indicating a putative function as linker proteins. All in all, the observed spherical shape of the pyrenoid, the observation of pyrenoid fission and identification of proposed Rubisco linker proteins, suggests that pyrenoids are formed by LLPS across algal lineages. Taken together these four primary lines of evidence suggest that pyrenoids across diverse taxa demonstrate qualitative hallmarks of LLPS and may be underpinned by similar assembly principles.

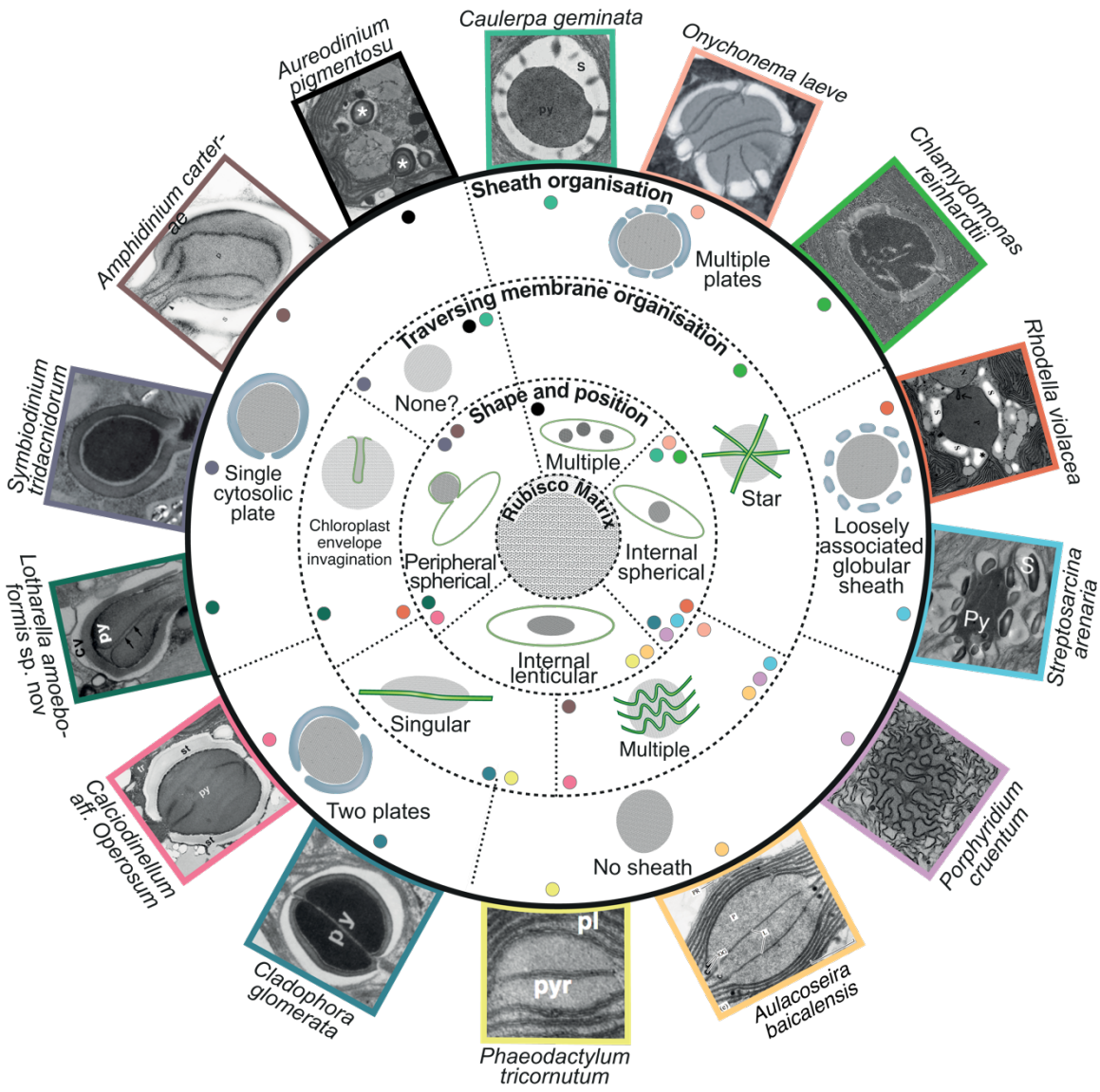


Figure 1-9 | Pyrenoids are structurally diverse.
Peripheral TEM images are used to demonstrate non-exhaustive examples of combinations of pyrenoid structural features. Starting centrally, with the defining pyrenoid matrix and radiating outwards, combinations of matrix shape/position, traversing membrane organization and sheath organization can be achieved. Observed combinations are loosely demarcated by dashed lines between and within rings, but no combinations are definitively excluded. Coloured dots indicate combinations of structural features, corresponding to image borders. Image labelling is from original publication: p/py/pyr/*, pyrenoid; s/st, starch; pl, plastid; cv, capping vesicle. References for images, clockwise from *Caulerpa geminata* (Calvert *et al.,* 1976), *Onchynonema laeve* (Goudet *et al.,* 2020), *Chlamydomonas reinhardtii* (authors collection), *Rhodella violacea* (Scott *et al.,* 2011), *Streptosarcina arenaria* (Mikhailyuk *et al.,* 2018), *Porphyridium cruentum* (McKay & Gibbs, 1990), *Aulacoseira baicalensis* (Bedoshvili *et al.,* 2009), *Phaeodactylum tricornutum* (De Martino *et al.,* 2011), *Cladophora glomerata* (McDonald & Pickett-Heaps, 1976), *Calciodinellum* aff. *Operosum* (Zinssmeister *et al.,* 2013), *Lotharella amoeboformis* sp. nov (Ishida *et al.,* 2000), *Symbiodinium tridacnidorum* (Lee *et al.,* 2015), *Amphidinium carterae* (Jenks & Gibbs, 2000), *Aureodinium pigmentosu* (Dodge, 1968).

1.8. Pyrenoid evolution and diversity

1.8.1. Pyrenoid evolution

Though pyrenoids occur in all algal lineages and in hornworts (Fig. 1-6), not all algae or hornwort species contain pyrenoids, consistent with the fact their molecular basis was not inherited from a common ancestor. Pyrenoid-less algae (*e.g.,* the extremophile rhodophyte class *Cyanidiophyceae*, members of the chlorophyte genera *Bathycoccus* and *Chloromonas*, the TSAR class chrysophyte (golden algae), and most species of the Eustigmatophyte genus *Nannochloropsis*) and hornworts are interspersed across the phylogenetic tree, suggesting that pyrenoids were gained multiple times during evolution (Morita *et al.,* 1998; Raven *et al.,* 2012; Villarreal & Renner, 2012), possibly with hundreds of evolutionary origins (Meyer, Goudet, *et al.,* 2020). The exact distribution of pyrenoids within each lineage is unknown since the anatomy of many algae has rarely been investigated thoroughly at even phylogenetic family levels. Perhaps occluding this, the occurrence of a pyrenoid in different algal species could depend on factors such as CO_2 abundance, light, and life-cycle stage. Thus, the apparent absence of pyrenoids in some species might be attributed to the metabolic state of the imaged cells, life-cycle stage or even missed due to insufficient imaging. Accordingly, the evolutionary history of pyrenoids is far from understood in most pyrenoid-containing lineages. Analysis of ultrastructural differences between the pyrenoids of algal species in different lineages has previously been used to constrain their evolution to timepoints following the divergence of the lineages and support the convergent evolution theory of pyrenoids (He *et al.,* 2023), though this approach is far from definitive. Elsewhere, direct evidence for this theory is sparse.

The only example of definitively traced pyrenoid evolution comes from hornworts, where pyrenoids have been shown to have evolved at least 5-6 times independently and have also been lost at least 5-6 times (Villarreal & Renner, 2012). The first hornwort pyrenoids appeared ~100 million years ago in a period that coincided with declining atmospheric CO_2 levels and increasing O_2 levels, consistent with their role. Other younger pyrenoid-containing hornwort clades originated during periods of similar atmospheric conditions and concurrently pyrenoids were apparently lost in hornwort clades during the same period (Villarreal & Renner, 2012). The lack of PYCO1 homologues between *Phaeodactylum*

tricornutum and other pyrenoid-containing diatoms that diverged in the last 90 million years (*e.g., Thalassiosira pseudonana*; ((Bowler *et al.,* 2008) might suggest a similar timepoint for the evolution of diatom pyrenoids. Likewise, the lack of sequence homologues of EPYC1 also constrains the evolution of pyrenoids to prior to the divergence of the earliest *Chlamydomonas* species ~240 Mya (Herron *et al.,* 2009), and provides molecular evidence for the convergent evolution of pyrenoids at least in *Phaeodactylum* and Chlamydomonas. The evolutionary timing of pyrenoids more generally remains an outstanding question, though previous approaches have suggested likely timelines based on environmental factors. Estimates have ranged: ~400 Mya (Griffiths *et al.,* 2017), ~300 Mya (Raven *et al.,* 2017), >360 Mya (Meyer, Goudet, *et al.,* 2020) and 275-400 Mya (Badger & Price, 2003), but are generally consistent with a late-Paleozoic origin for the evolution of pyrenoid-based CCMs.

1.8.2. Pyrenoid structural diversity across different algal lineages

Pyrenoid structure varies greatly between different algal and hornwort species (Fig. 1-6, Fig. 1-9). Common to all pyrenoids is that they consist of a dense Rubisco matrix, which is probably formed through LLPS, as discussed in section 1.7. Whereas most species have only one pyrenoid per chloroplast, some species have two or more (Gibbs, 1962; Hoffman, 1968a; Lokhorst & Star, 1980; Vaucher, 1803), which can even be ultrastructurally distinct (Lokhorst *et al.,* 1988). In many species the pyrenoid is localized centrally in the chloroplast amidst the thylakoids (common in glaucophytes and all green lineages; some rhodophytes and diatoms). In other species the pyrenoid is localized in peripheral protrusions of the chloroplast (common in all TSAR lineages and some rhodophytes). In species with a peripheral pyrenoid, the pyrenoid is tightly encircled by the chloroplast envelope and the protrusion is typically into the central cytosolic space. In many species the pyrenoid is traversed by one or more membranes that typically are continuous with the thylakoid membrane but can be derived from other cellular membranes. The membrane traversions are presumably important for the delivery of inorganic carbon as discussed for Chlamydomonas and postulated in the diatom *Phaeodactylum tricornutum*, whose single pyrenoid membrane traversion contains a carbonic anhydrase (Kikutani *et al.,* 2016).

Although it is currently assumed that the pyrenoid-traversing thylakoid membranes are important for the delivery of inorganic carbon to the pyrenoid, not all pyrenoid containing species have a matrix traversed by thylakoid membranes. In species with peripheral pyrenoids the pyrenoid-traversing thylakoid membranes are often missing. In these species, the thylakoid membrane often stops before the pyrenoid matrix begins (Deane *et al.,* 1998; Evans, 1966; Fresnel & Probert, 2005; Scott *et al.,* 2011) and, in some cases, the thylakoid tips extend into the matrix (Klöpper *et al.,* 2013). However, there are exceptions (*e.g.,* dinophytes) where thylakoid membranes cross the matrix of peripheral pyrenoids completely or even form networks in them (Scott *et al.,* 2006; Zinssmeister *et al.,* 2013). In species with peripheral pyrenoids without thylakoid membrane traversions, the chloroplast envelope can extend into the pyrenoid matrix by forming tubular intrusions, indicating that membranes traversing the pyrenoid is potentially a ubiquitous feature of pyrenoids. In striking examples of such envelope intrusions, the internal region of the pyrenoid seems to be directly connected to the cytosol, nucleus or mitochondria (Jouenne *et al.,* 2011), opening the possibility that photorespiratory CO_2 release could directly be driving photosynthetic carbon fixation. The function of all pyrenoid traversing chloroplast envelope intrusions remain unknown, but it seems likely that they are also involved in CO_2 delivery to the pyrenoid.

Even though different algal lineages use different carbohydrates for energy storage, there are example species from all algal clades that surround their pyrenoid with a layer of their storage material (hereafter referred to as starch sheath). This is even more astonishing considering that rhodophytes and algal lineages that inherited a “red” chloroplast through secondary endosymbiosis store their reserve material not in the chloroplast but in the cytosol. Consequently, only peripheral pyrenoids in “red” chloroplasts exhibit a starch sheath and central pyrenoids are always sheath-less in these lineages. Glaucophytes never have a starch sheath, and in the green lineages the pyrenoid is often, but not always, surrounded by a starch sheath. The morphology of the starch sheath also varies greatly between species (Fig. 1-9). The starch sheath can be formed by only one plate (Deane *et al.,* 1998; Oborník *et al.,* 2011; Van Thinh *et al.,* 1986) and in species where the pyrenoid matrix is crossed by a single thylakoid sheet, the starch sheath is sometimes formed by two plates (Gärtner *et al.,* 2015; McDonald & Pickett-Heaps, 1976), but

in most cases the starch sheath is formed by several plates. In some species there are broad gaps between the starch plates (Mikhailyuk *et al.,* 2018; Nam *et al.,* 2013; Scott *et al.,* 2011), whereas in others the plates sit tightly together, sometimes even in multiple layers (Mikhailyuk *et al.,* 2014).

Differences in pyrenoid structure across algae indicate that inorganic carbon flow from the external environment to Rubisco must differ from the described mechanism for Chlamydomonas. In species with pyrenoids lacking thylakoid membrane traversions, inorganic carbon must enter the pyrenoid matrix directly from the chloroplast stroma without entering the thylakoid lumen or even from the cytosol in the case of peripheral pyrenoids without even entering the chloroplast stroma, perhaps through chloroplast envelope intrusions that extend into the pyrenoid matrix. Species without a starch sheath around the pyrenoid potentially lose more CO_2 through leaking than species with a starch sheath. However, the presence of low electron dense CO_2 impermeable protein layers or membrane diffusion barriers created from adjacent thylakoids or the chloroplast envelope cannot be ruled out. Outside of Chlamydomonas, we currently lack a molecular understanding of the structural arrangement of the pyrenoid across algal lineages and hornworts and, consequently, mostly understand the operation of these CCMs by analogy to Chlamydomonas. Taken together the commonalities and structural variation observed across pyrenoids suggests that although the pyrenoid matrix is likely underpinned by LLPS, the assembly of the matrix with other ultrastructural features is likely orchestrated by different molecular mechanisms, in line with the apparent convergent evolution of pyrenoids. Characterisation of more diverse pyrenoids therefore may offer promising novel approaches for engineering pyrenoids into plants.

1.9. Engineering pyrenoid-based CCMs in plants

In modern agriculture where abiotic and biotic stresses such as water and nitrogen availability, pests and pathogens can be controlled, crop yield is often limited by CO_2 fixation (Ainsworth & Long, 2005). Calculations show that many C_3 crops, such as rice, wheat and soya are only reaching at maximum a third of their theoretical potential of conversion of solar energy capture to carbohydrate synthesis (Zhu *et al.,* 2008), and this results in losses in yield (Walker *et al.,* 2016). It is thought that photosynthetic performance has not been selectively improved through breeding programmes due to it being highly conserved within crop species, giving very little room for positive selection (Long *et al.,* 2019). As discussed in previous sections, a promising strategy for photosynthetic improvements is the engineering of a CCM into C_3 crops (see Adler *et al.,* (2022), Hennacy & Jonikas (2020); Mackinder (2018), Meyer *et al.,* (2016), Rae *et al.,* (2017) for recent detailed reviews). Modelling has shown that biophysical CCM engineering in the form of a pyrenoid- or carboxysome-based CCM could result in theoretical yield increases of 60% along with improvements in water and nitrogen use efficiencies (Long *et al.,* 2019; McGrath & Long, 2014). This is caveated with the disclaimer that predicting yield increases from photosynthetic improvements is complicated due to the interplay of multiple processes that determine crop yield. This has been demonstrated by cross-scale modelling that indicate that simultaneous improvements in Rubisco activity (*i.e.,* CCM presence), electron transport and mesophyll conductance may be required for significant yield improvements (Wu, Hammer, *et al.,* 2019). That said, field data in tobacco supports photosynthetic engineering as a promising approach, with significant increases in plant biomass seen with multiple approaches, including synthetic photorespiratory bypasses to reduce photorespiration (South *et al.,* 2019), enhanced photoprotection (Kromdijk *et al.,* 2016) combined improvements in RuBP regeneration and electron transport (López-Calcagno *et al.,* 2020) and expression of form ID Rubisco (Zhou *et al.,* 2023). Again, how biomass improvements translates to grain crop yields also remains unclear. However, the potential for increasing CO_2 supply to Rubisco through CCM engineering to translate into grain yield improvements is supported to some extent by in field data, where season-long CO_2 enrichment using Free-Air Concentration Enrichment (FACE) technology has demonstrated yield improvements on average of 17% across rice, wheat, cotton and sorghum (Ainsworth & Long, 2005).

Of the ongoing CCM engineering strategies, engineering of a pyrenoid-based CCM remains amongst the most promising and achievable for three key reasons. Firstly, in contrast to C_4- and carboxysome-based engineering approaches which will require previously outlined extensive modifications to each target host plant, no extraneous modifications to the host plant physiology are expected to be required to allow function of a pyrenoid-based CCM in C_3 plants. Secondly, for the same reasons, it is hoped that a "one size fits all" approach can be developed for the pyrenoid-based CCM, such that the same protein blueprints can be implemented in a range of plants with minimal host-specific modifications required. Thirdly, the closer phylogenetic relationship between donor pyrenoid-containing chlorophyte algae and the target host plants (Fig. 1-6), relative to that of carboxysome-containing cyanobacteria and plants, is expected to afford increased compatibility of pyrenoid CCM components in plants. One key drawback of engineering the Chlamydomonas pyrenoid-based CCM is that replacement of the RbcS in each host plant with that of Chlamydomonas is required to allow interaction with EPYC1 and RBM-containing proteins. Crop plants often contain multiple non-homologous copies of the RbcS gene, which are non-trivial to remove given their nuclear encoding in high ploidy genomes (Mao *et al.,* 2022).

Proof of concept engineering of the Chlamydomonas pyrenoid-based CCM is currently ongoing as part of the Combining Algal and Plant and Photosynthesis (CAPP) consortium. In support of the predicted cross-compatibility of algal components, multiple CCM components have been expressed and shown to correctly localize in *Arabidopsis thaliana* (Atkinson *et al.,* 2016). To prime plants for Rubisco condensation by EPYC1, *Arabidopsis* lines that have had the majority of their native Rubisco small subunit replaced with Chlamydomonas RbcS have been developed and the hybrid Rubisco shown to be functional (Atkinson *et al.,* 2017). Purified hybrid *Arabidopsis*/*Chlamydomonas* Rubisco has subsequently been shown to undergo LLPS *in vitro* (Atkinson et al., 2019). Expression of EPYC1 in *Arabidopsis* lines expressing hybrid *Arabidopsis*/Chlamydomonas Rubisco has recently resulted in *in planta* proto-pyrenoid assembly (*i.e.,* EPYC1-Rubisco condensation) (Atkinson *et al.,* 2020), with *Arabidopsis* proto-pyrenoids having a comparable size and internal mixing to cognate *Chlamydomonas* pyrenoids (Freeman Rosenzweig *et al.,* 2017). In contrast to *in planta* carboxysome assembly where Rubisco packaging results in severely reduced plant growth (Long *et al.,* 2018), proto-pyrenoid

expressing lines have a similar photosynthetic performance to wild-type (Atkinson *et al.,* 2020), suggesting condensation of Rubisco by EPYC1 has no detrimental impact on carbon fixation. Although plant proto-pyrenoid assembly is a major breakthrough, photosynthetic improvements most likely will only be realized once a complete CCM is assembled. Based on conserved structural features of pyrenoids across algae and our current knowledge of the Chlamydomonas CCM as well as recent modelling approaches (Fei *et al.,* 2022), a minimum CCM is expected to require: 1) Rubisco condensation around thylakoids; 2) inorganic carbon delivery to the matrix traversing thylakoids via HCO_3^- channels; 3) accelerated dehydration of HCO_3^- to CO_2 via a carbonic anhydrase localized in the matrix traversing thylakoids and 4) a diffusion barrier surrounding the Rubisco condensate (*e.g.,* starch). Further less certain requirements may include modification of thylakoids traversing the pyrenoid matrix, similar to pyrenoid tubule assemblies observed in Chlamydomonas and additional inorganic carbon accumulation systems at the chloroplast envelope (*i.e.,* LCIA) and pyrenoid periphery (*i.e.,* LCIB/LCIC complex). In addition, detailed understanding and control of pyrenoid assembly, regulation, and division within different plant leaf cell-types will be critical for successful function.

A major goal for pyrenoid engineering in plants is to discover or develop a linker protein that can phase-separate plant (specifically crop) Rubiscos without modification. This would overcome the primary drawback of the Chlamydomonas engineering strategy and remove the need for any modification of the native host plant genes. It is also hoped that understanding pyrenoid assembly across diverse algal lineages will offer additional approaches to engineering plants with pyrenoid CCMs. Given the wide ultrastructural diversity of pyrenoids (Fig. 1-9), it is expected that alternative, perhaps simpler mechanisms exist in nature and that through their characterisation alternative pyrenoid assembly proteins may discovered. Identification of the RBMs within these hypothesised proteins would presumably allow ‘mix and matching’ of assembly proteins by replacement of RBMs with those that are putatively plant-compatible. In summary, engineering of pyrenoid-based CCMs in plants is still very much in its infancy and there is huge scope to provide valuable insight for future engineering efforts past the proof-of-concept stage.

1.10. Aims of this study and framework of the thesis

The fundamental hypothesis of this study was that linker proteins commonly underpin the condensation of Rubisco in pyrenoids across the tree of life, as first supposed by Mackinder *et al.* (2016) and described in this introductory chapter. The motivation for the identification and characterisation of these hypothesised proteins was two-fold. The primary motivation was to discover novel approaches to pyrenoid assembly in nature that could benefit ongoing efforts to engineer pyrenoid-based CCMs in plants. Secondary to this goal it was hypothesised that the identification of a wide range of pyrenoid linker and assembly proteins from evolutionary distinct pyrenoid-containing algae would allow the currently uncertain timepoint for the evolution of most pyrenoids to be usefully narrowed, and that the hypothesised convergent evolution of pyrenoids could be molecularly proven or disproven.

Though not originally part of the framework of this study, the COVID-19 lockdowns enforced in the early stages of the study provided a 'unique' opportunity to explore the fundamental hypothesis bioinformatically, which is presented in Chapter 2. Following directly from the bioinformatic characterisations, Chapters 3 and 4 present a detailed molecular and structural characterisation of a linker protein from another unicellular green alga, *Chlorella sorokiniana*. Chapter 5 builds on Chapter 4's structural characterisation of the linker's interaction with Rubisco and presents surprising, yet hugely exciting cross-compatibility of the identified linker protein.

The majority of the results in Chapters 3, 4 and 5 form part of a submitted manuscript entitled 'A promiscuous mechanism to phase separate eukaryotic carbon fixation'. As such, some of the work is shared with co-authors as indicated by disclaimers throughout the thesis, though the overwhelming majority of results were produced and analysed by me. The results are rewritten in a more verbose format to allow for more discussion in the context of literature and provide a personal outlook on future efforts. Results from Chapter 2 are largely unformatted for publication though are part of ongoing characterisations building towards manuscripts of their own.

1.11. Fundamental methodologies

1.11.1. Disorder prediction with IUPred

IUPred is a computational approach that allows the rapid prediction of protein regions that are incompatible with ordered regions (Dosztányi *et al.,* 2005). The speed of the model is facilitated by its basis in a low-resolution statistical approach that is underpinned by simple arithmetic equations derived from an energy predictor matrix that is used to assess the energy of interactions available to a given residue within the sequence. The values defined in the energy predictor matrix were populated based on assessment of the expected pairwise energy of interacting residues in a set of globular proteins with known structures (Mészáros *et al.,* 2018; Thomas & Dill, 1996). Calculation of the predicted energy of each of the residues within the sequence is completed considering the surrounding residues in the local sequence environment (2-100 residues adjacent). Residues with unfavourable energies are predicted disordered and the disorder score is smoothed over a specified window to give a probability of disorder for the defined region of the protein sequence, which is reported as a value between 0 and 1.

1.11.2. Protein structure prediction with AlphaFold

AlphaFold 2 utilises a deep learning approach to predict protein structure from primary sequence (Jumper *et al.,* 2021). Fundamentally the system is comprised of three main parts. Firstly, a deep multiple sequence alignment (MSA) is constructed from the query sequence by comparison against extensive sequence databases. The MSA is the input for the iterative neural network that forms the second main part of the prediction system, denoted the 'Evoformer'. In this module the relation between residues, termed pair representations, within the alignments are assessed and compared with structural hypotheses based on these representations to identify likely interactions within the sequences. Principally, this assessment is based on the observation of co-evolution of stabilising interactions within protein sequences (Lupo *et al.,* 2022). Over the defined number of iterations, the neural network builds a model of predicted interactions within the protein, which are inputs for the final module in the system, the structural module. In addition to the pair representations, the sequence variation captured in the MSA is also an input for the structural module. The structural module also utilises an iterative approach to predict the final three dimensional structure. Initially, all residues are placed at the origin of the

coordinate system and subsequently a set of displacements and rotations are performed to the main chain of the sequence to form the overall structure that satisfies the predicted interactions from the Evoformer module. During the structural modelling, side chains are also positioned based on a conventional implementation of the permitted torsion angles for the residues. In addition to the predicted 3D coordinates of the protein, two further outputs are also obtained as scored metrics; the per-residue confidence, pLDDT, and the topological score, Predicted Aligned Error (PAE). The predicted local-distance difference test (pLDDT) score estimates how well the prediction would agree with an experimentally determined structure of the same sequence. To do so, a local distance difference test is performed on the Cα coordinates (lDDT-Cα) (Mariani *et al.,* 2013) such that the quality of the local atomic environment of the model is assessed. In this sense, the pLDDT score represents the confidence in the prediction of regions of substructure within the prediction but does not penalise deviations from alignment of the overall 3D structure. In contrast, the PAE score assesses the positional error based on the relative position of two residues within the predicted structure when aligned on a particular residue. PAE is therefore useful for assessing the overall 3D structure and accordingly the relative position and orientation of domains within.

1.11.3. Molecular clock analyses and phylogenetic chronograms

Molecular clock analyses are used to assess evolutionary rates of sequences over time which can be used as inputs to construct phylogenetic chronograms which inform on the time estimates of the divergence points of lineages (Langley & Fitch, 1974). Principally, molecular clock analyses function by assessing the mutation rate by comparison of homologous sequences in organisms that have time-calibrated divergence points in order to define evolutionary rates for those sequences. In most instances absolute time calibration is performed against fossil records, though other calibrations including for geological and biogeographic events can be performed (Ho & Duchêne, 2014). Early molecular clock analyses assumed homogeneous evolutionary rates of proteins across lineages, referred to as ‘strict’ analyses. The accuracy of strict molecular clock analyses has since been disproven (dos Reis *et al.,* 2016), and ‘relaxed’ analyses, utilising a Bayesian approach to integrate differences in evolutionary rates and uncertainty in calibration timepoints and tree topologies are now more commonplace

(Wertheim *et al.,* 2010). The output of these analyses are used to draw phylogenetic chronograms, in which branch lengths are defined in units of time and nodes are positioned with confidence estimates based on the Bayesian predictions.

1.11.4. Surface Plasmon Resonance (SPR)

The phenomena of Surface Plasmon Resonance (SPR) can be used to measure the interaction of biomolecules at a sensor surface. SPR occurs when electrons in the metallic sensor surface are excited by incident polarised light, creating plasmon waves that reduce the intensity of the reflected light at a particular angle, referred to as the resonance angle. The resonance angle is proportionate to the mass on the sensor that determines the refractive index of the surface. As such, changes in the resonance angle can be used to detect the mass change that results from the interaction of biomolecules at the sensor surface.

1.11.5. Single particle cryogenic electron microscopy (cryo-EM)

Cryogenic electron microscopy (cryo-EM) is rapidly developing as a powerful tool for determining the 3D structure of proteins in solution (Callaway, 2015). In cryo-EM, samples immobilised on grids are vitrified by rapid plunging into liquid ethane, which prevents the formation of crystalline ice and maintains the native state of proteins in the sample. The vitrified samples are subsequently subjected to Transmission Electron Microscopy (TEM) to capture two dimensional images of the protein from many different angles that are ultimately used to reconstruct a 3D volume of the protein (Henderson *et al.,* 1990). In TEM, electrons are accelerated in a high vacuum towards the sample, which gives rise to their elastic and inelastic scattering as they traverse and interact with the sample. The scattered electrons as well as the unscattered electrons that pass through the samples are subsequently focussed by electromagnetic lenses onto an electron detector. The interference of the unscattered electrons and the elastically scattered electrons, which have a known scattering angle, give rise to phase contrast at the electron detector. The characteristic phase contrast is directly influenced by the focus, and as such different defocus values result in different phase contrast images of the same sample at the detector. A Fourier transform is required to convert the digital measurement of the phase contrast image at the

electron detector to a reconstructed 2D projection of the sample. Following correction for beam-induced motion of the sample during imaging, Contrast Transfer Function (CTF) correction is performed to model the constructive, destructive and inverting interference in the phase contrast image. Practically, this is performed by estimation of the CTF in each of the micrographs (Rohou & Grigorieff, 2015; Zhang, 2016), and also accounts for aberrations in the imaging process, such as spherical aberration of the focussing electromagnetic lenses.

Following the manual or automated picking of single particles representing the species of interest, 2D images of different orientations of the sample are grouped by classification and erroneous particles are removed from the dataset. To reconstruct a 3D volume of the sample from the 2D images that represent information for central Fourier slices of the object, the orientations of the randomly oriented particles within the 2D images must then be assigned. Since direct assessment of the projection angles from the 2D images is precluded by the signal to noise ratio, CTF and particle picking errors, most 3D reconstruction methods rely on expectation maximisation approaches that can be completed *de novo* or with the assistance of a 3D reference model (Scheres, 2012; Zivanov et al., 2018). These methods use an iterative approach in which projection vectors are assigned, a 3D model is reconstructed from the assigned vectors and a comparison between the 2D image information, and the 3D model is performed. This process is repeated until the model and assigned vectors converge. Multiple rounds of 3D reconstruction and particle selection can be performed in a process termed 3D classification, in order to further refine particle selection or distinguish structural heterogeneity within the sample. Depending on the processing pipeline used, a further series of refinement, polishing and correction procedures can be performed to maximise the resolution of the 3D reconstruction (Punjani *et al.,* 2017; Zivanov *et al.,* 2018). To assess the quality and report a resolution of the final reconstruction the "gold-standard" Fourier Shell Correlation (FSC) curve of the map is most commonly used (Henderson *et al.,* 2012). The gold-standard FSC is the correlation between two 3D volumes reconstructed from independent halves of the data (half maps) and the resolution at which the FSC passes through 0.143 is most commonly reported as the estimated resolution.

2. A FAST LINKER IDENTIFICATION PIPELINE FOR PYRENOIDS

2.1. Introduction

Although pyrenoids exist across vast evolutionary taxa (He *et al.,* 2023; Meyer *et al.,* 2017), and have been hypothesised to be commonly underpinned by liquid-liquid phase separation (Barrett *et al.,* 2021), only two pyrenoid linker proteins have so far been identified (Mackinder *et al.,* 2016; Oh *et al.,* 2023). In line with the apparent commonality of LLPS in pyrenoids, it has previously been hypothesised that linker proteins are also commonly required for the condensation of Rubisco within pyrenoids (Mackinder *et al.,* 2016). The main limitation for the identification of hypothesised additional linker proteins lies in the supposed lack of sequence conservation between the linker proteins of different algal lineages. In support of this, default NCBI BLASTp searches of PYCO1 only return the PYCO1 sequence from *Phaeodactylum tricornutum*, and those of EPYC1 return sequences in species exclusive to the Chlamydomonadales order (Table 2-1). Further to this, the biochemical investment to identify both EPYC1 and PYCO1 was extensive in both cases, with authors completing multiple rounds of mass spectrometry (Mackinder *et al.,* 2016; Oh *et al.,* 2023). Mackinder *et al.,* (2016) previously attempted to shortcut these extensive characterisations by bioinformatically identifying functional analogues of EPYC1 with conserved properties, namely: a high isoelectric point, tandem repeats, a repeating disorder profile and lack of transmembrane domains. Though stringent filtering of input sequences was achieved using this approach, no hits were observed in pyrenoid-containing *Chlorella variabilis* or *Micromonas pusilla* species and there is no support for the role of the identified proteins in *Thalassiosira pseudonana*, *Phaeodactylum tricornutum* and *Emiliania huxleyi* (Mackinder *et al.,* 2016), suggesting the approach had considerable limitations. The aim of this chapter was accordingly to develop a bioinformatic pipeline inspired by the work of Mackinder *et al.,* (2016) to identify functional analogues of EPYC1 more accurately in pyrenoid-containing species.

Table 2-1: NCBI BLASTp hits for PYCO1 and EPYC1 with XSTREAM identified repeat number.

Query	Hit Species	E value	%Identity	XSTREAM Repeats	Accession
PYCO1	*Phaeodactylum tricornutum*	0	97.8	4.52	XP_002184702.1
EPYC1	*Chlamydomonas reinhardtii*	0	100	4.39	XP_001690584.2
EPYC1	*Chlamydomonas incerta*	4E-103	85.5	5.03	KAG2432954.1
EPYC1	*Chlamydomonas schloesseri*	5E-100	84.3	4.93	KAG2448775.1
EPYC1	*Pleodorina starrii*	3E-69	52.9	4.81	GLC41192.1
EPYC1	*Volvox reticuliferus*	4E-69	49.1	2.83	GIL71809.1
EPYC1	*Volvox africanus*	2E-66	59.7	4.28	GLI63967.1
EPYC1	*Volvox carteri*	3E-65	56.3	4.72	XP_002946604.1
EPYC1	*Edaphochlamys debaryana*	2E-51	56.6	3.69	KAG2501276.1
EPYC1	*Astrephomene gubernaculifera*	4E-45	62.6	5.32	GFR42532.1
EPYC1	*Gonium pectorale*	5E-44	66.3	3.15	KXZ46518.1
EPYC1	*Tetrabaena socialis*	9E-31	54.2	2.87	PNH11430.1

2.2. Construction and testing of a Fast Linker Identification Pipeline for Pyrenoids (FLIPPer)

Development of the bioinformatic pipeline was commenced in early 2020, prior to the presentation or publication of the pyrenoid linker protein in *Phaeodactylum tricornutum* (Phaeodactylum hereafter); PYCO1 (Oh *et al.,* 2023). Accordingly, all the initial development was completed based on the properties of the single confirmed pyrenoid linker protein at the time, belonging to Chlamydomonas; EPYC1. Hypotheses about the EPYC1 protein structure and function during the initial development of the pipeline were also respective of the available publications at the time of development but are discussed here in the context of current understanding.

2.2.1. Primary sequence analysis of EPYC1 and sequence pre-filtering

To determine an appropriate strategy and which programs to use for sequence filtering, an initial analysis of the EPYC1 primary sequence was completed. As noted by Mackinder *et al.,* (2016) the isoelectric point of the EPYC1 protein is exceptionally high (pI = 11.9) (Fig. 2-1), which determines a net positive charge of the protein in the pH ~8 stroma of the illuminated chloroplast where the pyrenoid resides (Fig. 1-3) (Heldt *et al.,* 1973; Werdan *et al.,* 1975). Prior to publication of the 3D structure of the EPYC1-Rubisco interaction (He *et al.,* 2020), the positive charge of EPYC1 and net negative charge of Chlamydomonas Rubisco (pI = 6.28) in the stroma were hypothesised to contribute to the interaction between the proteins and possibly underpin a complex coacervation mechanism for liquid-liquid phase separation (Wunder *et al.,* 2018). Given the largely conserved sequence and consequent isoelectric point of Rubiscos across phylogenetic taxa (Liu *et al.,* 2017), it was presumed the net negative surface charge of Rubiscos in the stroma of pyrenoid-containing algae is also conserved. The isoelectric point of hypothesised linker proteins was accordingly presumed to have to exceed the pH of the illuminated chloroplast stroma also to determine a net positive charge and allow interaction with Rubisco. This requirement was easily and efficiently integrated in the first pre-filtering step in the bioinformatic pipeline using the ProtParam module of the Biopython package (Fig. 2-3) (Cock *et al.,* 2009; Gasteiger *et al.,* 2005).

Table 2-2: EPYC1 primary sequence analysis.

	% of full length EPYC1 sequence																						
pI	Structure	Disorder	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
11.9	13	42	20	0	2	3	0	3	0	1	5	4	1	3	10	2	9	27	4	4	3	0	

In line with predictions (Mackinder *et al.,* 2016), EPYC1 has previously been shown to be a largely disordered protein in solution using circular dichroism spectroscopy of the purified protein (Wunder *et al.,* 2018). Prior to the release of AlphaFold for accurate structure prediction (Jumper *et al.,* 2021), PSIPRED v3.3 (Buchan *et al.,* 2013) and Phyre2 (Kelley *et al.,* 2015) were used by Mackinder *et al.,* (2016) to predict secondary structure within the disordered EPYC1 sequence. 4 predicted helices of 9 and 12 amino acids were present in each of the predictions respectively, representing 11% and 15% of the total amino acids in each case. The predicted helices were subsequently shown to contribute to the interaction of EPYC1 with Chlamydomonas Rubisco in yeast two-hybrid (Y2H) experiments at the time (Atkinson et al., 2019), a fact which has been demonstrated structurally since (He *et al.,* 2020). In line with secondary structure predictions, ProtParam analysis of the residue composition in the EPYC1 sequence indicated 13% of the sequence was hydrophobic and structure-promoting (Phe, Iso, Leu, Val, Trp, Tyr; (Pace & Scholtz, 1998)) (Table 2-2), and this was therefore used to roughly determine the fraction of each sequence likely involved in structural elements. The unstructured, disordered regions between the helical regions of EPYC1 were presumed of equal importance to the function of EPYC1 by allowing extension of adjacent helices over the length-scale required for Rubisco cross-linking (3-5 nm) (Engel *et al.,* 2015; Mackinder *et al.,* 2016). In agreement with the disordered nature of EPYC1, 42% of the EPYC1 primary sequence was comprised of disorder-promoting residues (Pro, Asn, Gly, Ser). Based on the assumption that both the structured and unstructured regions are essential to the function of EPYC1, a requirement that the sequences contain at least an equal percentage of disorder-promoting and structure-promoting residues was imposed in the pre-filtering step. The proportion of Serine and Alanine residues within EPYC1 was also notably high (Table 2-2), due to their presence in the disordered spacer regions of EPYC1 (Fig. 2-2). A final sequence pre-filtering step ensured the presence of these amino acids in the sequences (Fig. 2-3). Following these sequence pre-filtering steps,

~25% (7,679) of the 32,672 input sequences from the Chlamydomonas genome (*Chlamydomonas reinhardtii* CC-4532 v6.1) remained. Of the pre-filtering steps, filtering by isoelectric point had the largest impact on the number of remaining sequences (60% of the total 75% filtering efficiency), in line with the binodal distribution of protein isoelectric points in the Chlamydomonas genome (Fig. 2-1). Filtering by the ratio of structure- to disorder-promoting residues had less of an effect (12%), with Serine and Alanine filtering having little effect (<2% each). The sequence pre-filtering module was implemented using Biopython sequence handling and analysis, with subsequent compositional filtering completed using pandas dataframes.

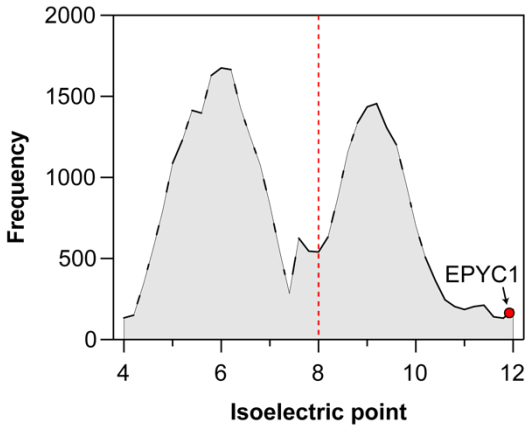


Figure 2-1 | Distribution of protein isoelectric points in the Chlamydomonas genome.
Histogram of the frequency of theoretical isoelectric points (pIs) of proteins in the Chlamydomonas genome. Data was binned in 0.2 pH units from pH 4.0 to pH 12.0. The pI of EPYC1 is indicated by the red dot.

2.2.2. Primary sequence repeat detection in EPYC1 and application to the pipeline

The multivalency afforded by the 4 quasi-identical and single C-terminal partial repeats of EPYC1 is essential to its function in the condensation of Rubisco (Wunder *et al.,* 2018), and was accordingly an important property of linker sequences to capture in the pipeline. In line with the distance between adjacent Rubiscos in the pyrenoid (Engel *et al.,* 2015), the repeats of EPYC1 are ~60 amino acids in length (Fig. 2-2), which has been shown to allow the required distance between Rubiscos to be spanned with minimal energetic cost (He *et al.,* 2020). To identify appropriately scaled repeats in the sequences, three repeat detection programs were initially compared. XSTREAM (Newman & Cooper, 2007), RADAR (Heger & Holm, 2000) and HHrepID (Biegert & Söding, 2008) all correctly identified 5 repeats of ~60 amino acids within the EPYC1 sequence (Fig. 2-2). Due to the hidden Markov model

methodology implemented by HHrepID, this program was comparatively much slower than XSTREAM and RADAR which both use seed-based approaches to rapidly remove non-repeating sequences. Although RADAR and XSTREAM were comparable in terms of accuracy and speed, XSTREAM was developed for large scale analysis and thus more appropriately suited in terms of output and optimisation to the analysis of the thousands of input sequences resulting from the sequence pre-filtering module. Using python handling, the output of XSTREAM could also be easily manipulated for use in the downstream pipeline (Fig. 2-3).

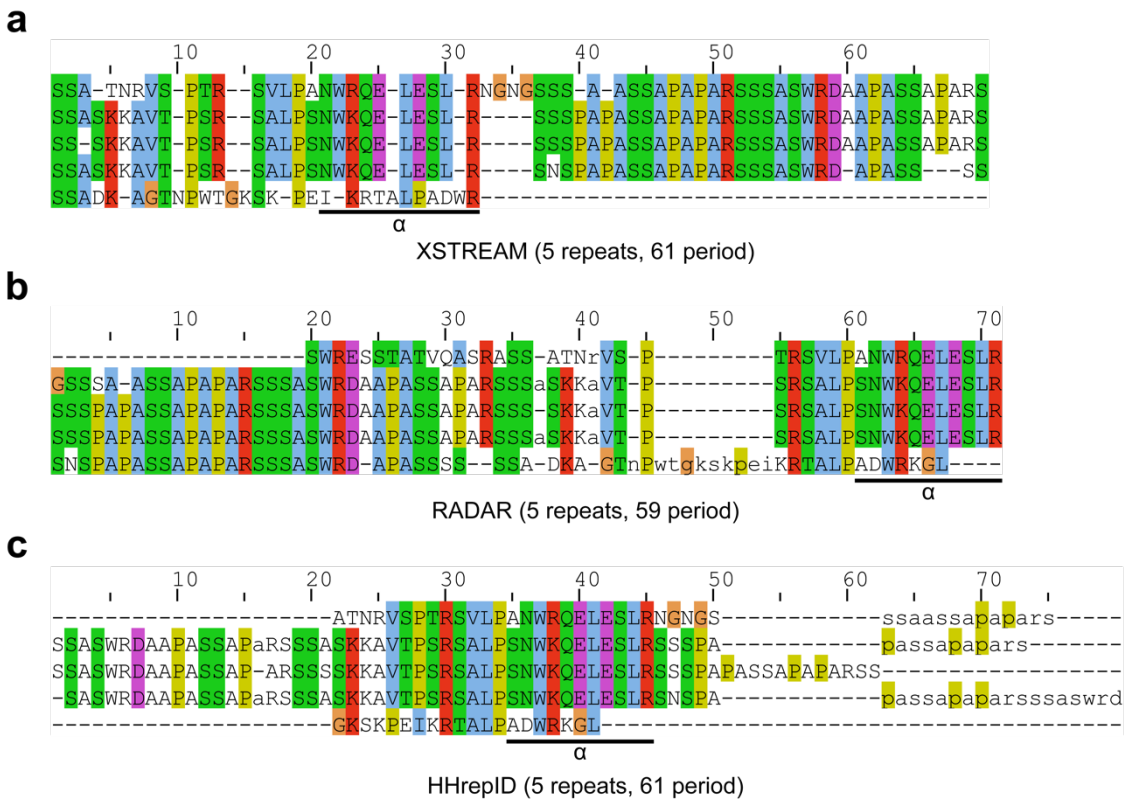


Figure 2-2 | Comparison of repeat detection software.
(a) Multiple sequence alignment of repeats detected by XSTREAM (see methods). (b) Alignment of tandem repeats detected by RADAR web tool run with default settings. (c) Alignment of tandem repeats detected by HHrepID web tool run with default settings. In each alignment the predicted helical region is indicated (α) and residues are coloured by property according to the Clustal X colour scheme

XSTREAM also allows a large degree of parameterisation to adjust the properties of the tandem repeats that are detected. As such, some of the default values were adjusted based on a set of pre-formed hypotheses. Whilst EPYC1 demonstrates near identical repeats of the same approximate length (Fig. 2-2), it was hypothesised that there may be variation in other sequences either due to genuine biological

differences, or errors introduced from genome assembly or protein annotation which is more challenging with low complexity repeat sequences (Peona *et al.,* 2021). Specifically, it was hypothesised the number of repeat regions was likely to be variable, which was supported by XSTREAM analysis of the EPYC1 homologues (Table 2-1). To account for this, the minimum number of complete repeat regions required for identification was set to 3, with no upper limit. Similarly, it was hypothesised other sequences may not maintain the high degree of repeat-level conservation observed for EPYC1 (Fig. 2-2), and accordingly the maximum number of permitted gaps in each repeat was increased to 50 (from a default value of 3). The minimum consensus and repeat match thresholds, which represent the percentage similarity of each repeat to the consensus repeat sequence and of each repeat to the others respectively, were also decreased to 40% (from a default value of 70%). It was also hypothesised that the total repeat length would exhibit some variance. As such, a lower threshold of 20 amino acids for the repeat length was set according to the minimum spacing observed between Rubiscos in the pyrenoid (2 nm), and the minimum feasible number of amino acids to bridge this length scale. An upper repeat length threshold of 120 amino acids was arbitrarily set based on double the length of the EPYC1 repeat length. Lastly, a requirement that the identified tandem repeats represent over 70% of the input sequences was enforced, based on the assumption that *bona fide* linker proteins would not be fused to other non-linker domains, but would have a cleaved chloroplast transit peptide that accounts for <30% of the total sequence length. XSTREAM analysis of the 7,679 pre-filtered Chlamydomonas sequences with these settings resulted in filtering to 0.9% (294) of the 32,672 input sequences. Crucially these sequences included EPYC1, in line with the previous performance of XSTREAM used under the same settings to detect EPYC1 repeats (Fig. 2-2). Whilst this represented a significant reduction from the number of input sequences, the high number of remaining sequences would be unfeasible to screen manually in *de novo* studies, and as such an additional filtering step was pursued.

2.2.3. Disorder assessment

At this point in the bioinformatic pipeline, protein sequences have been filtered to be net positively charged, contain some structure and some disorder, and contain greater than 3 repeat regions of 20-120 amino acids in length. Arguably these properties could all be possessed by globular proteins, should the

primary sequence be arranged appropriately. It was therefore necessary to introduce a filtering step to explicitly predict local secondary structure and filter for overall disorder of the protein sequences. Ideally this step would be completed as a whole sequence structural prediction using tools such as Phyre2 or at the time unpublished, AlphaFold2, though the time and computational expense of these approaches was unfeasible for the medium throughput application required here. Alternative 3D structure-independent approaches were instead pursued.

IUPred allows the rapid assessment of local disorder likelihood based on an energy estimation that characterises the likelihood of interaction between pairs of residues within a sequence (Dosztányi *et al.,* 2005). The algorithm employed by IUPred is particularly well suited to linker proteins as it is optimised for the prediction of disordered regions greater than 30 amino acids, consistent with the length of the disordered region interspaced between the helices of EPYC1 (Fig. 2-2). The approach of IUPred contrasts other, less appropriate disorder prediction software which are more suited to the identification of short regions of disorder within globular proteins (Kozlowski & Bujnicki, 2012; Linding *et al.,* 2003; Romero *et al.,* 2001; Walsh *et al.,* 2012). Accordingly, IUPred2A was employed for rapid disorder assessment of each of the protein sequences following XSTREAM repeat detection. The average IUPred2A score of the proteins was then filtered to a threshold of greater than 0.5, corresponding to >50% of the sequence being disordered. This filtering step reduced the number of sequences to 160 (~0.5%), from the 32,672 input sequences. This point represented the first iteration of the pipeline, here termed Fast Linker Identification Pipeline for Pyrenoids (FLIPPer) v0 (beta), which was built in a shell environment and was tested extensively in Windows environments but took significant porting to run successfully in UNIX environments. The time for analysis of the Chlamydomonas genome was ~5 minutes (M2 MacBook Air, 16GB RAM), which largely consisted of XSTREAM repeat detection of the large number of input sequences and highlighted the need for selection of an efficient repeat detection software. Whilst the performance with the Chlamydomonas genome was far from the desired level, it was appropriate at this point to determine the efficacy against the genomes of other pyrenoid-containing species.

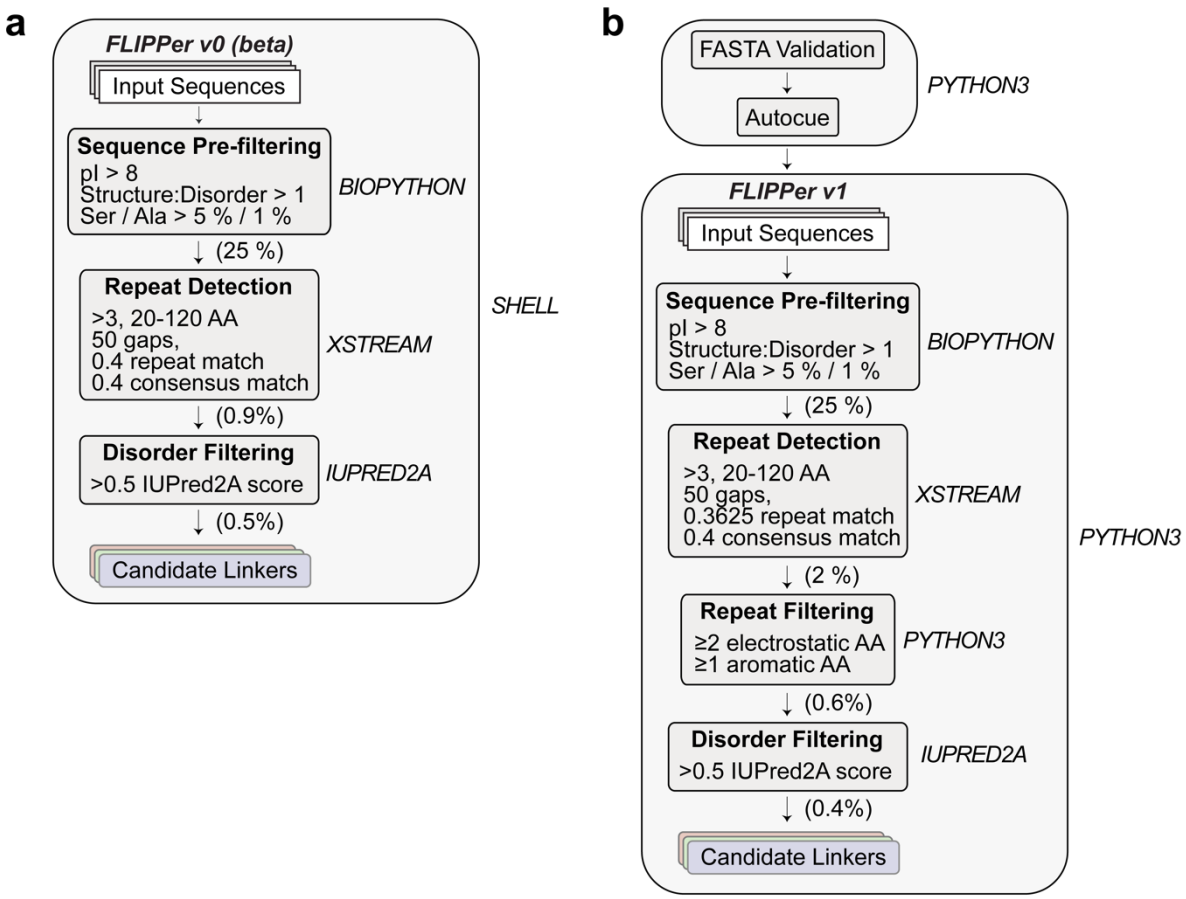


Figure 2-3 | Fast Linker Identification Pipeline for Pyrenoids (FLIPPer) v0 (beta).
(a) Schematic representation of v0 (beta) of FLIPPer. The general process completed at each stage is indicated in bold in each box, with the program used in italics adjacent. The key parameters for each program or filtering step are indicated in the boxes also, where other settings were left as default. The filtering percentage at each stage, relative to the number input sequences is shown based on the statistics of the *Chlamydomonas* genome. (b) Schematic representation of FLIPPer v1, labelled as in Fig. 2-3a.

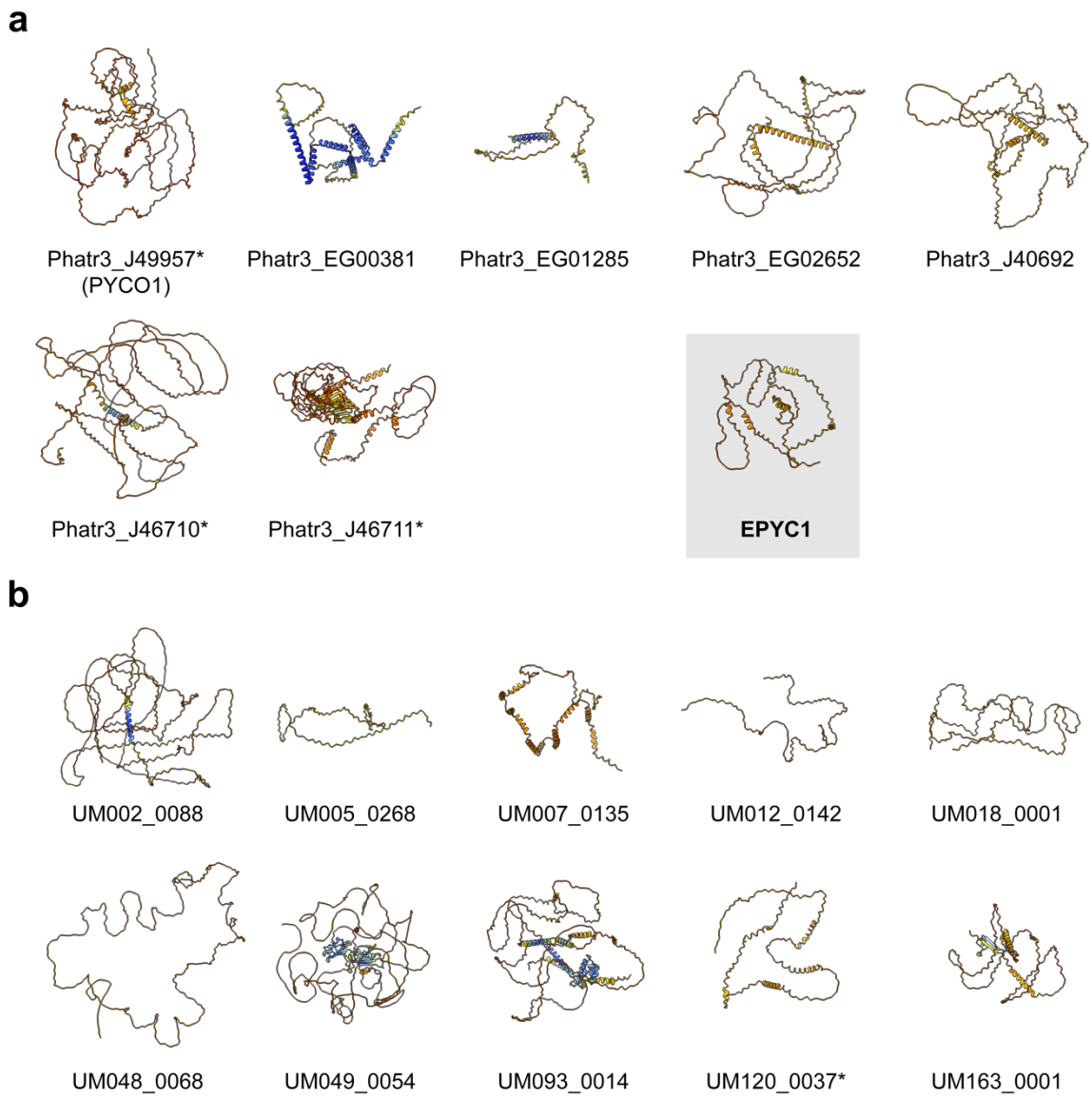
2.2.4. Initial trials of FLIPPer v0

Following the optimisation of FLIPPer v0 parameters with EPYC1 and the Chlamydomonas sequences, FLIPPer v0 was applied to the well-annotated genomes of a selection of pyrenoid-containing species. Surprisingly, the speed and selectivity of the pipeline was much greater for all non-Chlamydomonas genomes, indicating the Chlamydomonas genome was not a representative benchmark for the performance of the pipeline (Table 2-3). Besides the larger number of input sequences, the high average sequence length of the pre-filtered sequences in the Chlamydomonas genome appeared to have a large impact on the speed of the analysis, likely due to the seeded approach XSTREAM uses. The lower number of output sequences from the non-Chlamydomonas genomes was much more amenable to more intensive *in silico* characterisations. Accordingly, AlphaFold2 modelling of the candidate sequences

demonstrated that the output sequences largely contained the desired properties consistent with those of EPYC1 (*i.e.,* largely disordered, spaced repeating structural elements) (Fig. 2-4). Together these encouraging results demonstrated the efficacy of FLIPPer v0 in identification of EPYC1-like proteins.

Table 2-3: FLIPPer v0 benchmarks of genomes with high confidence linker candidates.

Species	Time (s)	Average Input Length	#Inputs	#Candidates	%Selectivity
Chlamydomonas reinhardtii	~300	783 AA	32,672	160	0.49
Phaeodactylum tricornutum	5	430 AA	12,178	7	0.06
Chlorella sorokiniana	23	524 AA	12,871	24	0.19
Ulva mutabilis	20	462 AA	12,924	10	0.08



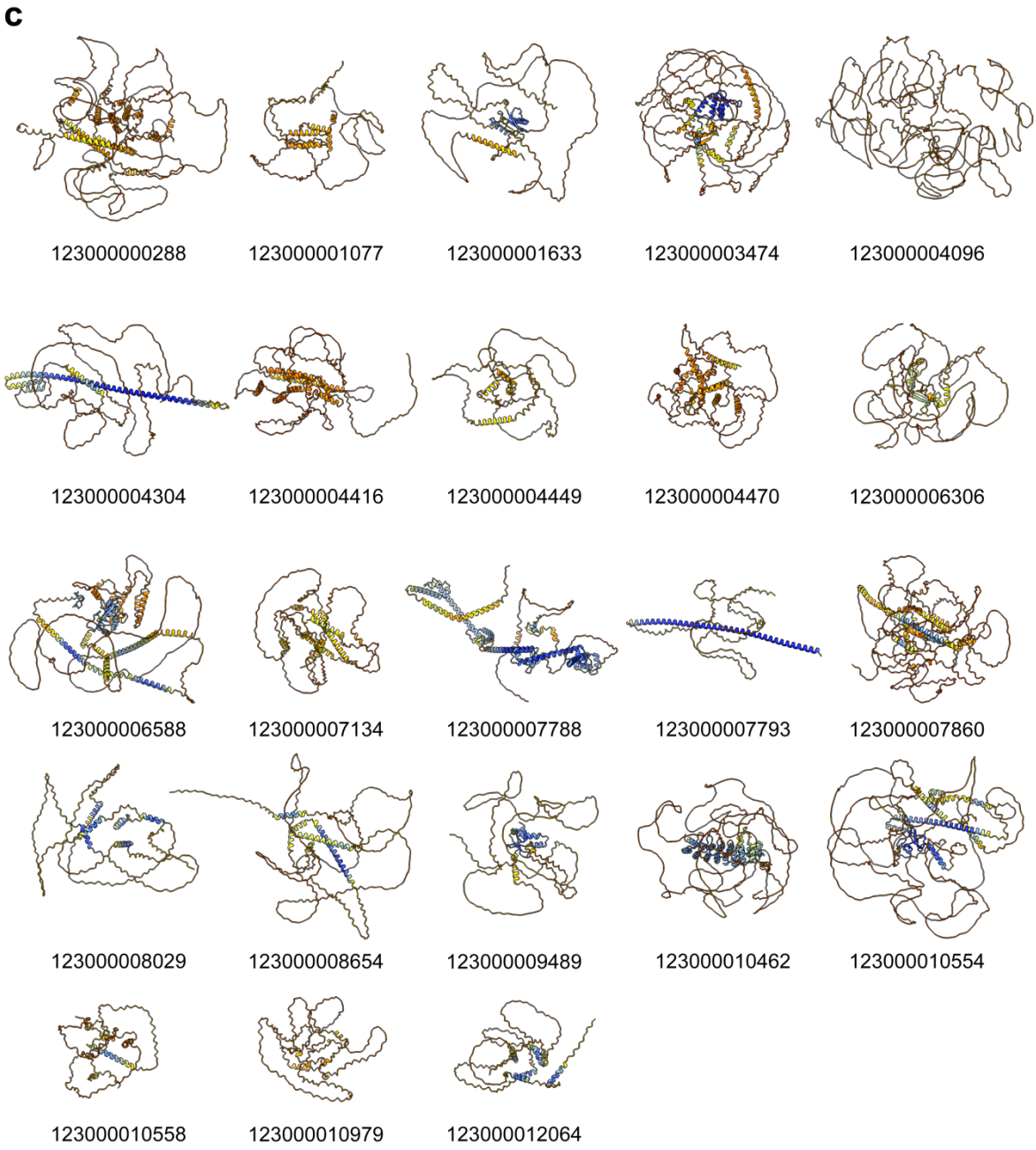


Figure 2-4 | AlphaFold2 models of candidate linker proteins identified by FLIPPer v0.
(a) Top ranked AlphaFold2 models of the 7 candidate proteins from the *Phaeodactylum tricornutum* v3 genome. The AlphaFold2 model of EPYC1 is also shown for comparison. The models are coloured according to per-residue pLDDT scoring, with blue indicating high and red indicating low scores. The gene names are shown below each model, with starred names indicating proteins discussed further in text. (b) Top ranked AlphaFold2 models of the 10 candidate proteins from the *Ulva mutabilis* genome, coloured as in Fig. 2-4a. UM120_0037 is the later-characterised putative linker sequence, and indicated by the asterisk. (c) Top ranked AlphaFold2 models of the 23 candidate proteins from the *Chlorella sorokiniana* UTEX1230 genome, coloured as in Fig. 2-4a.

2.2.5. Repeat filtering addition to FLIPPer v1

During the development of FLIPPer v0, preliminary data from the characterisation of the Phaeodactylum linker PYCO1 was generously shared by the authors at the 10th International Symposium on Inorganic Carbon Utilization by Aquatic Photosynthetic Organisms (CCM10). Encouragingly PYCO1 (Phatr3_J49957) was amongst the 7 candidate proteins identified by FLIPPer v0 analysis of the Phaeodactylum genome (Table 2-3 and Fig. 2-4a). The subsequently published characterisation of PYCO1 interaction with Rubisco (Oh *et al.,* 2023) allowed further comparison of the properties of the linker proteins from Chlamydomonas and Phaeodactylum. Enforcing the accuracy of FLIPPer v0, the proteins possessed largely similar isoelectric points, fraction of structure- and disorder-promoting residues, repeat length and number, and overall disordered structure (Table 2-4, Fig. 2-4a). 3D structures for both the Chlamydomonas linker EPYC1 (He *et al.,* 2020) and Phaeodactylum linker PYCO1 (Oh *et al.,* 2023) and their interactions with Rubisco were previously determined by cryogenic Electron Microscopy (cryo-EM), which highlighted the identity of the Rubisco-interacting motifs of the linker proteins. Comparison of the consensus Rubisco-interacting motifs from EPYC1 (**W**KQELESL**R**) and PYCO1 (K**W**SP**R**) indicated they both contain at least 2 electrostatic residues and an aromatic residue (Table 2-4). As such, an additional step was added to FLIPPer to filter for proteins in which the XSTREAM identified consensus repeat contains at least 2 electrostatic residues and 1 aromatic residue (Fig. 2-3b). Notably, this approach differs from sequence pre-filtering approaches that would identify the total number of electrostatic or aromatic residues in the sequence, and instead filters for the presence of these residues in the identified repeat sequences that likely contribute to the interaction with Rubisco. This filtering step was implemented in tandem with a slight modification to the XSTREAM consensus repeat match value (from 0.4 in v0 to 0.3625 in v1), based on further repeat detection testing (data not shown). Finally, a major overhaul was completed in which FLIPPer was rebuilt in a python3 environment to allow seamless use in both Windows and UNIX environments (Fig. 2-3b). During this overhaul an additional FASTA format validation module and file autocue module were implemented to allow high throughput usage of the pipeline, as well as a command line user interface to allow customisation of parameters within the program (Fig. 2-5). This marked the generation of FLIPPer v1 which was used for all subsequent characterisations. The performance of

FLIPPer v1 was improved over v0, with the selectivity increasing for both the Chlamydomonas (0.5% [v0] to 0.4% [v1]) and Phaeodactylum (0.06% [v0] to 0.04% [v1]) genomes whilst retaining identification of the *bona fide* linker proteins. The effect on the speed of operation of the pipeline was negligible, given the low computational expense of the implemented filtering steps in v1.

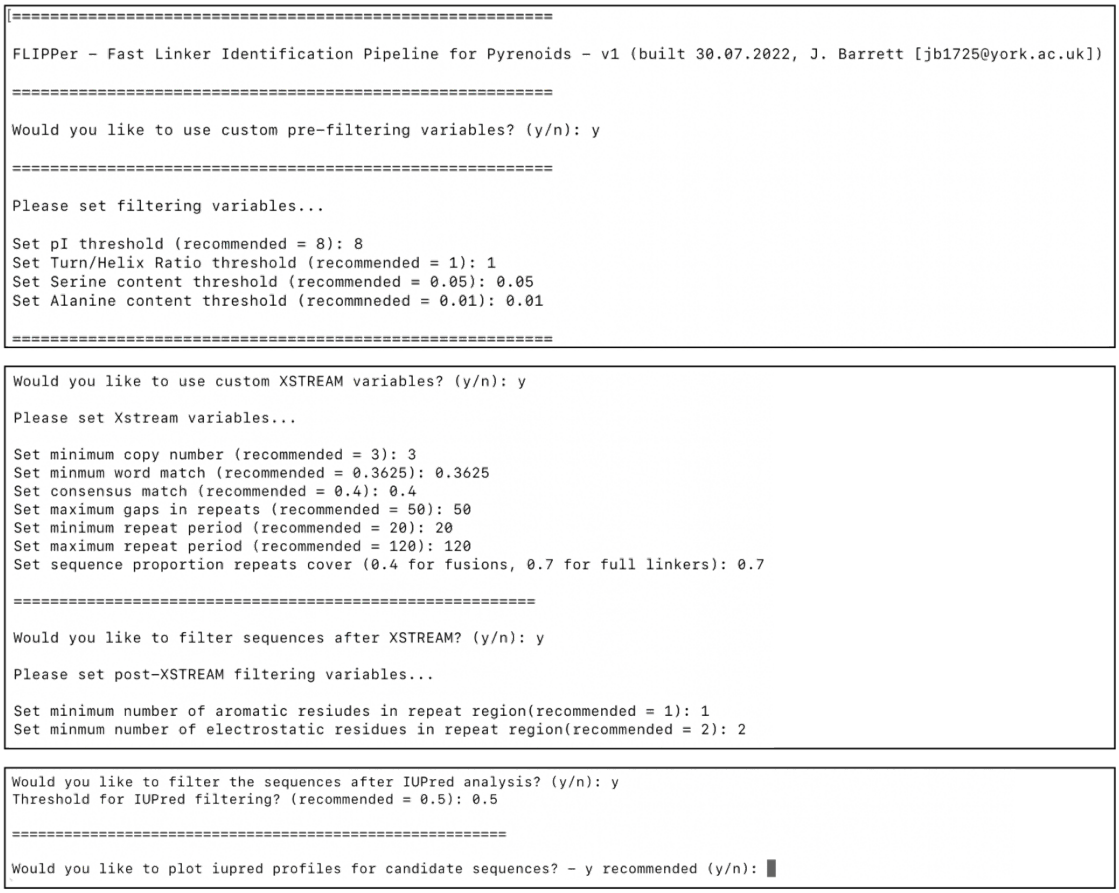


Figure 2-5 | Command line user interface for FLIPPer v1.
Screenshots from the operation of FLIPPer v1 using the command line interface for parameterisation of each stage of the pipeline. This setup phase is completed prior to the FASTA validation and autocue module shown in Fig. 2-3b and applied to all the subsequent analyses for each run. The recommended settings are shown in the interface, though the pipeline can be run in default mode with all the recommended settings and no user input.

Table 2-4: Comparison of EPYC1 and PYCO1 properties.

		% of full length sequence		XSTREAM analysis		# in interacting motif	
Linker	pI	Structure	Disorder	Repeat number	Repeat Length	Electrostatic	Aromatic
EPYC1	11.9	13	42	4.39	61	4	1
PYCO1	10.0	13	52	5.69	70	2	1

2.3. Implementation of FLIPPer v1 and validation of candidate proteins

Following the construction and parameterisation of FLIPPer v1, it was then aimed to implement the pipeline to discover novel pyrenoid linker proteins in the genomes of other pyrenoid-containing species. Experimental validation of the output of some of these analyses was also pursued in this work, as well as with the support of collaborators, for which gratuity is extended. Initially, annotated genomes of algal species were collated, and a literature review was performed to validate the presence of a pyrenoid in each species (Table 2-5). FLIPPer v1 was subsequently applied to the 96 available annotated genomes that represented coverage of a wide range of phylogenetic taxa. The selectivity of the pipeline was largely consistent with the small scale trial of FLIPPer v1 completed against the Phaeodactylum genome, with an average of 0.09% ± 0.13% S.D. of the input sequences outputted as candidates, leaving an average of 20 sequences per genome for further analysis. For each of the outputted candidate sequences, the XSTREAM identified repeats and IUPred disorder profile were manually interrogated to determine the likelihood of each sequence as pyrenoid linker proteins. A selection of the discoveries from implementation of FLIPPer v1 are discussed in the remainder of this chapter, as well as in the remainder of the thesis.

Table 2-5: FLIPPer v1 analysis of the annotated genomes of pyrenoid-containing species. The right 4 columns show the filtering efficiency after the indicated stage, relative to the number of input sequences.

Species and strain	Reference for pyrenoid presence	Class	Genome Source	Pre-filtered	XSTREAM Repeats	Filtered Repeats	IUPred Candidates
Pyropia yezoensis U-51	(Kim *et al.,* 2016)	Rhodaphyta	JGI	8.99%	0.15%	0.15%	0.08%
Porphyridium purpureum CCMP1328	(Nelson & Ryan, 1988)	Rhodaphyta	JGI	9.41%	0.02%	0.02%	0.02%
Rhodosorus marinus	(Krayesky-Self *et al.,* 2020)	Rhodaphyta	NCBI	6.74%	0.00%	0.00%	0.00%
Cyanophora paradoxa CCMP329	(Price *et al.,* 2019)	Glaucophyta	JGI	28.08%	2.97%	2.97%	1.18%
Prasinoderma coloniale	(Jouenne *et al.,* 2011)	Prasinodermophyta	JGI	9.60%	0.27%	0.27%	0.10%
Astrephomene gubernaculifera 4017	(Stein, 1958)	Chlorophyceae	JGI	23.13%	0.91%	0.26%	0.17%
Chlamydomonas eustigma NIES-2499	(Hirooka *et al.,* 2017)	Chlorophyceae	JGI	11.99%	0.07%	0.03%	0.00%
Chlamydomonas incerta SAG 7.73	(Pröschold & Darienko, 2018)	Chlorophyceae	JGI	20.25%	2.18%	0.52%	0.34%
Chlamydomonas priscuii UWO241	(Possmayer *et al.,* 2016)	Chlorophyceae	JGI	16.44%	0.82%	0.38%	0.19%
Chlamydomonas reinhardtii CC-4532	(Lacoste-Royal & Gibbs, 1987)	Chlorophyceae	JGI	23.50%	2.08%	0.61%	0.41%
Chlamydomonas schloesseri 11/173	(Pröschold *et al.,* 2018)	Chlorophyceae	JGI	22.07%	2.27%	0.55%	0.35%
Desmodesmus armatus UTEX B 2533	(Matusiak-Mikulin *et al.,* 2006)	Chlorophyceae	JGI	12.33%	0.54%	0.54%	0.17%
Dunaliella salina CCAP19/18	(Polle *et al.,* 2020)	Chlorophyceae	JGI	20.76%	0.51%	0.51%	0.15%
Edaphochlamys debaryana 11/70	(Yoshitomi *et al.,* 2020)	Chlorophyceae	JGI	20.59%	1.83%	0.55%	0.29%
Enallax costatus CCAP 276/31	(Cardon *et al.,* 2018)	Chlorophyceae	JGI	12.34%	0.41%	0.41%	0.06%
Gonium pectorale NIES-2863	(Arakaki *et al.,* 2013)	Chlorophyceae	JGI	13.56%	0.52%	0.21%	0.12%
Microglena sp. YARC	(Demchenko *et al.,* 2012)	Chlorophyceae	JGI	15.69%	0.40%	0.40%	0.14%
Monoraphidium minutum 26B-AM	(Krienitz *et al.,* 2001)	Chlorophyceae	JGI	26.11%	2.74%	2.74%	0.68%
Monoraphidium neglectum SAG 48.87	(Krienitz *et al.,* 2001)	Chlorophyceae	JGI	16.94%	0.80%	0.80%	0.36%
Pleodorina starii NIES-4479	(Nozaki *et al.,* 2006)	Chlorophyceae	JGI	22.47%	1.25%	0.35%	0.24%
Raphidocelis subcapitata NIES-35	(Pröschold *et al.,* 2011)	Chlorophyceae	JGI	22.42%	2.72%	2.72%	1.01%
Scenedesmus obliquus EN0004	(Miyachi *et al.,* 1986)	Chlorophyceae	JGI	13.90%	0.73%	0.73%	0.17%
Scenedesmus obliquus var. DOE0013	(Miyachi *et al.,* 1986)	Chlorophyceae	JGI	13.04%	0.41%	0.41%	0.08%

Scenedesmus obliquus var. UTEX2630	(Miyachi *et al.,* 1986)	Chlorophyceae	JGI	13.65%	0.42%	0.42%	0.11%
Scenedesmus obliquus UTEX 393	(Miyachi *et al.,* 1986)	Chlorophyceae	JGI	17.07%	0.89%	0.89%	0.14%
Scenedesmus obliquus UTEX 3031	(Miyachi *et al.,* 1986)	Chlorophyceae	JGI	13.03%	0.38%	0.38%	0.18%
Scenedesmus obliquus UTEX 1450	(Miyachi *et al.,* 1986)	Chlorophyceae	JGI	14.21%	0.54%	0.54%	0.13%
Tetrabaena socialis NIES-571	(Arakaki *et al.,* 2013)	Chlorophyceae	JGI	20.46%	0.79%	0.33%	0.15%
Tetradesmus obliquus UTEX 72	(Ahiahonu *et al.,* 2022)	Chlorophyceae	JGI	14.72%	0.52%	0.52%	0.07%
Volvox africanus NIES-4468	(Kochert & Olson, 1970)	Chlorophyceae	JGI	20.56%	0.48%	0.14%	0.10%
Volvox carteri	(Kochert & Olson, 1970)	Chlorophyceae	JGI	25.03%	0.44%	0.20%	0.11%
Volvox reticuliferus NIES-3785	(Kochert & Olson, 1970)	Chlorophyceae	JGI	22.70%	0.55%	0.13%	0.11%
Bryopsis sp. KO-2023	(Satoh *et al.,* 1985)	Ulvophyceae	NCBI	10.95%	0.15%	0.15%	0.09%
Caulerpa lentillifera	(Miyamura *et al.,* 1996)	Ulvophyceae	JGI	5.78%	0.02%	0.02%	0.02%
Ulva mutabilis Føyn	(Katsaros *et al.,* 2017)	Ulvophyceae	JGI	14.69%	0.35%	0.35%	0.12%
Asterochloris sp. Cgr/DA1pho T	(Friedl, 1989)	Trebouxiophyceae	JGI	7.38%	0.13%	0.13%	0.08%
Botryococcus braunii Race B	(Suzuki *et al.,* 2013)	Trebouxiophyceae	JGI	16.33%	0.35%	0.35%	0.11%
Chlorella ohadii	(Treves *et al.,* 2016)	Trebouxiophyceae	NCBI	10.58%	0.28%	0.28%	0.13%
Chlorella sorokiniana DOE1412	(Azaman *et al.,* 2017)	Trebouxiophyceae	JGI	19.02%	0.98%	0.98%	0.33%
Chlorella sorokiniana UTEX 1230	(Azaman *et al.,* 2017)	Trebouxiophyceae	JGI	14.56%	0.64%	0.64%	0.25%
Chlorella sorokiniana Str1228	(Azaman *et al.,* 2017)	Trebouxiophyceae	JGI	15.68%	0.71%	0.71%	0.27%
Chlorella variabilis NC64A	(Hoshina *et al.,* 2010)	Trebouxiophyceae	NCBI	11.24%	0.72%	0.72%	0.26%
Chlorella variabilis NC64A	(Hoshina *et al.,* 2010)	Trebouxiophyceae	JGI	9.49%	0.33%	0.33%	0.10%
Chlorella vulgaris 211/11P	(Němcová & Kalina, 2000)	Trebouxiophyceae	NCBI	14.72%	0.41%	0.41%	0.19%
Helicosporidium sp. ATCC 50920	(Yaman & Radek, 2005)	Trebouxiophyceae	NCBI	7.86%	0.41%	0.41%	0.18%
Micractinium conductrix SAG 241.80	(Hoshina *et al.,* 2010)	Trebouxiophyceae	JGI	13.11%	0.49%	0.49%	0.19%
Tetraselmis striata	(Hori *et al.,* 1986)	Chlorodendrophyceae	JGI	13.10%	0.32%	0.32%	0.21%
Micromonas commoda NOUM17	(van Baren *et al.,* 2016)	Mameillophyceae	JGI	13.28%	0.43%	0.43%	0.19%
Micromonas pusilla CCMP1545	(van Baren *et al.,* 2016)	Mameillophyceae	JGI	15.97%	0.45%	0.45%	0.24%
Micromonas pusilla CCMP1545	(van Baren *et al.,* 2016)	Mameillophyceae	ORCAE	14.04%	0.43%	0.43%	0.20%
Micromonas sp. AD1	(van Baren *et al.,* 2016)	Mameillophyceae	JGI	15.54%	0.06%	0.06%	0.00%

Chlorokybus atmophyticus	(Gartner & Ingolic, 1989)	Chlorokybophyceae	JGI	11.41%	0.05%	0.05%	0.03%
Mesostigma viride CCAC1140	(Manton & Ettl, 1965)	Mesostigmatophyceae	JGI	16.87%	0.28%	0.28%	0.10%
Klebsormidium nitens NIES-2285	(Mikhailyuk *et al.,* 2014)	Klebsormidiophyceae	JGI	13.84%	0.18%	0.18%	0.06%
Closterium sp. NIES-67	(Weber, 1965)	Zygneomophyceae	NCBI	19.99%	0.64%	0.64%	0.16%
Mesotaenium endlicherianum	(Busch & Hess, 2022)	Zygneomophyceae	JGI	15.42%	0.20%	0.20%	0.10%
Mesotaenium kramstae NIES-657	(Busch & Hess, 2022)	Zygneomophyceae	JGI	14.71%	0.32%	0.32%	0.11%
Zygnema cf. cylindricum SAG 698	(Feng *et al.,* 2021)	Zygneomophyceae	JGI	7.48%	0.13%	0.13%	0.06%
Zygnema circumcarinatum SAG 698	(Feng *et al.,* 2021)	Zygneomophyceae	JGI	21.12%	0.27%	0.27%	0.07%
Z.circumcarinatum UTEX 1559	(Feng *et al.,* 2021)	Zygneomophyceae	JGI	22.27%	0.20%	0.20%	0.10%
Z. circumcarinatum UTEX 1560	(Feng *et al.,* 2021)	Zygneomophyceae	JGI	20.51%	0.14%	0.14%	0.04%
Baffinella frigidus CCMP2293	(Daugbjerg *et al.,* 2018)	Cryptophyta	JGI	21.03%	0.76%	0.76%	0.21%
Guillardia theta	(Deschamps *et al.,* 2006)	Cryptophyta	JGI	9.19%	0.14%	0.14%	0.08%
Chrysochromulina parva Lackey	(Manton & Leedale, 1961)	Haptophyta	JGI	12.49%	0.42%	0.42%	0.00%
Chrysochromulina tobin CCMP291	(Manton & Leedale, 1961)	Haptophyta	JGI	8.28%	0.32%	0.32%	0.23%
Emiliania huxleyi	(Uwizeye *et al.,* 2021)	Haptophyta	JGI	16.40%	0.53%	0.53%	0.14%
Phaeocystis antarctica CCMP 1374	(Decelle *et al.,* 2019)	Haptophyta	JGI	27.39%	1.07%	1.07%	0.46%
Phaeocystis globosa	(Decelle *et al.,* 2019)	Haptophyta	JGI	25.38%	1.68%	1.68%	0.49%
Bigelowiella natans CCMP2755	(Hirakawa & Ishida, 2015)	Chlorachniophyta	ENSMBL	16.94%	0.15%	0.15%	0.11%
Bigelowiella natans CCMP2755	(Hirakawa & Ishida, 2015)	Chlorachniophyta	JGI	11.96%	0.24%	0.24%	0.19%
Chaetoceros tenuissimus NIES-3715	(Kimura & Tomaru, 2015)	Bacillariophyta	JGI	7.75%	0.19%	0.19%	0.06%
Cyclotella cryptica CCMP332	(Hoops & Floyd, 1979)	Bacillariophyta	JGI	12.44%	0.42%	0.42%	0.14%
Fistulifera solaris JPCC DA0580	(Sunaga *et al.,* 2015)	Bacillariophyta	JGI	5.73%	0.08%	0.08%	0.04%
Fragilariopsis cylindrus	(von Quillfeldt, 2001)	Bacillariophyta	JGI	10.16%	0.15%	0.15%	0.08%
Nitzschia inconspicua GAI-293	(Drum, 1963)	Bacillariophyta	JGI	9.66%	0.07%	0.07%	0.05%
Nitzschia putrida NIES-4239	(Drum, 1963)	Bacillariophyta	JGI	9.36%	0.15%	0.15%	0.06%
Phaeodactylum tricornutum	(Tachibana *et al.,* 2011b)	Bacillariophyta	ENSMBL	8.93%	0.07%	0.07%	0.04%
Pseudo-nitzschia multiseries CLN-47	(Lundholm & Moestrup, 2002)	Bacillariophyta	JGI	12.83%	0.19%	0.19%	0.08%
Seminavis robusta D6	(Aspar, 2013)	Bacillariophyta	JGI	11.07%	0.20%	0.20%	0.07%

Thalassiosira oceanica CCMP1005	(Nawaly *et al.,* 2023)	Bacillariophyta	ENSMBL	26.16%	0.19%	0.19%	0.08%
Thalassiosira pseudonana 1335	(Nawaly *et al.,* 2023)	Bacillariophyta	ENSMBL	8.79%	0.12%	0.12%	0.09%
Ectocarpus siliculosus	(Evans, 1966)	Phaeophyta	JGI	20.01%	0.45%	0.45%	0.26%
Ectocarpus siliculosus	(Evans, 1966)	Phaeophyta	ORCAE	25.84%	0.57%	0.57%	0.32%
Monodopsis C141	(Santos & Leedale, 1995)	Eustigmatophyta	JGI	11.89%	0.14%	0.14%	0.00%
Monodopsis C73	(Santos & Leedale, 1995)	Eustigmatophyta	JGI	11.88%	0.16%	0.16%	0.00%
Vischeria C74	(Gärtner *et al.,* 2012)	Eustigmatophyta	JGI	9.90%	0.11%	0.11%	0.00%
Ochromonas sp. CCMP1393	(Andersen, 2011)	Chrysophyta	JGI	9.52%	0.47%	0.47%	0.09%
Paraphysomonas imperforata 1604	(O'Kelly *et al.,* 2003)	Chrysophyta	JGI	6.69%	0.08%	0.08%	0.03%
Aureococcus anophagefferens 1984	(Sieburth *et al.,* 1988)	Pelagophyta	JGI	5.86%	0.24%	0.24%	0.11%
Vitrella brassicaformis CCMP3155	(Oborník *et al.,* 2012)	Chromerida	JGI	12.28%	0.05%	0.05%	0.04%
Breviolum minutum	(Shinzato *et al.,* 2014)	Dinophyta	JGI	8.63%	0.10%	0.10%	0.04%
Cladocopium goreaui SCF055-01	(Yang, Wei, *et al.,* 2023)	Dinophyta	JGI	5.62%	0.05%	0.05%	0.03%
Polarella glacialis CCMP1383	(Montresor *et al.,* 1999)	Dinophyta	JGI	13.09%	0.34%	0.34%	0.16%
Polarella glacialis CCMP2088	(Montresor *et al.,* 1999)	Dinophyta	JGI	13.11%	0.35%	0.35%	0.17%
Symbiodinium microadriaticum 2467	(Kevin *et al.,* 1969)	Dinophyta	JGI	7.76%	0.10%	7.76%	0.10%
Averages				**14.4±5.6%**	**0.5±0.6%**	**0.18±0.2%**	**0.09±0.13%**

2.3.1. *Robust sequence-independent identification of EPYC1 homologues in the Chlamydomonadales using FLIPPer v1*

Conventional homology BLAST approaches previously identified direct sequence homologues of EPYC1 in closely related species of the Chlamydomonadales order (Table 2-1). Subsequent XSTREAM analysis demonstrated a degree of variability in the linker repeat number of the EPYC1 homologues which was factored during the construction of the pipeline (Fig. 2-3b). FLIPPer v1 analysis of the same Chlamydomonadales genomes identified all the BLAST identified EPYC1 homologues (Table 2-1, Table 2-5) and demonstrated the robustness of FLIPPer in the sequence-independent identification of linker proteins despite variation in their repeat number and structure (Fig. 2-6). In addition to the varying number of repeating predicted helical RBMs (3-6), the repeat length also varied from 44 to 84 amino acids, reinforcing the need for leniency in parameterisation of XSTREAM repeat detection within the pipeline. The selectivity of FLIPPer v1 against the genomes of the Chlamydomonadales species was of the lowest in the 96 genomes analysed (Table 2-5; 0.22 % ± 0.11 % S.D.), in line with previous observations of FLIPPer v0 analysis of the Chlamydomonas genome (Table 2-2). Clearly, identification of the EPYC1 linker homologues from the moderate number of candidates identified by FLIPPer v1 analysis of the non-Chlamydomonas genomes was facilitated by the sequence homology with EPYC1 of *Chlamydomonas reinhardtii*, but nevertheless the presence of the *bona fide* linkers in the candidate sequences was encouraging for the implementation of FLIPPer in unexplored genomes. The greater selectivity of FLIPPer v1 against other genomes also suggested great promise for analysis of these other genomes.

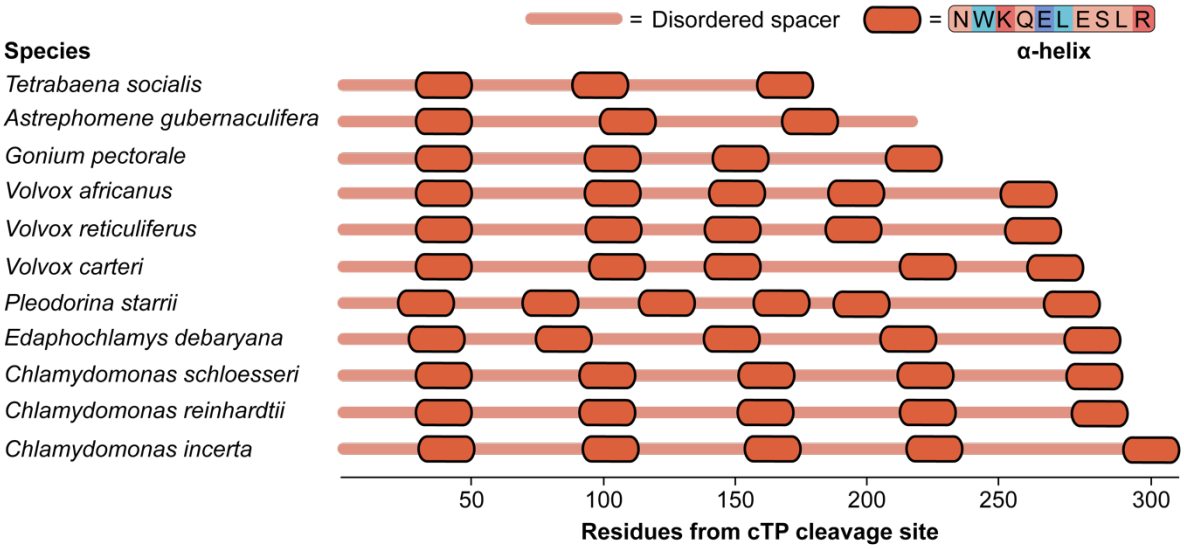


Figure 2-6 | EPYC1 homologues identified in the Chlamydomonadales using FLIPPer v1.
Scaled schematic representation of EPYC1 homologues identified by FLIPPer v1 in species of the Chlamydomonadales. The sequences are shown following removal of the chloroplast transit peptide based on alignment with the *Chlamydomonas reinhardtii* sequence (EPYC1), and prior experimental work (Atkinson *et al.,* 2019). The helices are shown as oblongs with the consensus sequence indicated. The helices are shown linked by disordered spacer regions.

2.3.2. In silico characterisation of Phaeodactylum linker candidate proteins

As mentioned, FLIPPer v1 showed greater selectivity against the Phaeodactylum genome than FLIPPer v0, with only 4 candidate sequences outputted from v1 (Table 2-2, Table 2-5), including the sequence of PYCO1. As expected, the 6 RbcS-interacting 'KWSPR' motifs in the PYCO1 sequence were identified and aligned correctly by XSTREAM (Fig. 2-7b). In addition to the RbcS-interacting KWSPR motif, the 3D structure of a fragment of PYCO1 in complex with Phaeodactylum Rubisco previously demonstrated that a second motif ('AAEWGSMNQ') within PYCO1 is responsible for a separate interaction with the Rubisco large subunit (RbcL) (Fig. 2-7a,d) (Oh *et al.,* 2023). Interestingly, one of the 3 other candidate proteins identified by FLIPPer v1 (Phatr3_J46710) contains 3 motifs with a high degree of sequence similarity to the RbcL-interacting motif of PYCO1 (Phatr3-J49957) (Fig. 2-7d). Like PYCO1, the motif in Phatr3_J46710 were separated at a length-scale feasible for Rubisco cross-linking as demonstrated by AlphaFold2 modelling (Fig. 2-7c). Default NCBI BLASTp analysis of the Phatr3_J46710 sequence identified a homologous protein (Phatr3_J46711), which was also identified by FLIPPer v0 analysis (Fig. 2-4a), in addition to a non-homologous protein from *Plasmodium coatneyi*. Both Phatr3_J46710 and Phatr3_J46711 showed properties largely consistent with that of PYCO1 (Fig.

2-7a,c), suggesting they may also play roles in assembly of the Phaeodactylum pyrenoid and certainly warrant further investigation. It is possible these proteins could play redundant roles in the assembly of the Phaeodactylum pyrenoid. In the absence of functional demonstration or abundance quantification of PYCO1 *in vivo* it is difficult to rule out that other proteins such as those identified here could also be contributing to Rubisco condensation. This discovery further enforced the selectivity of FLIPPer and potential for discovery of novel pyrenoid assembly proteins in the genomes of other species.

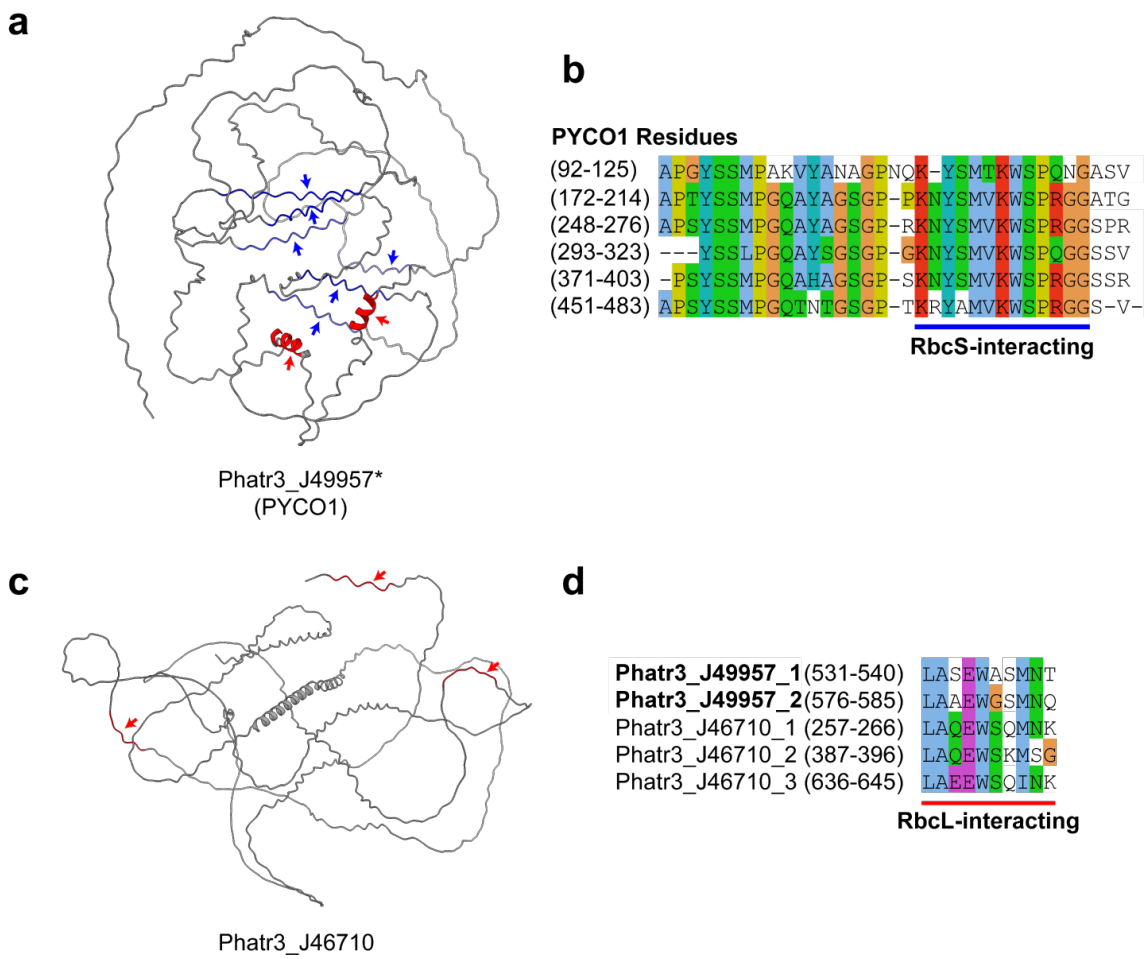


Figure 2-7 | *In silico* characterisation of Phaeodactylum FLIPPer v1 candidates.
(a) Top ranked AlphaFold2 model of PYCO1. RbcS-interacting motifs identified by Oh *et al.,* (2023) are shown in blue and indicated by blue arrows. The RbcL-interacting motifs, as shown in Fig. 2-7d are shown in red and indicated by red arrows. (b) Alignment of the repeat sequences identified by XSTREAM analysis of the protein sequence. The RbcS-interacting motif is indicated by the blue bar and residues are coloured by property according to the Clustal X colour scheme (c) Top ranked AlphaFold2 model of Phatr3_46710. RbcL-interacting motifs are indicated in red, with corresponding red arrows. (d) Alignment of the RbcL-interacting motifs identified in Phatr3_J49957 (PYCO1) and Phatr3_J46710. The RbcL-interacting motif is indicated by the red bar.

2.3.3. *Characterisation of a candidate Porphyridium linker protein*

The characterisation of EPYC1 and PYCO1 have demonstrated analogous mechanisms of condensing form IB and form ID Rubiscos respectively (Oh *et al.,* 2023; Wunder *et al.,* 2018), suggesting that analogous mechanisms may be widespread in pyrenoid-containing species across plastids of red and green origins. *Porphyridium* is a genus of unicellular red algae that lies phylogenetically between *Phaeodactylum* and *Chlamydomonas* as a member of the Archaeplastida that possesses form ID ('red') Rubisco (Fig. 1-6) (Oh *et al.,* 2022; Pichard *et al.,* 1997). Akin to the pyrenoid of Phaeodactylum, the pyrenoid of *Porphyridium* is not surrounded by a starch sheath as observed in Chlamydomonas, owing to the lack of starch in the plastids of red origin (Viola *et al.,* 2001). Contrastingly, the pyrenoid of *Porphyridium* is traversed by a complex network of thylakoids that also appear to delimit its boundary (Fig. 1-9). FLIPPer v1 analysis of the *Porphyridium purpureum* CCMP1328 genome initially returned only a single sequence that was not a convincing linker analogue. When the threshold for the sequence covered by XSTREAM identified repeats was lowered from 70% to 67.5% a second candidate protein was identified. AlphaFold2 modelling of the sequence (KAA8498780) showed a repeating structural motif that was consistent with the 10 repeating motifs within the sequence (Fig. 2-8a-c). The repeats had an average length of 44 residues with a range of 34-54. Two additional C-terminal helices which did not contain the repeated motif were predicted in the AlphaFold2 model of KAA8498780, explaining the lack of detection with the 70% sequence coverage threshold. The second helix had a high probability of being transmembrane as assessed by TMHMM 2.0 (Fig. 2-8d), perhaps suggesting a role in interaction with the traversing and surrounding thylakoid membranes. Default NCBI BLASTp analysis of the KAA8498780 sequence revealed 2 additional proteins in the *Porphyridium purpureum* CCMP1328 genome, with no additional sequence homologues outside *Porphyridium* (Table 2-6), consistent with the distributions of EPYC1 and PYCO1 (Table 2-1).

Table 2-6: BLAST results for the putative *Porphyridium* linker.

Query	Hit Species	E value	%Identity	Accession
KAA8498780	*Porphyridium purpureum*	0	100.0	KAA8498780
KAA8498780	*Porphyridium purpureum*	1E-25	34.9	KAA8499344
KAA8498780	*Porphyridium purpureum*	9E-20	52.1	KAA8498779

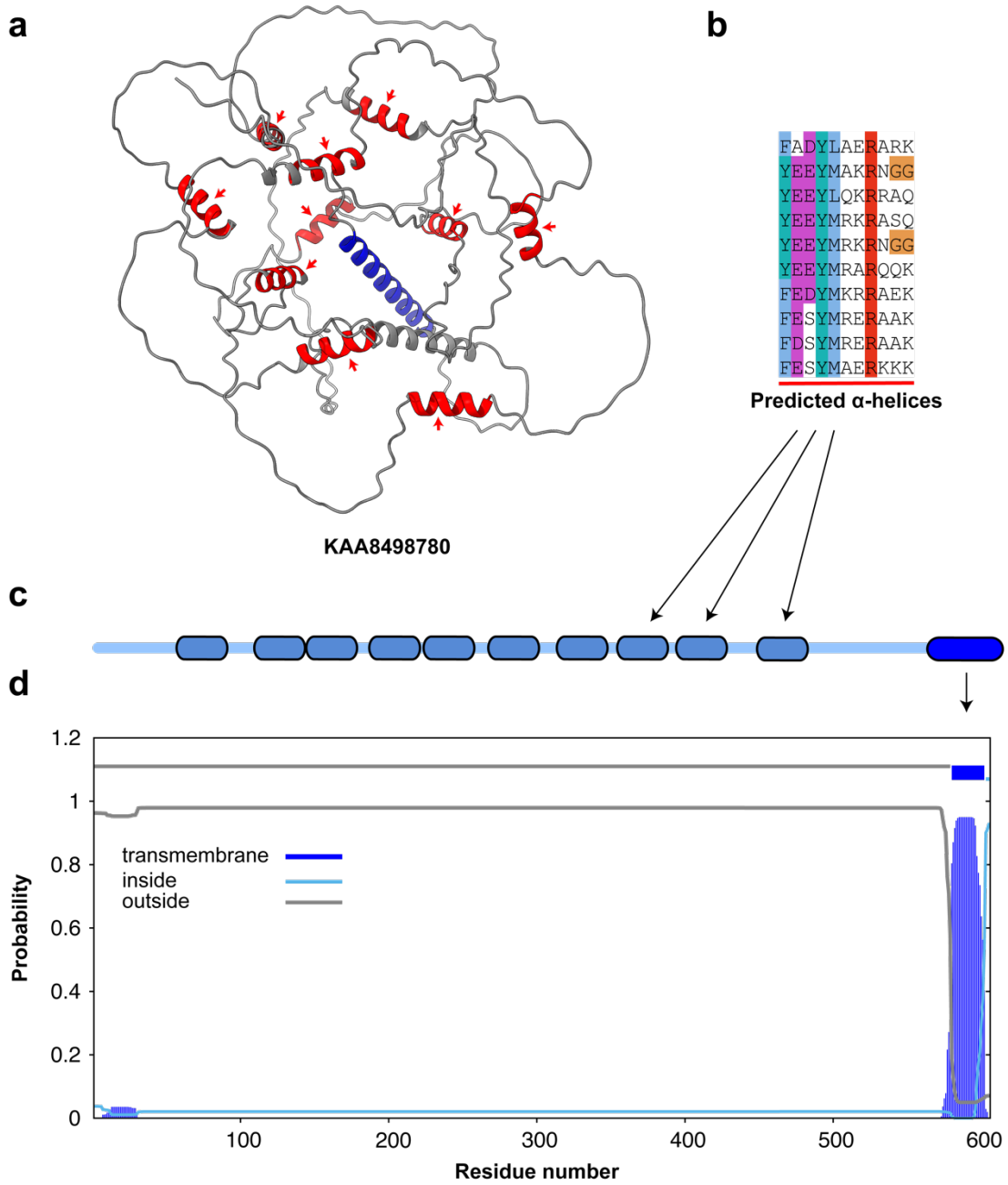


Figure 2-8 | *In silico* characterisation of the *Porphyridium* linker candidate.
a) Top ranked AlphaFold2 model of KAA8498780. The predicted α-helical regions are shown in red and indicated by red arrows. The predicted transmembrane helix is shown in blue. (b) Alignment of repeated α-helical regions in the KAA8498780 sequence. Residues are coloured by property according to the Clustal X colour scheme. (c) Scaled schematic representation of the full length KAA8498780 protein. Pale blue oblongs represent the repeating predicted helical regions, connected by disordered sequence. The predicted transmembrane helix is shown in dark blue at the C-terminus. (d) TMHMM 2.0 prediction of the sequence, the predicted transmembrane is shown in blue.

Whilst the validation of this protein as a *bona fide* linker in the *Porphyridium* pyrenoid is very much in its infancy, some evidence for its role is presented for clarity with permission from Sabina Musiał (University of York) who completed the work. To validate interaction of the native putative linker protein with Rubisco, co-immunoprecipitation (co-IP) mass spectrometry experiments were completed on the lysate of *Porphyridium purpureum* cells using an antibody raised to a surface exposed region of the Rubisco large subunit (RbcL) (Fig. 2-9). The co-IP experiment demonstrated enrichment of both the Rubisco large subunit (RbcL) and Rubisco small subunit (RbcS), confirming pulldown of the intact holoenzyme. Both the chloroplast- and nuclear-encoded Rubisco activase proteins (CbbX(C) and CbbX(N) respectively) were also enriched, in line with their essential role in reactivating Rubisco (Mueller-Cajar *et al.,* 2011). The putative *Porphyridium* linker protein was enriched to a similar degree as both Rubisco and CbbX, consistent with complex formation between the two proteins *ex vivo* and increasing the likelihood of the hypothesised role of the protein as a *bona fide* linker protein. Further work to confirm the localisation of the protein *in vivo*, as well as the interaction between the putative linker and Rubisco *in vitro* should be completed to validate the role of KAA8498780 as a pyrenoid linker protein.

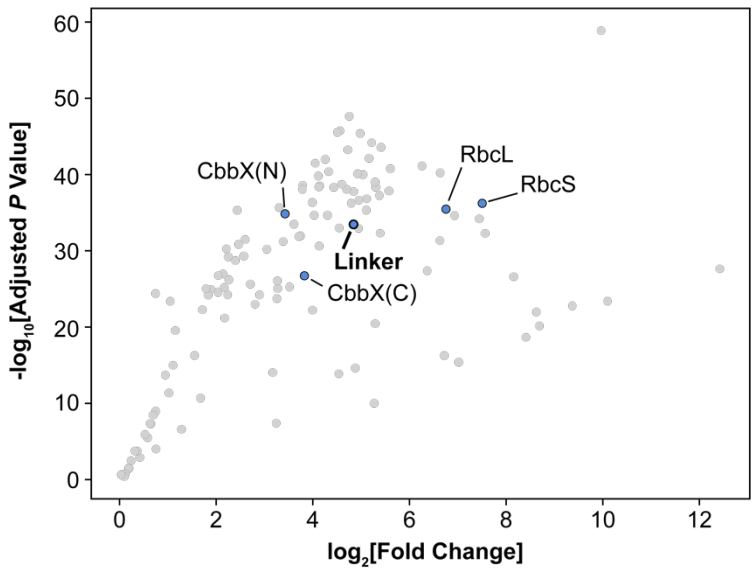


Figure 2-9 | *Ex vivo* validation of *Porphyridium* linker interaction with Rubisco.
Scatter plot of co-immunoprecipitation mass spectrometry results from experiments completed with an antibody raised to the Rubisco large subunit (RbcL) relative to a blank control experiment. Only values with a positive enrichment relative to the control are shown. Proteins of interest are shown in filled blue circles and annotated. The nuclear and chloroplast forms of the Rubisco activase CbbX are indicated (N and C respectively).

2.3.4. Characterisation of candidate Ulva linker proteins

Owing to the high quality genome assembly (De Clerck *et al.,* 2018), and rapid development of molecular tools (Blomme *et al.,* 2021), the globally abundant green seaweed *Ulva mutabilis* (Ulva hereafter) is rapidly emerging as a model organism (Blomme *et al.,* 2023). In contrast to Chlamydomonas, Ulva is a multicellular alga with pyrenoids that remain almost entirely uncharacterised. More generally, pyrenoids in multicellular organisms also remain uncharacterised at a molecular level. As such, there is great interest in the characterisation of the Ulva pyrenoid. Ulva and Chlamydomonas both belong to the Ulvophyceae, Trebouxiophyceae, Chlorophyceae (UTC) clade within the chlorophytes, and though they diverged ~800 million years ago (Mya) (Cortona *et al.,* 2020), the ultrastructure of their pyrenoids is largely similar. 12 candidate linker sequences were identified by FLIPPer v1 analysis of the Ulva genome (Table 2-5). Of the 12 outputted sequences, one sequence (UM120_0037) was clearly most analogous to EPYC1 as it demonstrated 6 quasi-identical repeats with electrostatic and aromatic-containing motifs that mapped to predicted structural features within the largely disordered protein (Fig. 2-10a,b). Additionally, the repeating helices of UM120_0037 are spaced at a distance similar to those of EPYC1 (~50 residues) (Fig. 2-10c), consistent with a hypothesised role in cross-linking Rubiscos. BLAST analysis of UM120_0037 identified sequence homologues of UM120_0037 in the draft sequences of the genomes of *Ulva linza and Ulva prolifera* (Fig 2-10c), with no homologous sequences identified outside of *Ulva* species, consistent with the limited conservation of EPYC1 within the Chlamydomonadales and PYCO1 in Phaeodactylum (Table 2-1). The putative *Ulva* linker sequences from the 3 species demonstrated highly similar arrangement of the predicted helical regions (Fig. 2-10c), consistent with their hypothesised conserved role in Rubisco-cross linking.

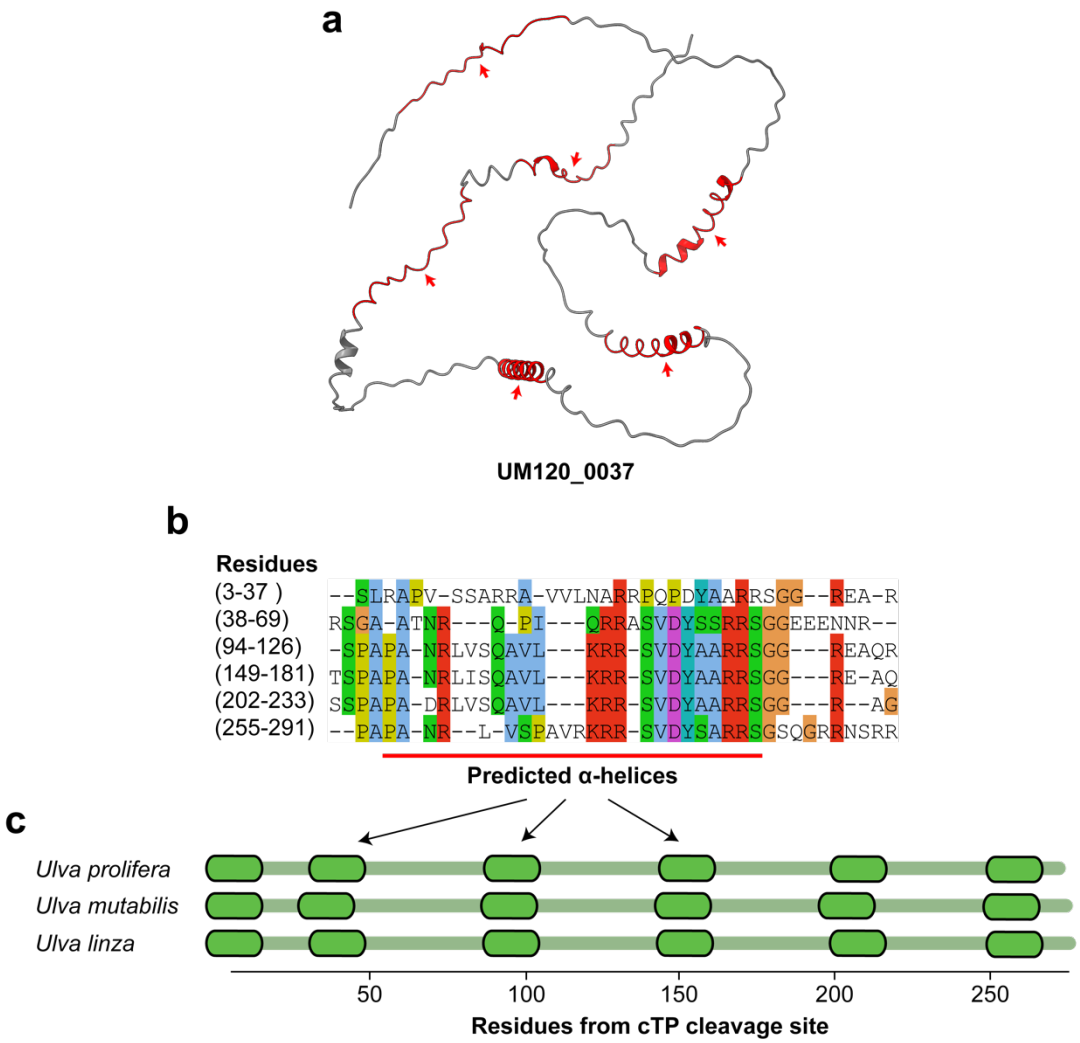


Figure 2-10 | *In silico* characterisation of the hypothesised *Ulva* linker protein.
(a) Top ranked AlphaFold2 model of UM120_0037. The predicted α-helical regions are shown in red and indicated by red arrows. (b) Alignment of the repeat sequences identified by XSTREAM analysis of the UM120_0037 sequence. The predicted α-helices are indicated by the red bar. Residues are coloured by property according to the Clustal X colour scheme. (c) Scaled schematic representation of putative *Ulva* linker sequences identified by BLAST in the three *Ulva* genomes. The sequences are shown following cleaving of the TargetP predicted chloroplast transit peptide. Green oblongs represent the predicted helical regions, connected by disordered sequence.

To validate the role of the putative *Ulva* linker sequence *in vivo*, YFP-tagged UM120_0037 was localised in *Ulva mutabilis* by Jonas Blomme (University of Ghent) who previously established a robust tagging methodology (Blomme *et al.,* 2021). Consistent with the putative role of UM120_0037 in Rubisco condensation, the YFP signal was localised almost exclusively to the pyrenoid (Fig. 2-11a). The same localisation was observed for YFP-tagged Rubisco small subunit (RbcS; UM023_0100 – Fig. 2-11b), consistent with both the putative linker and Rubisco localising in the pyrenoid. Preliminary qualitative Fluorescence Recovery After Photobleaching (FRAP) experiments also completed by Jonas

Blomme demonstrated the mobility of the protein within the pyrenoid (Fig. 2-11c). Whilst this evidence is not sufficient to distinguish the candidate sequence as a *bona fide* linker protein, the similarity of the protein structure and extent of conservation, as well as the localisation and mobility in the pyrenoid are consistent with the putative role. Further characterisation by knockout of the protein *in vivo* as well as *in vitro* characterisation of the putative interaction with Rubisco should be pursued to empirically validate UM120_0037 as a *bona fide* linker protein.

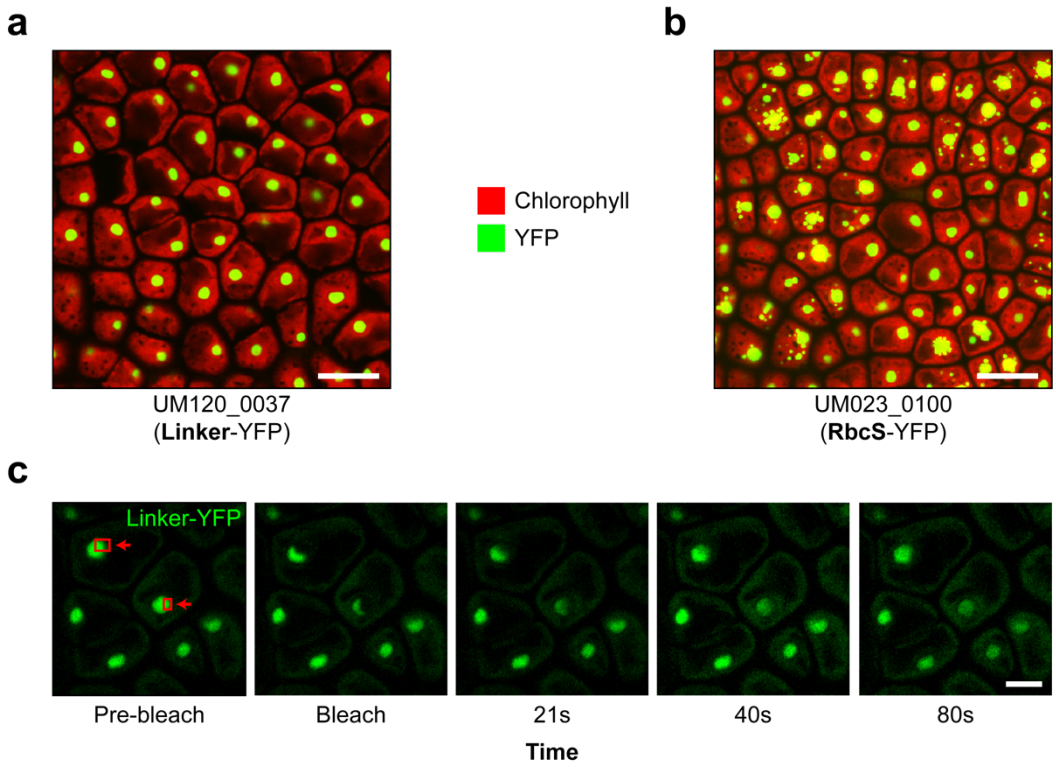


Figure 2-11 | *In vivo* validation of the Ulva linker sequence.
(a) Confocal fluorescence microscopy image of YFP-tagged Ulva linker (UM120_0037) expressed in *Ulva mutabilis*. Red is chlorophyll autofluorescence, and green is YFP fluorescence. Scale bar = 5 µm. (b) Confocal fluorescence microscopy image of YFP-tagged Ulva Rubisco small subunit (RbcS) (UM023_0100) expressed in *Ulva mutabilis*. Red is chlorophyll autofluorescence, and green is YFP fluorescence. (c) Snapshots from a representative FRAP experiment of YFP-tagged linker in the pyrenoid. The red boxes, indicated by arrows represent the bleached region. Scale bar = 2 µm.

2.3.5. *Characterisation of candidate Chlorella linker proteins*

Chlorella are unicellular green algae that belong to the Trebouxiophyceae class within the UTC clade and represent a well-studied genus (Krienitz *et al.,* 2015), with over 15 sequenced genomes. Despite huge interest in the bio-production of *Chlorella* for a range of applications (Safi *et al.,* 2014), very little molecular characterisation of the pyrenoid has been completed (Villarejo *et al.,* 1998). Given the structural analogy of EPYC1 and the putative Ulva linker (Fig. 2-10), it was expected a similar protein would be found in the genomes of *Chlorella* species. In the large scale implementation of FLIPPer v1, analysis was completed against 7 annotated *Chlorella* genomes and the genome of a closely related Chlorellaceae species, *Micractinium conductrix* (Table 2-5). Of the 8 genomes analysed, the lowest number of candidate sequences was obtained from that of *Chlorella vulgaris* CCAP211/11P (Cecchin *et al.,* 2019), in which 7 candidates were identified. AlphaFold2 modelling of the sequences, and comparison with the XSTREAM identified repeats demonstrated one sequence (KAI3425152) with clear analogous features to those of EPYC1 and the putative Ulva linker (Fig. 2-12a). As with EPYC1, PYCO1 and the putative *Porphyridium* and Ulva linkers, the predicted helical motifs of the KAI3425152 sequence also contained an aromatic residue and multiple electrostatic residues (Fig. 2-12b). Homologues of this protein were also identified by FLIPPer in the other 6 *Chlorella* genomes and the *Micractinium* genome, albeit with lower selectivity (Table 2-5, Fig. 2-12c). As observed with EPYC1 homologues in the Chlamydomonadales, homologues of the putative *Chlorella* linker demonstrated variation in repeat number and length. The range of repeat length across all sequences was 30-136 residues, with a range of 3-8 repeated helical regions also present, which exemplified the need for the previously completed parameterisation of FLIPPer v1. Default NCBI BLASTp analysis of KAI3425152 did not identify any homologous proteins outside of *Chlorella* or *Micractinium* species (Table 2-7), again reinforcing the specificity of distribution of the proteins.

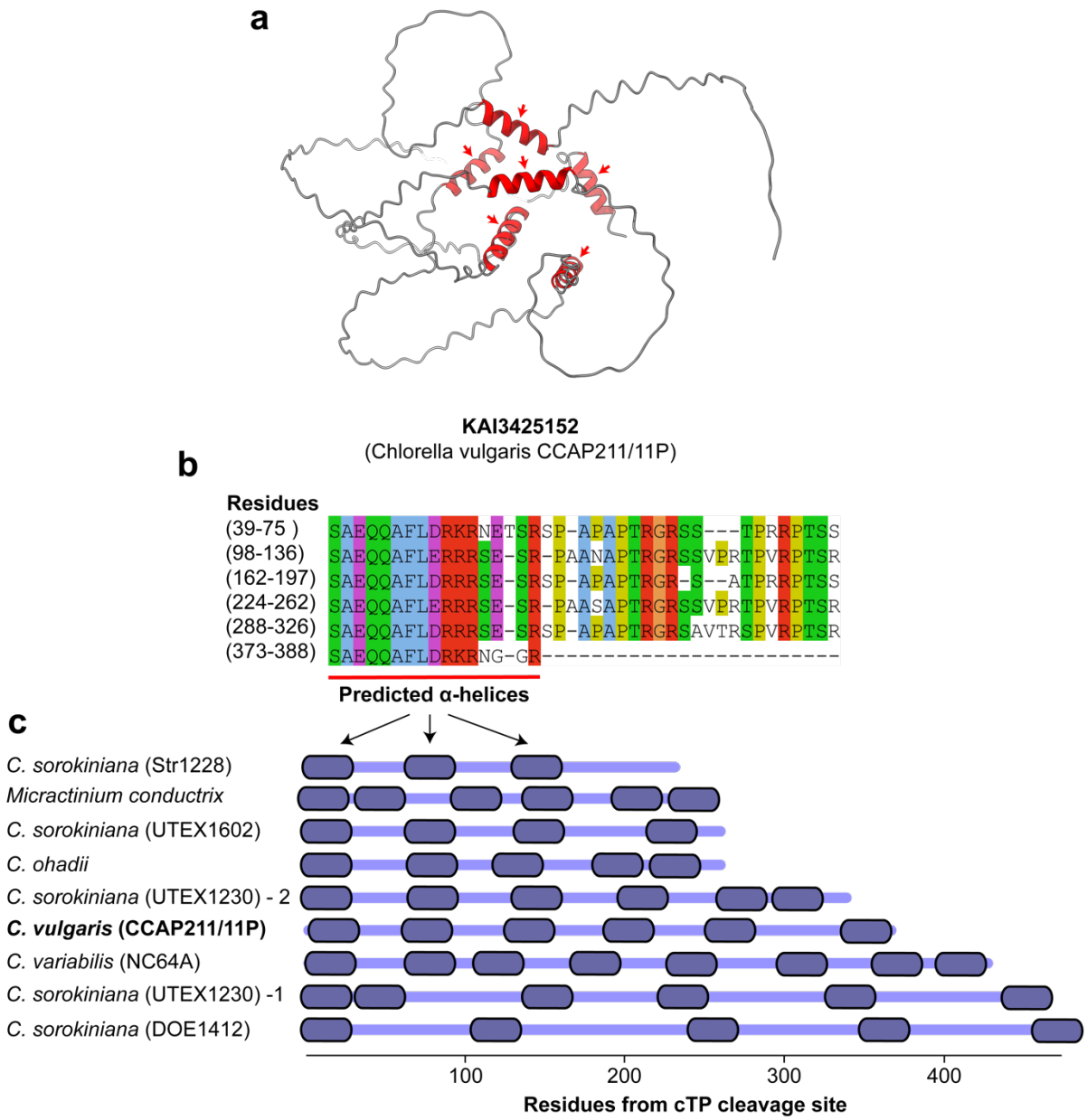


Figure 2-12 | *In silico* characterisation of the putative *Chlorella* linker protein.
Top ranked AlphaFold2 model of KAI3425152 identified by FLIPPer from the genome of *Chlorella vulgaris* CCAP211/11P. The. Predicted α-helical regions are shown in red and indicated by red arrows. (b) Alignment of the repeat sequences identified by XSTREAM analysis of the KAI3425152 sequence. The predicted α-helices are indicated by the red bar. Residues are coloured by property according to the Clustal X colour scheme. (c) Scaled schematic representation of putative *Chlorella* linker sequences identified by FLIPPer in the *Chlorella* genomes. The sequences are shown following cleaving of the TargetP 2.0 predicted chloroplast transit peptide. Purple oblongs represent the predicted helical regions, connected by disordered sequence.

Table 2-7: BLAST results for the putative *Chlorella* linker.

Query	Hit Species	E value	%Identity	Accession
KAI3425152.1	*Chlorella vulgaris*	0	100.00%	KAI3425152.1
KAI3425152.1	*Chlorella vulgaris*	6E-102	72.89%	KAI3429225.1
KAI3425152.1	*Chlorella vulgaris*	2E-91	70.47%	KAI3429224.1
KAI3425152.1	*Chlorella variabilis*	8E-53	46.67%	XP_005851975.1
KAI3425152.1	*Micractinium conductrix*	1E-44	43.15%	PSC68468.1
KAI3425152.1	*Micractinium conductrix*	7E-41	43.80%	PSC68430.1
KAI3425152.1	*Chlorella sorokiniana*	6E-35	48.99%	PRW44308.1
KAI3425152.1	*Micractinium conductrix*	7E-31	46.74%	PSC72476.1
KAI3425152.1	*Micractinium conductrix*	9E-31	46.74%	PSC72477.1
KAI3425152.1	*Chlorella vulgaris*	4E-23	52.09%	KAI3430698.1
KAI3425152.1	*Micractinium conductrix*	5E-22	45.40%	PSC69993.1
KAI3425152.1	*Micractinium conductrix*	1E-21	45.09%	PSC69992.1
KAI3425152.1	*Chlorella sorokiniana*	4E-21	50.41%	PRW50772.1
KAI3425152.1	*Chlorella variabilis*	4E-19	44.27%	XP_005850249.1
KAI3425152.1	*Chlorella variabilis*	9E-16	37.81%	XP_005844057.1
KAI3425152.1	*Chlorella vulgaris*	7E-15	43.80%	KAI3427970.1
KAI3425152.1	*Chlorella ohadii*	9E-13	38.43%	KAI7840023.1

A complete characterisation of the *Chlorella* linker is presented in Chapters 3, 4 and 5 but some validation is presented in this section for clarity. As previously demonstrated with EPYC1 and Chlamydomonas Rubisco (Wunder *et al.,* 2018), when purified *Chlorella* linker was mixed with purified *Chlorella* Rubisco the two proteins spontaneously demixed into micron-scale droplets akin to *in vivo* pyrenoids. The droplets also contained both Rubisco and the linker protein (Fig. 2-13). *Ex vivo* co-immunoprecipitation mass spectrometry, *in vivo* absolute quantification mass spectrometry and extensive biochemical characterisation all presented in subsequent chapters also supported the hypothesised role of the identified *Chlorella* protein as a *bona fide* pyrenoid linker.

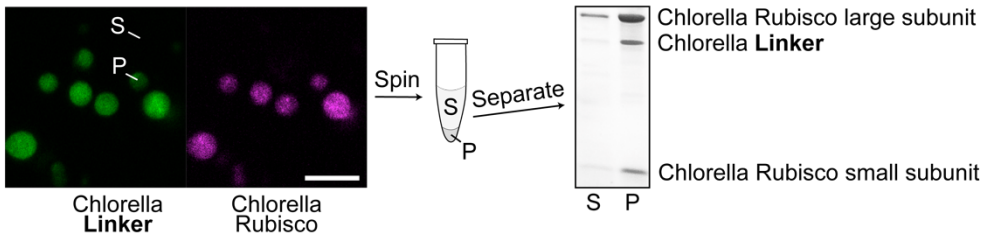


Figure 2-13 | *In vitro* validation of the *Chlorella* linker candidate.
Confocal fluorescence microscopy image (left) of *Chlorella* linker (green) and *Chlorella* Rubisco (magenta) co-localised in droplets (P) demixed from solution (S). Scale bar = 2 μm. Droplets were centrifuged (middle) to separate the supernatant (S) and pelleted droplet (P) phase which were separated by SDS-PAGE (right) and demonstrated the presence of the *Chlorella* linker candidate and Rubisco in the droplet phase (see Chapter 3 methods).

2.3.6. General observations from FLIPPer v1 usage

In addition to the success of FLIPPer v1 in the identification of at least 3 apparent novel linker proteins, several reflections from the wide scale implementation of the pipeline can also be made. Although relative success was achieved, with FLIPPer robustly identifying EPYC1 structural analogues in the genomes of diverse species, there was also a large number of genomes in which no convincing proof of analogous pyrenoid linker proteins could be obtained. Possible explanations for this are discussed with examples here.

Given the success of FLIPPer v1 in identification of PYCO1 and possible additional linker-like proteins in Phaeodactylum (Fig. 2-7), it was expected that identification of analogous linker proteins in the high quality genome of the related model centric diatom *Thalassiosira pseudonana* would be equally trivial. By contrast, no convincing linker candidate proteins were identified with FLIPPer v1 run in default mode, nor when an extensive parameter search space was explored. Comparison to the results of a Rubisco co-IP experiment completed by Dr. Onyou Nam concurred with this observation, that no clear linker candidate could be identified. It should be noted however that the biochemical identification of PYCO1 was also reportedly challenging, given its scarcity in Rubisco co-IP experiments (Oh *et al.,* 2023), and so may also preclude identification of a hypothesised *Thalassiosira* linker protein by this method. This observation leaves an open question as to whether an analogous mechanism is utilised by the diatom *Thalassiosira* to assemble the pyrenoid, or whether an alternative mechanism to condense

Rubisco has evolved. The lack of conservation of PYCO1 in *Thalassiosira* despite their relatively recent divergence in the last 100 My (Sims *et al.,* 2006), would certainly suggest a relatively recent, independently evolved mechanism. Critically, the direct role of PYCO1 in functional condensation of Rubisco *in vivo* has not been demonstrated and thus it is also possible that although PYCO1 exhibits phase separation of Rubisco, but that this is not essential to condensation and function of Rubisco in the pyrenoid. Further work to understand the biogenesis, division, and dynamics of diatom pyrenoids will be essential to understand the role and evolution of linker proteins and Rubisco phase separation in their function.

Notably, no strong linker candidates were identified in the genomes of species belonging to prasinophytes; a polyphyletic group of flagellate, unicellular, green algal species that anciently diverged from the chlorophytes (Fawley *et al.,* 2001; Guillou *et al.,* 2004). The average genome size of prasinophytes is approximately one third that of chlorophyte species (Li *et al.,* 2020; Worden *et al.,* 2009), and accordingly the number of annotated proteins is also roughly a third of the number in Chlamydomonas (Craig *et al.,* 2023). Given that 4 *Micromonas* genomes and a *Prasinoderma* genome were analysed by FLIPPer v1, it was surprising that no candidates were obtained. This again suggested an alternative mechanism for Rubisco condensation, that was not detected by FLIPPer has also evolved in this species – perhaps explained by the minimal genome size of these species. Despite their abundance in the Chlorophyta, the pyrenoids of prasinophyte algae are almost undiscussed and certainly warrant further investigation. Notably, prasinophyte pyrenoids have a wholly simple ultrastructure (Fig. 1-9) (van Baren *et al.,* 2016) and possess possible invagination of mitochondria-containing outer membranes (Jouenne *et al.,* 2011). Other notable exemptions from linker detection included the high quality genomes of the halophile *Dunaliella salina* and the acidophile *Chlamydomonas eustigma*, which are both closely related *Chlamydomonas* in the Volvacales (Craig *et al.,* 2021), and diverged less than 500 Mya (Davidi *et al.,* 2023). The extremophilic nature of these species may suggest an alternative mechanism required for the condensation of Rubisco that was again not detected by FLIPPer. Elsewhere, the FLIPPer outputs required much further analysis beyond the scope of this project to decipher the identity of linker proteins with a confidence high enough to progress.

Taken together, the results of FLIPPer v1 analysis demonstrated the pipeline performed its intended role in the identification of proteins with the key features of EPYC1. Whilst this was productive in the identification of analogues in *Chlorella*, *Ulva* and *Porphyridium* species, it may also have limited the identification of proteins that only possessed some of the features of EPYC1, or those that represent an entirely different mechanism for Rubisco condensation. Although a large degree of useful flexibility is achievable within the pipeline through customisation using the terminal user inputs (Fig. 2-5) as exemplified in the identification of the *Porphyridium* linker, some limitations were imposed by the programs used and construction of the pipeline. Given that XSTREAM was designed to identify head-to-tail (tandem) repeats, it was not suited for the identification of repeat sequences that were not continuously repetitive. This became apparent upon the analysis of the *Pleodorina starii* and Phaeodactylum genomes. In the *Pleodorina* genome, the EPYC1 homologue contains variable spacer regions (Fig. 2-7), which precluded identification of the final repeat region by XSTREAM. Similarly, the lack of 'tandem' repeat nature in the Phatr3_J46711 sequence precluded its identification in FLIPPer v1 analysis. This scenario is eminently possible in other hypothesised linker proteins and may have contributed to some of the missing candidates where they were expected. The other considered repeat detection software, RADAR (Fig. 2-2), is not reliant on the identification of tandem repeats and thus could be further explored as an alternative to XSTREAM to remedy this drawback. That said, most of the accurate selectivity of the pipeline relies on the correct parameterisation of XSTREAM and this would need to be carefully considered and tested when exploring the implementation of alternative repeat detection software. More generally, exploring a more integrated approach to analysis of the outputted results would vastly improve the objectivity of the final analysis. In FLIPPer v1, a degree of subjectivity is introduced in the analysis of the outputs to decide which candidates to explore further *in silico* where the number of candidates precludes total exploration. Integrating the common downstream analyses (BLAST, disorder repeat structure, structural mapping of repeats and structural modelling) into the pipeline prior to user analysis should be prioritised in future developments of the pipeline. Future iterations of FLIPPer (v2) should also aim to reassess the search criteria based on the outcome of ongoing characterisations and novel pyrenoid biology.

2.4. Assessment of RBM-guided pyrenoid assembly in Ulva and Chlorella

In Chlamydomonas, formation of the pyrenoid matrix is underpinned by multivalent interaction between the repeating helices of EPYC1 termed Rubisco-binding motifs (RBMs), and the small subunit of Rubisco (He *et al.,* 2020; Mackinder *et al.,* 2016). In addition to EPYC1, several other Chlamydomonas proteins that contained additional structural features were also found to possess similar RBMs that bound Rubisco (Meyer, Itakura, *et al.,* 2020). It was hypothesised that a subset of these proteins guide assembly of the pyrenoid matrix with the surrounding starch sheath and traversing thylakoid material through RBM-guided interaction with Rubisco in the matrix and separate interactions with the respective ultrastructural features using the additional structural domains. Meyer *et al.,* (2020) highlighted two carbohydrate binding module family 20 (CBM20)-containing proteins, StArch Granules Abnormal 1 and 2 (SAGA1 and SAGA2) that possess multiple copies of the RBM. A mutant of the SAGA1 protein was previously shown to have irregularly shaped starch plates surrounding the pyrenoids, which were also abnormally numerous in the chloroplast, consistent with the hypothesised role of SAGA1 in Rubisco-guided assembly of the starch sheath with the pyrenoid (Itakura *et al.,* 2019). Given the high CO_2-requiring phenotype of the SAGA1 mutant, it was proposed that SAGA1 and SAGA2 play essential, non-redundant roles in assembly of the starch sheath with the pyrenoid matrix. Two RBM-containing proteins were also proposed to underpin interaction between the pyrenoid matrix and the traversing thylakoids. The first, Rubisco-binding membrane protein 1 (RBMP1) otherwise known as bestrophin-like protein 4 (BST4), contains a bestrophin-like domain in addition to an extended C-terminus containing multiple RBMs. Although exclusive localisation of the protein to the traversing thylakoid material has been shown to be dependent on the RBMs, no growth phenotype or effect on thylakoid traversion was demonstrated in a mutant of the protein (Adler *et al.,* 2023). The second Rubisco-binding membrane protein (RBMP2) contains a split Rhodanese domain and weakly predicted transmembrane character in addition to the multiple RBMs. No evidence exists to date to support the hypothesised function of this protein. A host of other RBM-containing proteins were also highlighted, consistent with the role of RBMs in targeting proteins to the pyrenoid also (*e.g.,* the chloroplast epimerase CSP41a) (Meyer, Itakura, *et al.,* 2020). Given the apparently analogous evolved features of pyrenoid linker proteins in Chlamydomonas (Fig. 2-6), Ulva (Fig. 2-10) and

Chlorella (Fig. 2-12), it was hypothesised that Ulva and Chlorella may also have evolved similar RBM-guided pyrenoid assembly mechanisms. To this end, it was aimed to determine whether analogous proteins exist in these two species and to assess the likelihood of RBM-guided assembly having evolved as a conserved trait.

2.4.1. Identification of RBM-containing proteins in Ulva

Given the structural analogy of EPYC1 (Fig. 2-4, Fig. 2-6) and the putative Ulva linker (Fig. 2-10), it was reasonable to assume the interaction between the linker protein and Rubisco would also be underpinned by the predicted helical region as observed for EPYC1 (He *et al.,* 2020). To determine the search parameters for the putative RBM in Ulva, the predicted helices of the linkers from the 3 *Ulva* species were initially compared, and demonstrated complete conservation of several residues within the predicted helices (Fig. 2-14a). Consistent with the motif composition in EPYC1 (Table 2-3), 4 electrostatic residues and a central aromatic tyrosine residue were completely conserved across the 18 aligned helices. A python-implemented regular expression (regex) search of the Ulva genome using the conserved residues of the predicted helices ([**R/K**][**R**][X][X][X][**D/S**][**Y**][X][X][R][**R**]) highlighted 5 additional motif-containing proteins (Table 2-8). Interestingly an annotated Rubisco activase-like protein (UM003_0241) was identified amongst these and contained a single conserved motif in an exposed N-terminal region, suggesting a role in interaction with Rubisco (Fig. 2-14b). Alongside the putative linker protein (UM120_0037), 3 of the other proteins contained multiple putative RBMs, of which 2 contained no additional sequence annotations (UM004_0280 and UM0047_0002). The remaining protein with multiple RBMs (UM0016_026) contained a high confidence predicted transmembrane domain at the C-terminus (Fig. 2-14e), suggesting a possible role as a tether protein between the traversing thylakoid material and the pyrenoid matrix, analogous to the proposed role of RBMP2 in Chlamydomonas (Meyer, Itakura, *et al.,* 2020). Jonas Blomme kindly tagged this protein in Ulva and demonstrated a sub-pyrenoid localisation consistent with the hypothesised interaction with the traversing thylakoid material (Fig. 2-14g). Closer inspection of the UM0016_026 sequence revealed multiple variations of the putative RBM (Fig. 2-14d), that were also present in the UM004_0280 and UM047_002 sequences (Fig. 2-14c), suggesting some variation is permitted within the RBM sequence.

Localisation of UM004_0280 in Ulva demonstrated uniformity throughout the pyrenoid, in line with the lack of annotated domains (Table 2-7, Fig. 2-14f). This result suggests there may be redundant function of multiple motif-containing linker-like proteins in Ulva as also proposed in Phaeodactylum. Further work to characterise the relative expression and abundance *in vivo* as well as the *in vitro* Rubisco condensation properties of the linker-like proteins should be completed to decipher their exact role in pyrenoid assembly.

Table 2-8: Regular expression search results for RBM-containing proteins in *Ulva*.

ID	Number of motifs	Annotation
UM003_0241.1	1	Rubisco activase
UM004_0280.1	5	None
UM016_0026.1	6	Transmembrane (thylakoid tether?)
UM047_0002.1	7	None
UM119_0040.1	1	None
UM120_0037.1	**6**	**Putative linker**

Given the apparent conservation of an analogous thylakoid-matrix tether protein in Ulva, it was expected that a starch-matrix tether akin to SAGA1/2 in Chlamydomonas would also be present. By contrast, no clear starch-interacting domains were identified in the RBM-containing sequences identified by regex search with the canonical RBM, or more lenient searches with non-canonical RBM sequences identified in other pyrenoid-localised proteins (Fig. 2-14c). Similarly, BLAST analysis of the motif in the Ulva genome returned the same proteins identified by regex search of the canonical RBM (Table 2-8). BLAST of the CBM20 domain of SAGA1 from Chlamydomonas in the Ulva genome returned several CBM20-containing proteins, though none possessed RBM or RBM-like sequences, suggesting a direct SAGA analogue is not present in Ulva. Taken together, the results of this section suggest that RBM-containing proteins likely play a role in assembly of the Ulva pyrenoid, but that total conservation of an analogous RBM-driven assembly mechanism proposed in Chlamydomonas is unlikely. Further work to characterise the proteome of the Ulva pyrenoid and decipher whether Ulva has evolved an alternate mechanism to arrange the starch sheath around the matrix would be of particular interest moving forward. Similarly, empirical evidence for the functions of the identified RBM-containing proteins should also be prioritised.

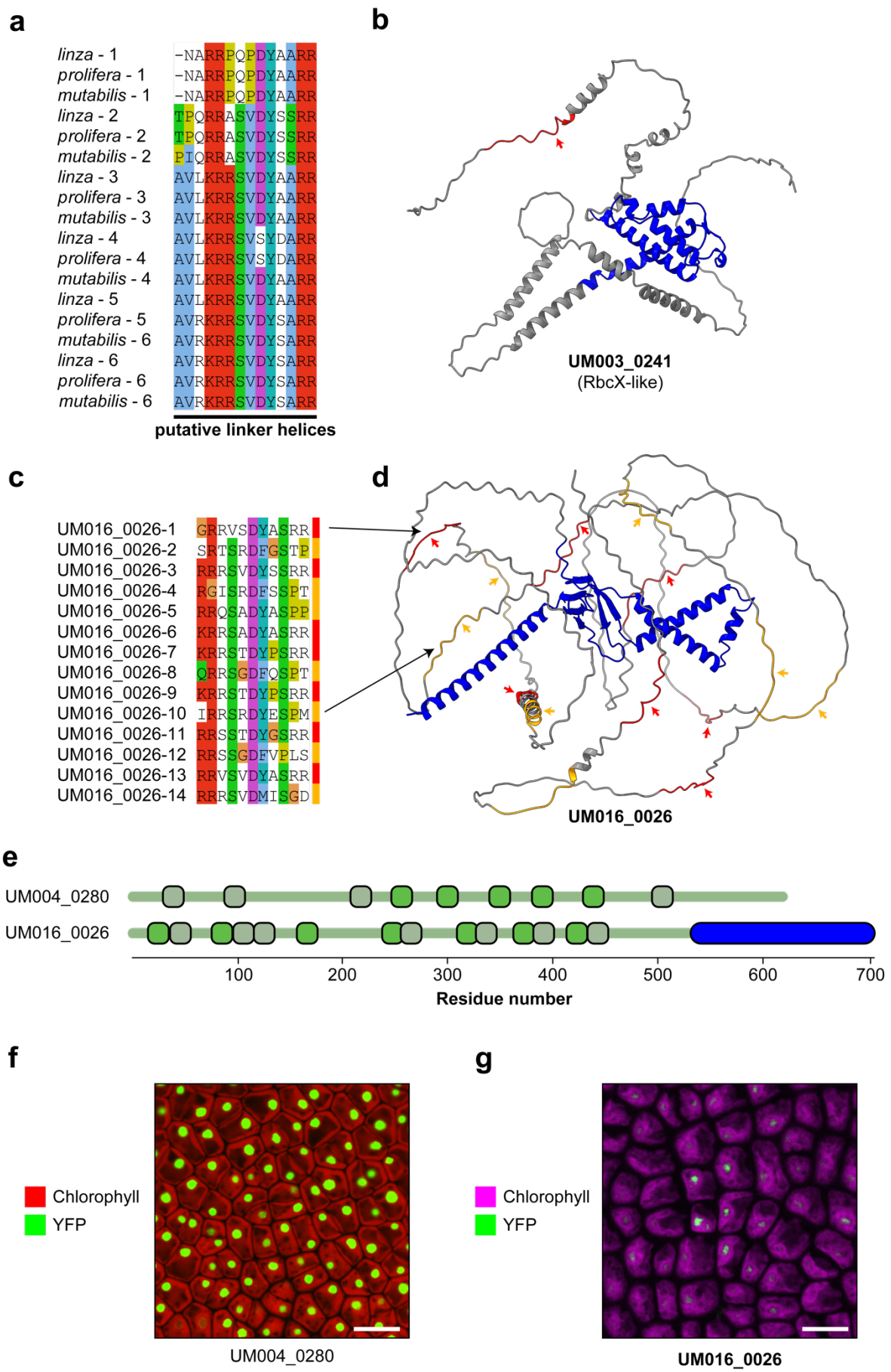

a
linza - 1 -NARRPQPDYAARR
prolifera - 1 -NARRPQPDYAARR
mutabilis - 1 -NARRPQPDYAARR
linza - 2 TPQRRASVDYSSRR
prolifera - 2 TPQRRASVDYSSRR
mutabilis - 2 PIQRRASVDYSSRR
linza - 3 AVLKRRSVDYAARR
prolifera - 3 AVLKRRSVDYAARR
mutabilis - 3 AVLKRRSVDYAARR
linza - 4 AVLKRRSVSYDARR
prolifera - 4 AVLKRRSVSYDARR
mutabilis - 4 AVLKRRSVDYAARR
linza - 5 AVLKRRSVDYAARR
prolifera - 5 AVRKRRSVDYSARR
mutabilis - 6 AVRKRRSVDYSARR
linza - 6 AVRKRRSVDYSARR
prolifera - 6 AVRKRRSVDYSARR
mutabilis - 6 AVRKRRSVDYSARR
putative linker helices
b
UM003_0241
(RbcX-like)
c
UM016_0026-1 GRRVSDYASRR
UM016_0026-2 SRTSRDFGSTP
UM016_0026-3 RRRSVDYSSRR
UM016_0026-4 RGISRDFSSPT
UM016_0026-5 RRQSADYASPP
UM016_0026-6 KRRSADYASRR
UM016_0026-7 KRRSTDYPSRR
UM016_0026-8 QRRSGDFQSPT
UM016_0026-9 KRRSTDYPSRR
UM016_0026-10 IRRSRDYESPM
UM016_0026-11 RRSSTDYGSRR
UM016_0026-12 RRSSGDFVPLS
UM016_0026-13 RRVSVDYASRR
UM016_0026-14 RRRSVDMISGD
d
UM016_0026
e
UM004_0280
UM016_0026
100
200
300
400
500
600
700
Residue number
f
Chlorophyll
YFP
UM004_0280
g
Chlorophyll
YFP
UM016_0026

Figure 2-14 | Characterisation of RBM-containing proteins in *Ulva*.
(a) Alignment of the predicted helical regions from the putative linker proteins of the 3 *Ulva* species (*linza*, *prolifera* and *mutabilis*). Residues are coloured by property according to the Clustal X colour scheme. (b) Top ranked AlphaFold2 model of the RbcX-like protein UM003_0241. The predicted RbcX-like fold is shown in blue. The putative RBM is shown in red and indicated by a red arrow. (c) Alignment of the putative RBMs from UM016_0026. The coloured bar adjacent to the sequence demonstrates canonical (red) or non-canonical (yellow) RBMs. Residues are coloured by property according to the Clustal X colour scheme. (d) Top ranked AlphaFold2 model of UM016_0026. The canonical RBMs are shown in red indicated by red arrows, with the non-canonical RBMs in orange indicated by orange arrows. The predicted transmembrane domain is shown in blue. (e) Scaled schematic representation of the multiple motif-containing proteins. Green oblongs represent the canonical RBMs as present in the UM120_0037 linker (Fig. 2-14a) and grey oblongs represent non-canonical RBMs (yellow in Fig. 2-14d). The predicted transmembrane domain in UM016_0026 is shown in blue. (f) Confocal fluorescence microscopy image of YFP-tagged UM004_0280 in Ulva completed by Jonas Blomme. (g) Confocal fluorescence microscopy image of YFP-tagged UM016_0026 in Ulva completed by Jonas Blomme.

2.4.2. Identification of RBM-containing proteins in Chlorella

Given the presence of RBM-containing proteins in Ulva and the related phylogeny of *Chlamydomonas*, *Ulva* and *Chlorella* species in the UTC clade, it was hypothesised that similar RBM-containing proteins could also be present in the genomes of *Chlorella* species. Given the high number of genomes available for *Chlorella* species (Table 2-5), it was also more likely a complete picture of the evolution and conservation of the hypothesised RBM-containing proteins could be obtained. As such, a similar regex search was completed, using the conserved features of the 49 putative RBMs aligned from the 9 identified linker sequences in *Chlorella* species (Fig. 2-12c, 2-15a). Searching the regular expression of [**D/E**][**Q**][X][X][**F**][**L/M/V**][X][**R**][**K/R**] against all 8 *Chlorella* genomes returned an average of ~9 motif-containing sequences in each genome (Table 2-9).

Table 2-9: Regular expression search results for RBM-containing proteins in *Chlorella species*.

Species	Number of sequences	Average number of motifs per sequence
Chlorella sorokiniana (Str1228)	8	2.1
Micractinium conductrix	11	2.8
Chlorella sorokiniana (UTEX1602)	13	1.5
Chlorella ohadii	6	1.6
Chlorella sorokiniana (UTEX1230)	8	3.1
Chlorella vulgaris (CCAP211/11P)	13	3.6
Chlorella variabilis (NC64A)	7	4.3
Chlorella sorokiniana (DOE1412)	7	2.7

Each of the *Chlorella sorokiniana, Chlorella ohadii* and *Micractinium conductrix* genomes contained a single clear CBM20 domain containing sequence with multiple RBMs that presented clear analogy to the SAGA proteins of Chlamydomonas (Fig. 2-15b). *Chlorella variabilis* and *Chlorella vulgaris* possessed multiple CBM20 domain containing sequences amongst the identified RBM-containing sequences, which all demonstrated homology to those of *sorokiniana, ohadii* and *Micractinium*. The clear structural similarity of the *Chlorella* CBM20-containing proteins with those of Chlamydomonas suggests the proteins have separately evolved the same function in tethering the starch sheath to the pyrenoid matrix. As such, the SAGA analogues in *Chlorella* are hereafter termed Pyrenoid Matrix Starch Tether (PMST) proteins. The structural similarity of the PMST proteins between species was in agreement with the phylogeny of *Chlorella* (Fig. 2-15b) (Wang *et al.,* 2020), with the early diverged *sorokiniana* and *ohadii* sequences demonstrating conserved, but clearly different arrangement than the later diverged *Micractinium* sequence. The *vulgaris* and *variabilis* PMST sequences showed further differences, in line with their even more recent divergence. This suggests that PMST proteins were present in a common ancestor of extant *Chlorella* species, and that the sequence evolved divergently over the last ~165 My, after *Chlorella sorokiniana* and *ohadii* first diverged from the remaining extant species.

Although the PMST proteins are all structurally analogous to the SAGA proteins of Chlamydomonas, they exist more as pseudo-linker sequences fused to the CBM20 domain. That is, the sequence between the RBMs of the PSMTs is more akin to the spacer sequence of the putative linker proteins, in contrast to SAGA1/2 in Chlamydomonas, where the spacer sequence possesses predicted coiled-coil structure. Further work to determine whether the spacer regions in the PMSTs and the putative linker sequences of each species are of the same origin may shed light on the evolution of RBM-containing proteins, possibly providing evidence for the hypothesis that they evolved from the duplication of an early linker sequence. The *in vivo* abundance and *ex vivo* Rubisco interaction of PMST from *Chlorella sorokiniana* UTEX1230 was subsequently confirmed in work completed by Mihris Naduthodi (University of York), and presented briefly in Chapter 3, though further empirical proof of the role of these proteins *in vivo* and *in vitro* should also be pursued.

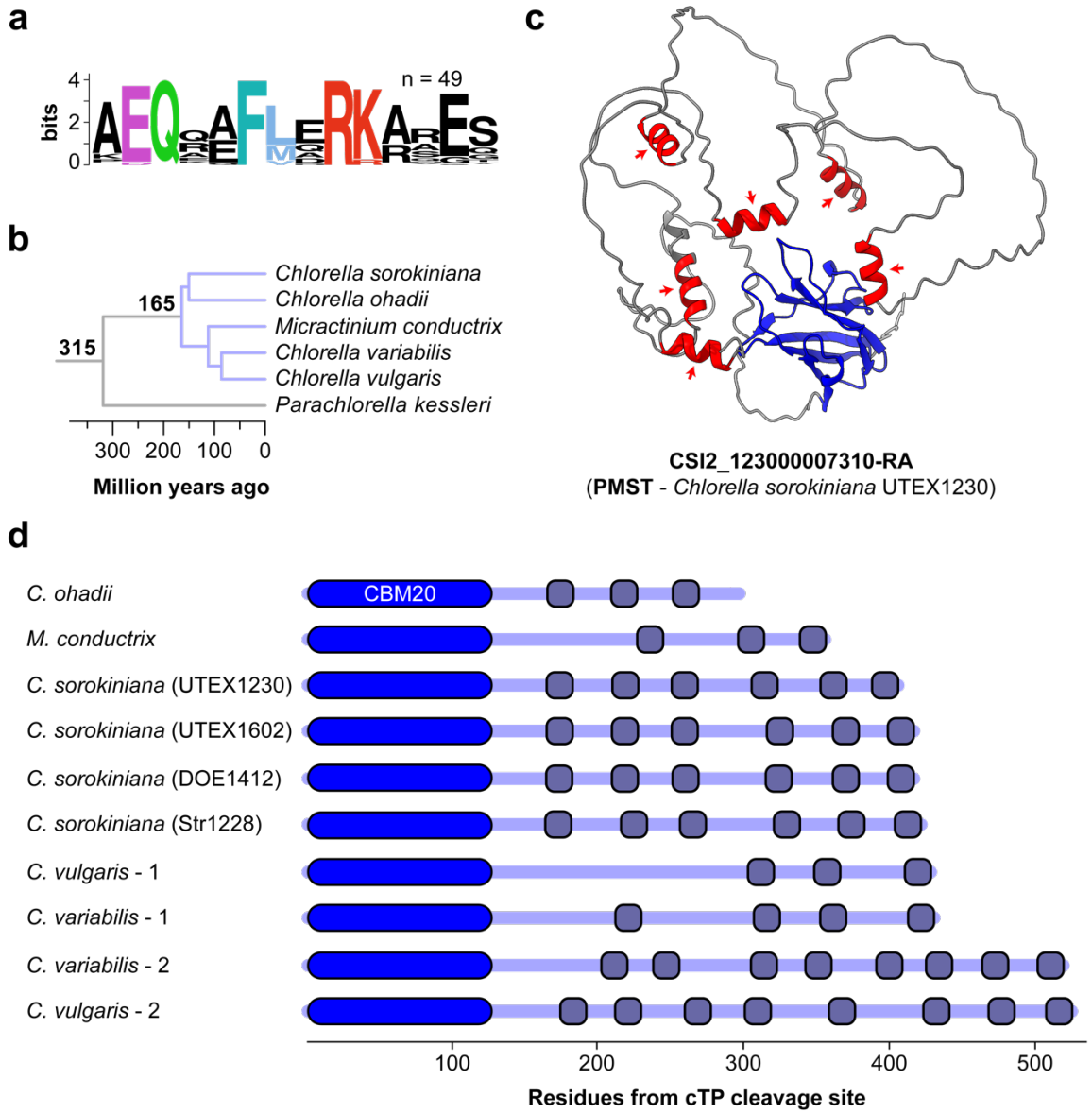


Figure 2-15 | Characterisation of CBM20 domain containing RBM proteins in *Chlorella*.
(a) Consensus logo based on the alignment of the 49 predicted helices in the 9 identified *Chlorella* linker sequences. The conserved residues are coloured according to the Clustal X colour scheme. The size of the one letter amino acid code in each position indicates the level of conservation at each position. (b) Time-calibrated phylogenetic chronogram of *Chlorella* species taken from (Wang *et al.,* 2020). (c) Top ranked AlphaFold2 model of the Pyrenoid Matrix Starch Tether (PMST) from *Chlorella sorokiniana* UTEX1230 (CS1230000007310-RA). The putative RBMs are shown in red and indicated with red arrows. The CBM20 domain is shown in blue. (d) Scaled schematic representation of the PMST proteins identified in the *Chlorella* genomes. The TargetP 2.0 predicted cTP was removed prior to alignment. The CBM20 domain is shown in blue and the putative RBMs are shown as purple squares.

Given the highly conserved nature of the PMST proteins across the *Chlorella* species (Fig. 2-15c), it was expected there would be a similar level of conservation of any thylakoid matrix tether proteins analogous to RBMP1/2 of Chlamydomonas in the *Chlorella* species. By contrast, amongst the RBM-containing sequences identified in each of the species analysed, there existed a range of potential thylakoid interaction mechanisms. Given the putative role in tethering the pyrenoid matrix to the traversing thylakoid network of these mechanisms, these proteins are hereafter termed pyrenoid matrix thylakoid tether (PMTT) proteins.

In the genome of *Chlorella sorokiniana* UTEX 1230, there exists a multiple RBM-containing protein (CSI2_123000003237) fused to 2 additional structural domains; an N-terminal coiled-coil and a C-terminal rhodanese-like domain (Fig. 2-16a,c). Interestingly, the protein is predicted to have a Twin-Arginine Translocation (TAT) thylakoid luminal transfer peptide immediately followed by a potential transmembrane domain (Fig. 2-16c,d), which would presumably anchor the protein in the thylakoid membrane, leaving the C-terminal RBMs and Rhodanese exposed in the chloroplast stroma (Fig. 2-16b). The stromal exposed RBMs could then presumably interact with condensed Rubisco in the pyrenoid matrix and act to tether the condensate to the thylakoid. The potential role of the Rhodanese domain, or how the PMTT protein could be positioned correctly within the thylakoid network are unclear from the domain composition.

An analogous but structurally and sequence distinct PMTT protein was also identified in the genome of *Micractinium conductrix* (rna-C2E20_4272), in which a PsbP-like domain is present fused to the multiple RBMs and a C-terminal Rhodanese domain (Fig. 2-17). In line with the function of PsbP domains at the thylakoid luminal side of photosystem II (Ifuku *et al.,* 2008), the identified protein also contains a predicted TAT thylakoid luminal transfer peptide. Although there is no clearly predicted transmembrane domain following the PsbP domain, it is likely the RBMs and Rhodanese domain are somehow exposed in the chloroplast stroma also to allow interaction with Rubisco (Fig. 2-17b). The conservation of the Rhodanese domain between the PMTT proteins of *Micractinium conductrix and Chlorella sorokiniana* suggests the domain plays a role in the functionality of the protein.

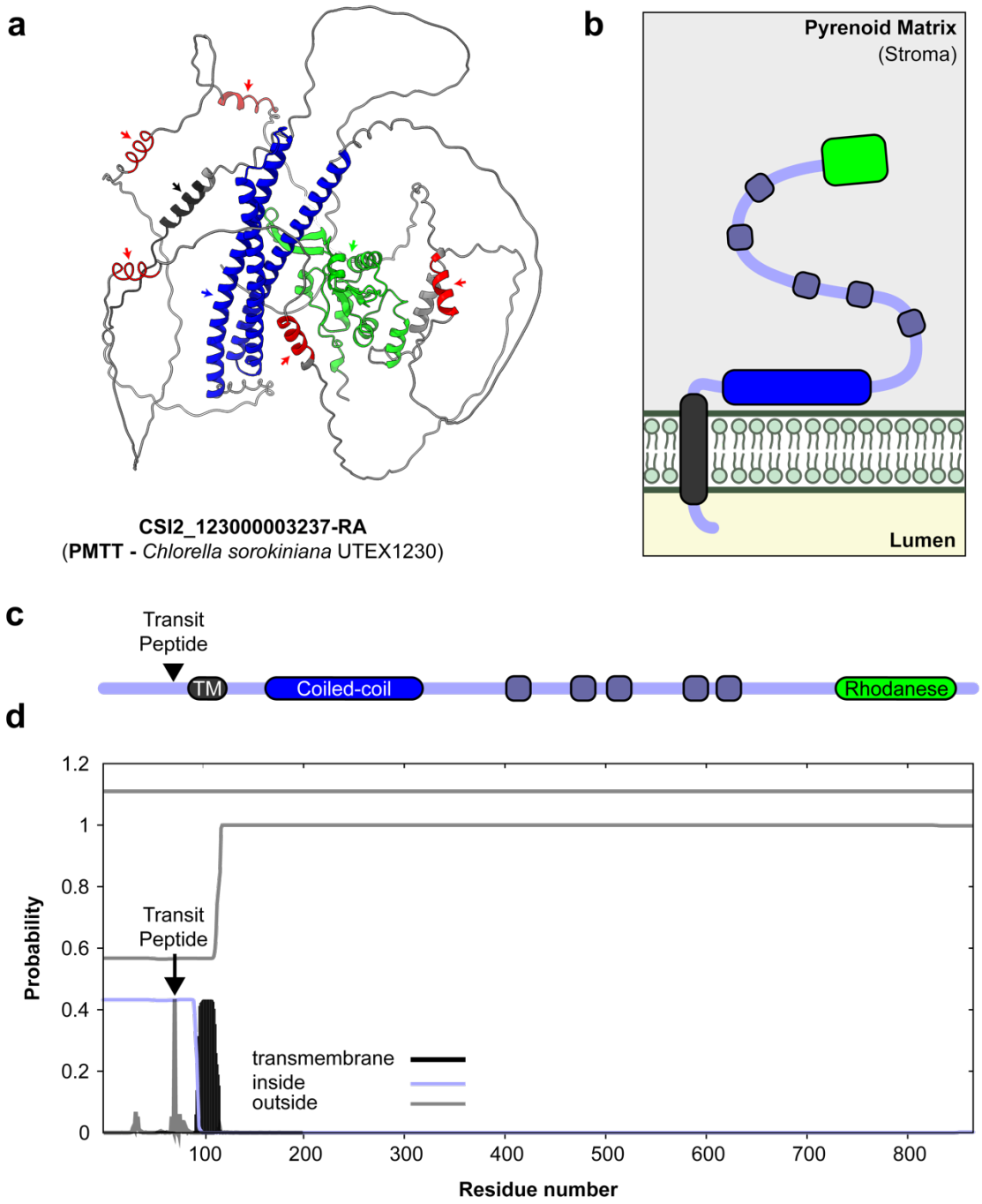


Figure 2-16 | Characterisation of a potential thylakoid matrix tether from *Chlorella sorokiniana* UTEX1230. (a) Top ranked AlphaFold2 model of the putative pyrenoid matrix thylakoid tether (PMTT) from *Chlorella sorokiniana* UTEX1230 (CSI2_123000003237-RA). The predicted transmembrane is shown in black and indicated by a black arrow. The predicted coiled-coil is shown in blue, the putative RBMs in red, and the C-terminal Rhodanese in green. (b) Schematic model for PMTT function at the pyrenoid matrix thylakoid interface. (c) Scaled schematic of PMTT shown annotated with the transit peptide, transmembrane domain (blue), RBMs (purple squares) and Rhodanese (green) corresponding to Fig. 2-16a. (d) Overlaid TMHMM (grey) and TargetP 2.0 predictions. The predicted thylakoid luminal transfer peptide is shown in grey and indicated by an arrow. The predicted transmembrane helix is shown in black.

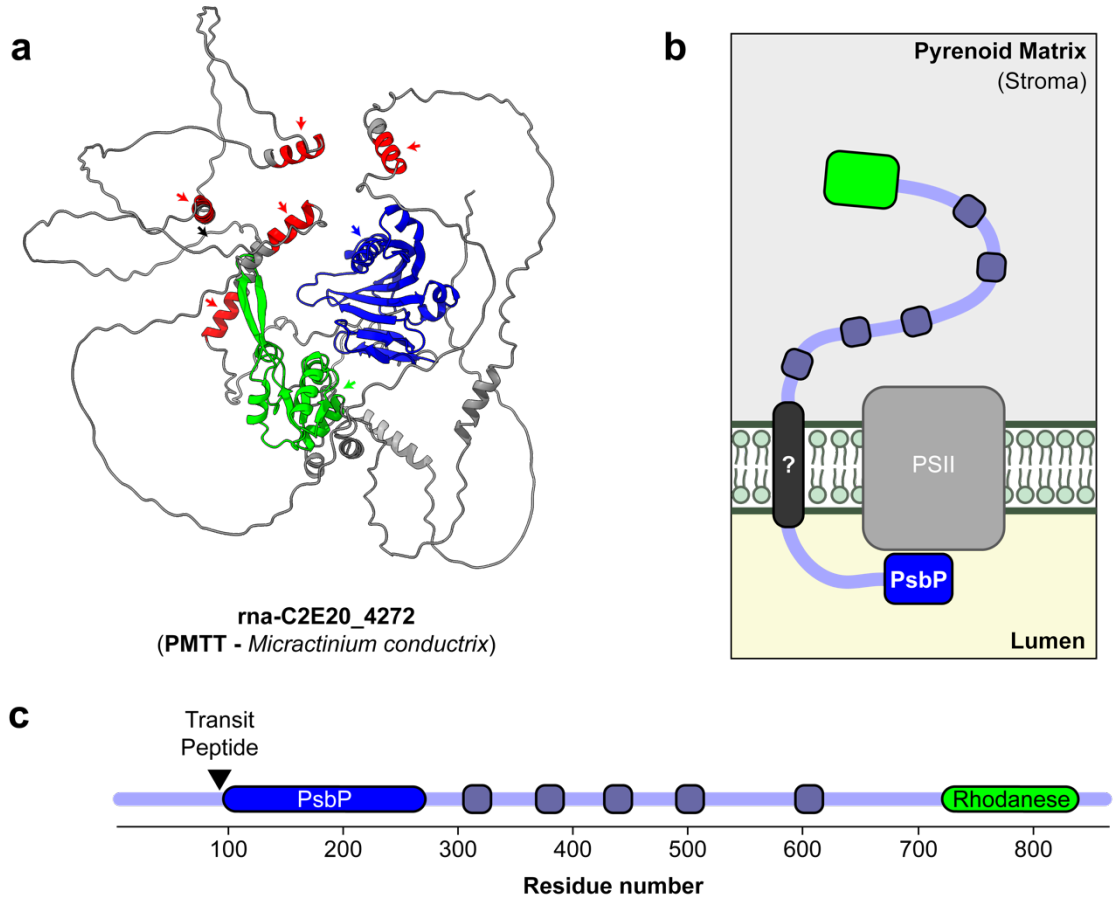


Figure 2-17 | Characterisation of a potential thylakoid matrix tether from *Micractinium conductrix*.
(a) Top ranked AlphaFold2 model of the putative pyrenoid matrix thylakoid tether (PMTT) from *Micractinium conductrix* (rna-C2E20_4272). The predicted PsbP is shown in blue, the putative RBMs in red, and the C-terminal Rhodanese in green. (b) Schematic model for *Micractinium* PMTT function at the pyrenoid matrix thylakoid interface. The proposed transmembrane location is indicated by the question mark. Photosystem II (PSII) is also indicated in the thylakoid membrane. (c) Scaled schematic of PMTT shown annotated with the transit peptide, PsbP domain (blue), RBMs (purple squares) and Rhodanese (green) corresponding to Fig. 2-17a.

A third class of putative PMTT analogue was identified in the genome of *Chlorella vulgaris* (KAI3429225), in which a predicted methyltransferase appears fused to multiple C-terminal RBMs (Fig. 2-18). The protein also contains a predicted N-terminal transmembrane domain, which would presumably anchor the protein in the thylakoid membrane (Fig. 2-18b). A VAST search of the AlphaFold2 predicted structure of the methyltransferase domain revealed homology to a Magnesium Protoporphyrin IX Methyltransferase (PDB: 4QDK), which has previously been shown to localise to the thylakoid network, possibly by embedding in the outer leaflet of the membrane (Block *et al.,* 2002). Together this evidence suggests the protein could be membrane-embedded with the C-terminal RBMs exposed in the chloroplast stroma (Fig 2-18b).

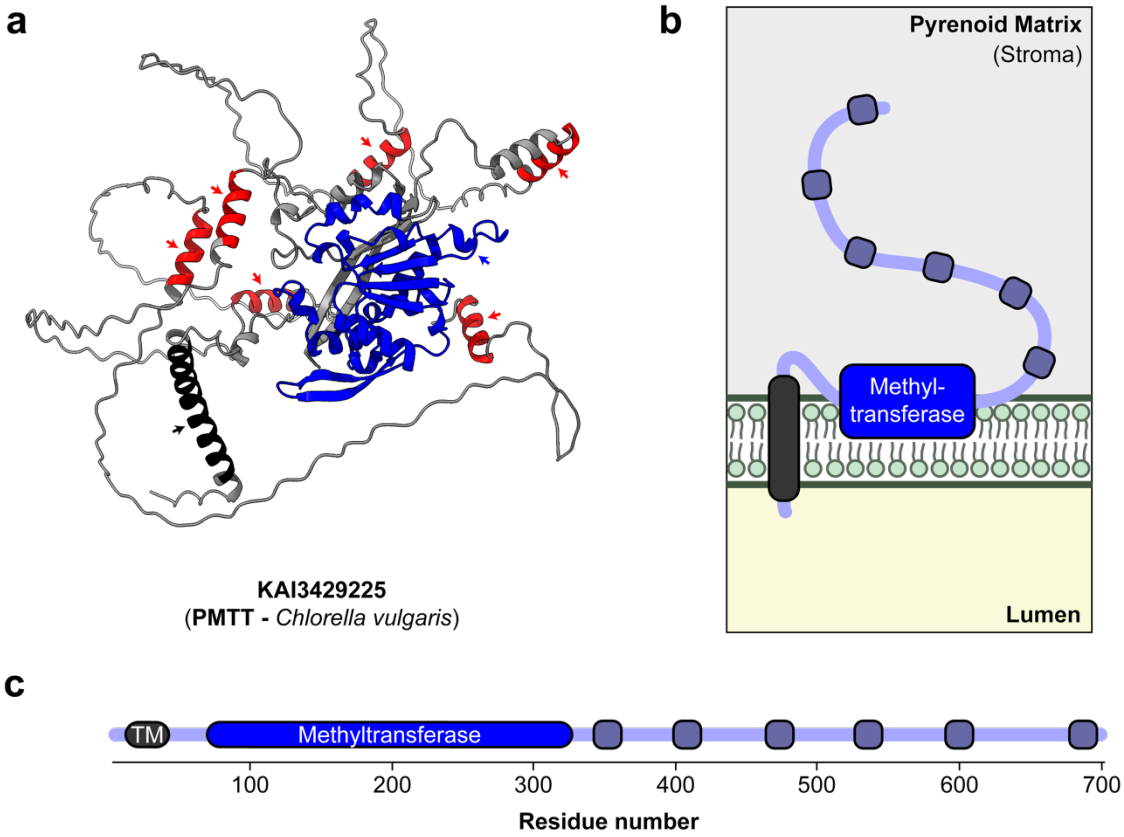


Figure 2-18 | Characterisation of a potential thylakoid matrix tether from *Chlorella vulgaris*.
(a) Top ranked AlphaFold2 model of the putative pyrenoid matrix thylakoid tether (PMTT) from *Chlorella vulgaris* (KAI3429225). The predicted transmembrane domain is shown in black, the predicted Magnesium Protoporphyrin IX Methyltransferase is shown in blue and the putative RBMs in red. (b) Schematic model for *Chlorella vulgaris* PMTT function at the pyrenoid matrix thylakoid interface. Homologues of the methyltransferase in other species insert into one leaf of the thylakoid membrane, as depicted here. (c) Scaled schematic of PMTT shown annotated with the transmembrane (black), Magnesium Protoporphyrin IX Methyltransferase (blue) and RBMs (purple squares) corresponding to Fig. 1-18a.

Taken together, the identification of three apparently analogous PMTT proteins in the *Chlorella* species suggest that the general approach of RBM-guided assembly of the pyrenoid with the traversing thylakoid may be conserved, but that the specifics of the approach used have evolved separately within each of the species. This appears to contrast with the putative linker and PMST proteins which demonstrated largely similar sequence and structural arrangement (Fig. 2-12, Fig. 2-15), consistent with their inheritance from a common ancestor and subsequent divergent evolution over the last ~165 My. The distinct structural arrangement of the PMTT proteins implies they each evolved following the divergence from their closest ancestor with a different PMTT analogue, constraining their evolution to the last ~165 My when *Chlorella sorokiniana* and *Micractinium conductrix* diverged (Fig. 2-15c).

Across all 3 analogous PMTT proteins the spacer region between the RBMs was akin to that of the linker sequence. As discussed with the PMST proteins, it is unclear whether the RBM-containing PMTT sequences evolved from a pseudo-linker sequence that was commonly inherited then fused to thylakoid-interacting domains or whether the sequences arose *de novo* within the respective species. Further work to decipher this mechanism would provide great insight into the evolution of pyrenoid assembly proteins, and more generally the evolution of functional phase-separated condensates in biology. Of course, empirical evidence for the functional role of these proteins is still required, but the *in silico* analyses suggest two interesting hypotheses. Firstly, it seems likely that the evolution of PMTT proteins followed the evolution of the linker and PMST proteins, which were presumably also evolved in a stepwise manner. This suggests a possible stepwise assembly of the pyrenoid throughout evolutionary history, in which a linker protein first evolved to condense Rubisco into the pyrenoid matrix ('naked'), followed by evolution of PMST proteins to encase the condensate in starch ('sheathed') and finally evolution of PMTT proteins to allow traversion of a thylakoid through the matrix ('traversed') (Fig. 2-19). Presumably for this hypothesised stepwise assembly to be feasible, each stage must have provided some physiological advantage (or at least no physiological detriment) for the respective hypothesised alga. In line with this hypothesis, the diversity of pyrenoid ultrastructure across the eukaryotic tree of life certainly suggests these 'intermediate' assembly steps are feasible to support CO_2-concentrating mechanisms, with 'naked', 'sheathed' and 'traversed' pyrenoids all observed in extant species (Fig. 1-9) (Barrett *et al.,* 2021). In support of this hypothesis, modelling approaches have suggested that condensation of Rubisco with or without a co-condensed carbonic anhydrase would provide a marked increase in Rubisco carboxylation rate (Long *et al.,* 2021). Separate modelling approaches have also suggested a marked benefit to adding a diffusion barrier surrounding the Rubisco condensate to prevent CO_2 leakage (Fei *et al.,* 2022).

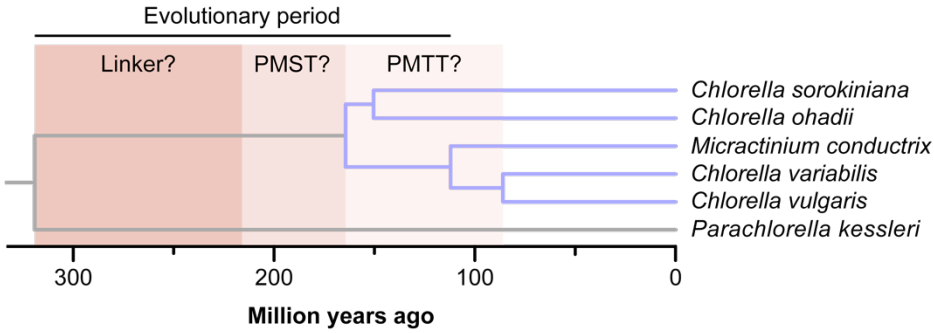


Figure 2-19 | Proposed evolutionary periods for pyrenoid components in *Chlorella*.
Phylogenetic chronogram of *Chlorella* linker-containing species relative to *Parachlorella kessleri*. The proposed evolutionary periods of pyrenoid assembly periods are indicated by the red shading.

The second hypothesis that can be derived from the identified sequences is based on the properties of the RBM-containing regions of the proteins. As discussed, these regions appear to present as pseudo-linkers with largely regular spacing of the putative RBMs by sequence that has properties akin to that of the linker. This is in contrast to the spacer sequences between RBMs in the RBMP and SAGA proteins of Chlamydomonas which show largely different properties to those of EPYC1 (Meyer, Itakura, *et al.,* 2020). The pseudo-linker nature of the *Chlorella* PMST and PMTT proteins suggest that they may have evolved from a common origin linker sequence as discussed in the context of stepwise assembly. Though little evidence for this hypothesis is presented here, future comparison of the spacer properties within the respective species may uncover a molecular clock that would allow distinction of the timing and origin of evolution of the RBM-containing proteins. Likewise, similar analyses with the RBMP and SAGA proteins within the Chlamydomonadales genomes would allow an independent point of reference for the evolution of RBM-containing proteins.

From the identification of analogous assembly proteins, an *in silico* derived model of pyrenoid assembly in *Chlorella sorokiniana* UTEX1230 could be proposed and compared with that of Chlamydomonas (Fig. 1-20). Chlamydomonas possesses a relatively complex pyrenoid ultrastructure, with numerous starch plates surrounding the matrix and a reticulated 'thylakoid knot' at the centre that also contains so-called 'minitubules' (Fig. 2-20a) (Engel *et al.,* 2015). Comparatively, the *Chlorella* pyrenoid is surrounded by only two starch plates, which are split by the traversion of two thylakoid sheets through

the matrix (Fig. 2-20b). In line with the simpler pyrenoid morphology and smaller genome size of *Chlorella*, the number of proteins required to assemble the pyrenoid with its ultrastructural features appears to be fewer than that of Chlamydomonas. The potentially simpler assembly blueprints of the *Chlorella* pyrenoid relative to that of Chlamydomonas are of great interest with respect to ongoing efforts to engineer functional pyrenoid-based CCMs in plants (Adler *et al.,* 2022; Atkinson *et al.,* 2020). In the context of work presented in Chapters 4 and 5, engineering of the *Chlorella* pyrenoid *in planta* accordingly becomes a tangible goal with much less engineering of the host plant required for potential function, relative to the current efforts with the Chlamydomonas system. The identification of clearly analogous parts from different species of *Chlorella* that likely utilise the same putative RBM also presents an opportunity to combine parts from different *Chlorella* species, as well as with those of Chlamydomonas presuming the RBMs can be functionally replaced. In this sense, the comprehensive assessment of assembly proteins from a range of *Chlorella* species was a valuable approach that allowed hypotheses of evolutionary origin and timing as well as providing a catalogue of potential pyrenoid assembly proteins.

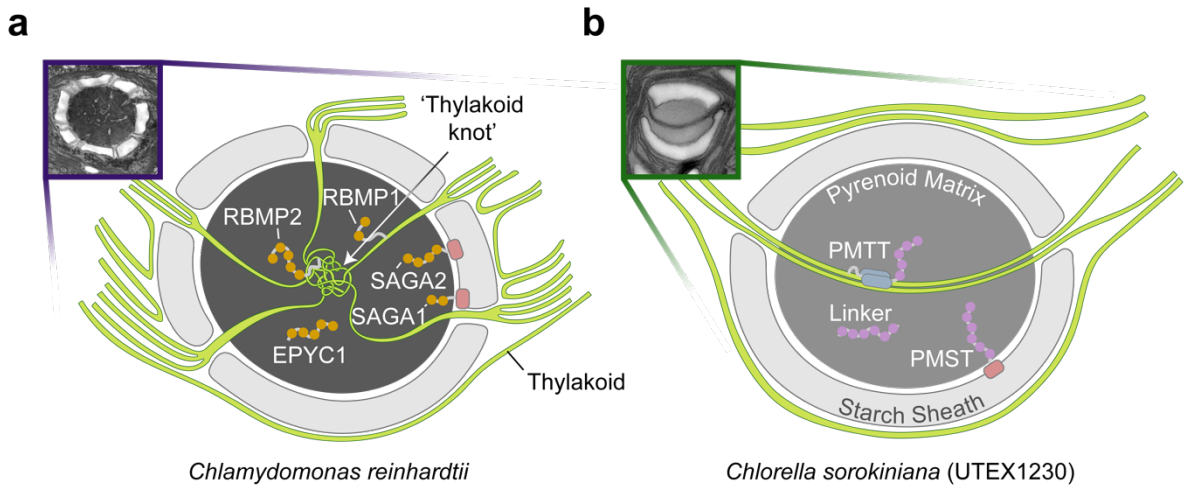


Figure 2-20 | Comparison of the *in silico* assembled pyrenoid of *Chlorella* with that of *Chlamydomonas*. (a) Schematic model of the *Chlamydomonas reinhardtii* pyrenoid shown relative to a transmission electron micrograph of a pyrenoid *in vivo* (top left) (Jessi Hennacy – Princeton University). The RBM-containing proteins hypothesised by Meyer *et al.,* (2020) to guide assembly of the pyrenoid matrix are shown schematically in the localised positions with the RBMs indicated by yellow circles. The thylakoids are shown in green, with the starch sheath in grey, surrounding the pyrenoid matrix (dark grey) (b) Schematic model of the *Chlorella sorokiniana* UTEX1230 pyrenoid shown relative to a transmission electron micrograph of a pyrenoid *in vivo* (top left) (Huang *et al.,* 2021). The identified RBM-containing proteins hypothesised to guide assembly of the pyrenoid matrix are shown schematically in the hypothesised positions with the RBMs indicated by purple circles. The thylakoids are shown in green, with the starch sheath in grey, surrounding the pyrenoid matrix (dark grey).

2.5. Dating linker evolution in the UTC clade

The identification of novel sequence-disparate linker analogues provided the unique opportunity to make more general comparisons about the distribution of the respective sequences with respect to their phylogenetic evolution. The identification of putative linkers in *Ulva* and *Chlorella* species provided representative sequences in the Ulvophyceae and Trebouxiophyceae classes respectively which, alongside EPYC1 of Chlamydomonas within the Chlorophyceae, gave representative sequences from each the classes within the UTC clade for comparison. UTC algae are exceptionally well represented with respect to genome sequences relative to other algal clades where linkers were identified (Rhodophyta, Bacillariophyta), making them particularly suited to subsequent analysis.

2.5.1. Construction of a phylogenetic chronogram for the UTC clade

To compare the likely evolutionary timepoint of each of the linker sequences, a consensus time-calibrated phylogenetic chronogram of sequenced pyrenoid-containing algae in the UTC clade was constructed. To do so, published phylogenetic chronograms were collated and a combined tree was built, with the values of key nodes presented as an average of published values (Fig. 2-21) (see methods). In some cases, published values precluded construction of the downstream tree and were excluded from the construction (asterisks in figure). A large proportion of the tree building was made possible by a recently published phylogenetic chronogram of green algae (Cortona *et al.,* 2020), with reliance on more focussed phylogenies of *Chlamydomonas* (Herron *et al.,* 2009) and *Chlorella* (Wang *et al.,* 2020) species for their construction. Generally, the phylogenies of each of the representative phylogenetic orders of *Ulva*, *Chlorella* and *Chlamydomonas* species were well-represented in the tree, relative to other species where the number of sequenced closely related species was fewer. In agreement with recent estimates (Cortona *et al.,* 2020), the chronogram indicated divergence of the Ulvophyceae, Trebouxiophyceae and Chlorophyceae ~800 million years ago (Mya), which themselves basally diverged from a hypothetical ancestral green flagellate ~1000 Mya.

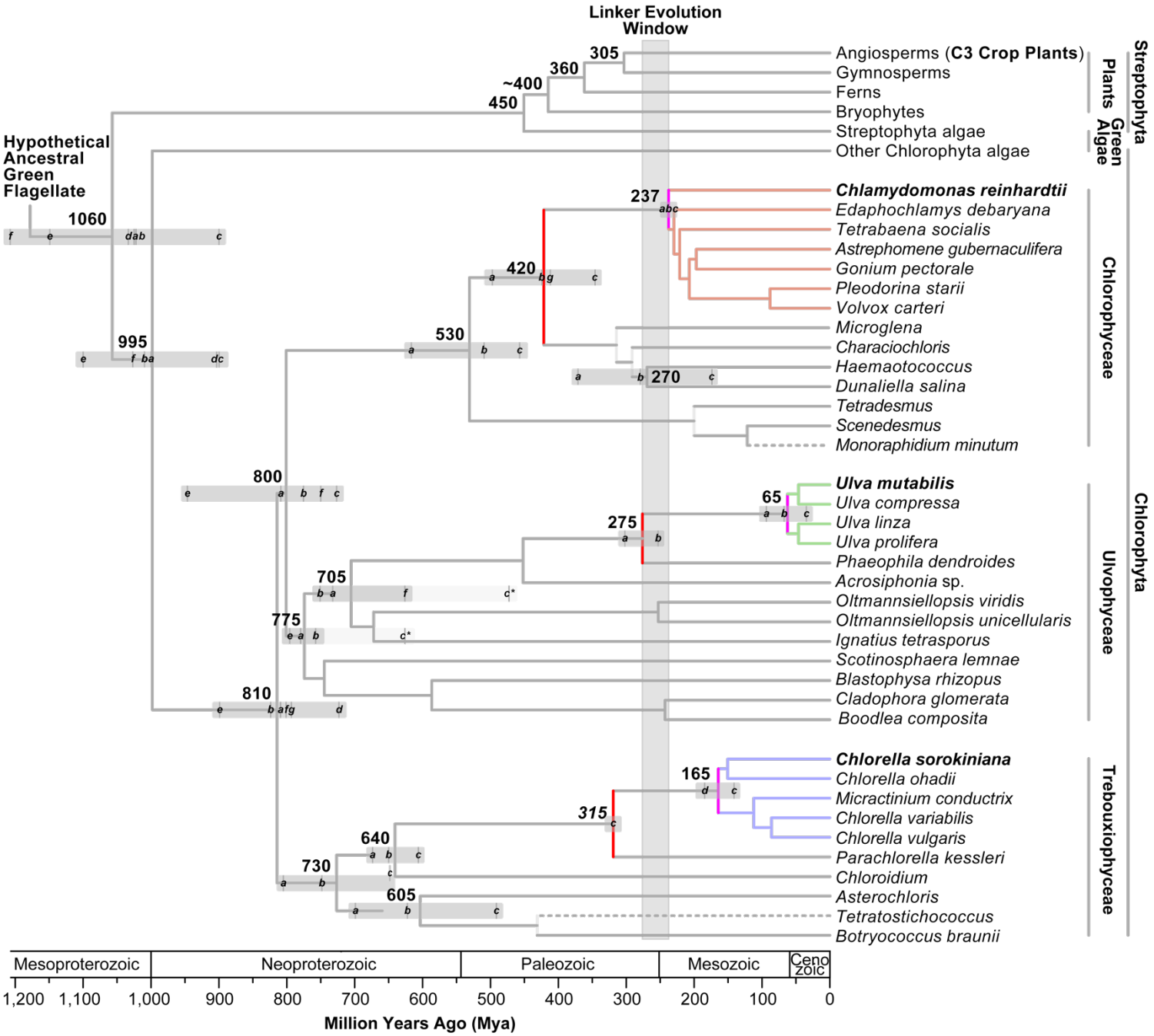


Figure 2-21 | Consensus phylogenetic chronogram of the UTC clade.
Consensus phylogenetic chronogram of sequenced pyrenoid-containing species in the UTC clade relative to other key taxa drawn from existing literature values, indicated by the letters positioned at branch points. Data for points a and b were taken from Cortona *et al.,* (2020) figure 1 and figure S9b respectively, c from Jackson *et al.,* (2018), d from Wang *et al.,* (2020), e from Munakata *et al.,* (2016), f from Herron *et al.,* (2009) and g from Davidi *et al.,* (2023). Major divergence points are positioned at an average value with a grey window bounded by the range of literature values, shown as lines within the window for those species. Asterisks indicate values that were excluded from averaging as they prevented assembly of the downstream tree when included. Dashed lines indicate phylogenetic placements that are uncertain or disputed in literature. Light grey divergence points indicate timepoints that are undefined and estimated from non time-calibrated reference phylogenies. For each of the key linker sequences identified, the 'representative' species (*Chlamydomonas reinhardtii*, *Ulva mutabilis* and *Chlorella sorokiniana*) is bolded. The extent of the phylogenetic sequence conservation of the linker sequence from that species, as determined by BLAST analysis is indicated by the colouring of the adjacent phylogenetic branches. For the detailed phylogeny of these groups, data was used from Herron *et al.,* (2009) for *Chlamydomonas*, Cortona *et al.,* (2020) for *Ulva* and Wang *et al.,* (2020) for *Chlorella*. The linker evolution window is defined as the common period in which each of the representative species was diverged from its closest ancestor without the linker sequence (red branches), but not diverged from its closest ancestor possessing the linker sequence (magenta branches), namely 237-275 Mya.

2.5.2. Identification of a shared timepoint for linker evolution

From the phylogenetic chronogram an evolutionary timepoint for each of the linker sequences could be hypothesised based on two previous observations. Firstly, the lack of sequence homology outside of closely related species for each of the linker sequences (Table 2-1, Table 2-7) suggested that the linkers evolved following divergence from the closest ancestor that did not contain the same linker sequence (closest divergence branch to the base of each group of sequences in Fig. 2-21, shown in red). Equally, the structural and sequence similarity of each of the linker sequences within the closely related species suggested that the linker sequence was inherited from a common ancestor to all extant species which predated the divergence of the two earliest diverged species within each of the closely related linker-containing species (earliest divergence branch in each group in Fig. 2-21, shown in magenta). From these two limitations, a hypothetical linker evolution window could be determined for each of the sequences (Table 2-10). For each of the respective linker sequences, the timepoints for the divergence of the closest ancestor without the respective linker sequence was largely agreeable (red branches; 275-420 Mya). Likewise, the divergence timepoint for the closest ancestor with the respective linker sequences was agreeable for *Chlorella* and *Chlamydomonas* (magenta branches; 165-237 Mya) though appeared much later for the *Ulva* sequence (65 Mya), likely due to the lack of sequence information for earlier diverged species in the Ulvales (Fig. 2-21). Taking the latest of the 3 timepoints for the divergence of the species without the respective linker (latest red branch - *Ulva*; 275 Mya), and the earliest divergence timepoint for the species with the respective linker (earliest magenta branch - *Chlamydomonas*; 237 Mya), a minimal shared hypothetical 'linker evolution window' could be derived (237-275 Mya) (Fig. 2-21). Clearly this hypothesised window is based on several assumptions and is determined by timepoints which themselves come with a large degree of error, though together the results suggest a likely timepoint for linker evolution at the end of the Paleozoic Era. This hypothesis could be further strengthened with two types of additional data.

Firstly, the number of sequenced genomes in *Ulva* species limits the analysis for the closest ancestor with the identified linker sequence. Therefore, assessment of linker sequence conservation in species of the Ulvales that diverged between the 65-275 Mya timeframe would allow a more accurate revision of

the latest timepoint at which the putative *Ulva* linker could have evolved (the magenta branch in the *Ulva* group). Likewise, assessment of species that diverged closer than the current sequenced closest ancestor without each respective linker would allow more accurate assessment of the earliest timepoint at which each respective linker could have evolved (red branches in each group). The second type of data that would allow testing of the hypothesised linker evolution window would be to identify and validate linker sequences in species that diverged within the hypothesised linker evolution window. For example, assessing whether *Dunaliella salina* and *Haematococcus* species, which diverged ~270 Mya (Fig. 2-21), contain the same linker sequence would either support or allow revision of the hypothesised linker evolution window. Likewise, identification and validation of additional linker sequences in species that diverged either before or after the hypothesised window would allow further assessment. Regardless, the identification of a shared window for linker protein evolution in the late Paleozoic Era in 3 separate classes of algae that themselves diverged ~800 Mya is unlikely coincidental and certainly suggests a common pressure to evolve this mechanism.

Table 2-10: Hypothesised timepoints for linker evolution.

		Divergence timepoint of closest ancestor	
Genus	Phylogenetic Order	**without** respective linker	**with** respective linker
Ulva	Ulvophyceae	275 Mya	65 Mya
Chlorella	Trebouxiophyceae	315 Mya	165 Mya
Chlamydomonas	Chlorophyceae	420 Mya	237 Mya

2.5.3. *Linker evolutionary timing in the context of historical environmental conditions*

To try and provide a rationale for the proposed linker evolutionary window derived from sequence analysis, it is useful to look to past analyses of atmospheric conditions as well as other environmental factors.

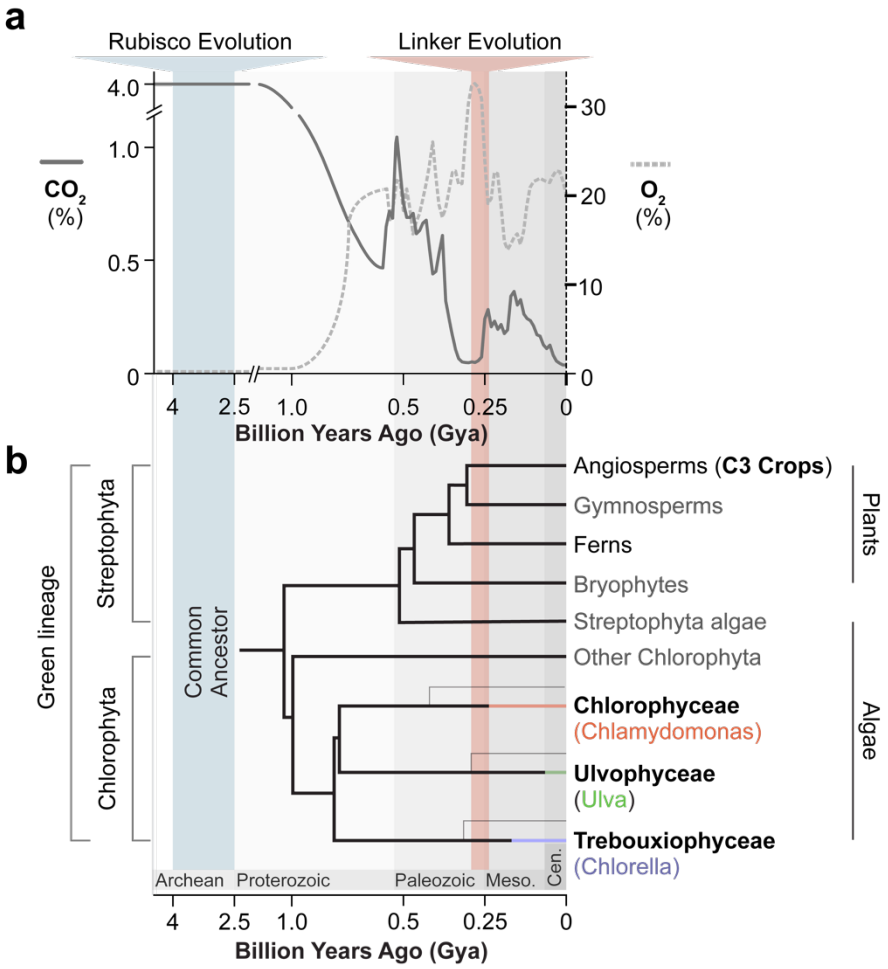


Figure 2-22 | Atmospheric conditions in the context of linker phylogeny and evolution.
(a) Atmospheric CO_2 (solid line) and O_2 (dashed line) levels over the last 4 billion years. An estimate for pre-Paleozoic atmospheric conditions from Catling and Zahnle (2020) is depicted. Estimates for the last ~500 million years were taken from GEOCARB III and GEOCARBSULFO models (Berner, 2009; Berner & Kothavala, 2001). (b) Simplified consensus phylogenetic chronogram calibrated to the timescale of Fig. 1-22a. For the three primary Chlorophyta classes (bolded), the coloured and grey diverged branches represent the closest ancestors with and without sequence homologues of the representative sequences as presented in Fig. 1-21. Red shading indicates the hypothesised linker evolution window. Blue shading indicates the evolutionary window for Rubisco.

Form I Rubisco almost certainly evolved in the Archean Eon ~2.5-4 billion years ago (Gya) (Shih *et al.,* 2016) when estimates suggest atmospheric CO_2 was in vast excess over O_2 (Fig. 2-22a) (Catling & Zahnle, 2020). Presumably due to evolution of oxygenic photosynthesis (Kasting, 2013), over the following 2 billion years (0.5-2.5 Gya) the levels of atmospheric oxygen rose to current day levels,

whilst concurrently CO_2 levels dropped to ~0.5%. Over the last 500 million years, CO_2 levels generally reduced, with a minimum reached in a window at the end of the Paleozoic Era (~260-330 Mya) that was comparable to present day atmospheric levels and has been accredited to the burial of organic matter (Berner & Kothavala, 2001). O_2 levels over the same time period were relatively consistent, with a major peak at the end of the Paleozoic Era (260-300 Mya) accredited to the radiation of vascular plants (specifically gymnosperm and angiosperm trees) that increased chemical weathering of silicates (Berner, 2006). Interestingly the O_2:CO_2 maxima at the end of the Paleozoic Era (~260-300 Mya) aligns well with the hypothesised linker evolution window (Fig. 2-22), suggesting atmospheric conditions during this timepoint provided significant evolutionary pressure to evolve compensatory mechanisms.

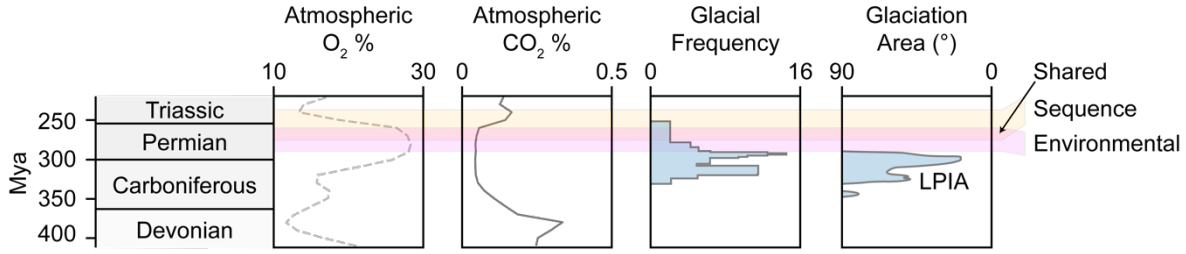


Figure 2-23 | Environmental conditions in the context of linker evolution.
Atmospheric and ice conditions in the late-Paleozoic early-Mesozoic Eras. The atmospheric O_2 and CO_2 are presented as in Fig. 2-22. Glacial frequency (number of major glacial events at each time point) and glaciation area (presence of ice at the indicated paleolatitude) are adapted from (Montañez & Poulsen, 2013).The hypothesised linker evolution window from sequence analysis is indicated by the yellow horizontal box ('Sequence') and the feasible evolutionary window from environmental conditions is indicated by the magenta box ('Environmental'). The crossover is indiated.

The proposed linker evolution window also follows the demise of the Late Paleozoic Ice Age (LPIA), in which glaciation events almost altogether diminished from ~275-290 Mya in all but the southern hemisphere (Fig. 2-23) (Montañez & Poulsen, 2013). Presumably evolution during the LPIA would have been somewhat slowed by the freezing temperatures and global ice coverage, especially for aquatic organisms, suggesting the post-LPIA Paleozoic atmosphere (~250-275 Mya) may have been the earliest environment in which pyrenoid-based CCMs could have evolved in these species. Melting of ice following the demise of LPIA may also have provided ecological niches for these algal CCMs to evolve, and the locale of the early LPIA demise presumably excluded evolution of the linker sequences in the southern hemisphere. Evolution of pyrenoid-based CCMs following the Carboniferous period (~300-

360 Mya) also agrees with previous arguments that evolution and retention of eukaryotic CCMs through earlier periods of high CO_2 would have been unlikely (Raven *et al.,* 2017). Similar arguments have also been made for the evolution of prokaryotic carboxysome-based CCMs in the same period (Badger *et al.,* 2002). Finally, it seems logical that evolution of the pyrenoid-based CCM would be constrained to the period of highest evolutionary pressure, *i.e.,* before the rise of CO_2 and fall in O_2 at the end of the Permian period (Fig. 2-23). From these observations, a proposed evolutionary window for linker proteins within the Paleozoic Era can be proposed from the environmental conditions alone. Namely, the period 260-290 Mya when the O_2:CO_2 was maximal and the LPIA had demised in all but the southern hemisphere ('Environmental' in Fig. 2-23). Combining this window with the window derived from sequence analysis alone ('Sequence' in Fig. 2-23), the most likely window for linker and pyrenoid evolution in the UTC clade appears to be 260-275 Mya ('Shared' in Fig. 2-23).

The identification and determination of the time-constrained evolutionary timepoints of 3 independently evolved pyrenoid linker proteins in the UTC clade provides some of the strongest support for the previously hypothesised window of pyrenoid-based CCM evolution in chlorophyte algae. In contrast, other pyrenoid-based CCMs, such as those in hornworts, appear to have evolved much later, with the majority proposed to have evolved in the last 35 My (Villarreal & Renner, 2012). Likewise, the lack of PYCO1 conservation between *Phaeodactylum* and *Thalassiosira*, which diverged ~90 Mya, presumably suggests a similarly recent origin for evolution of pyrenoids in the Bacillariophyta. Further evidence is required to make conclusions about evolution of pyrenoids in red algae, though it seems likely a similar evolutionary timepoint as observed for the Chlorophyte algae is conserved, given their similarly ancient divergence (Brawley *et al.,* 2017). Similar analyses as presented here in these pyrenoid-containing species may shed light on the evolutionary drivers for pyrenoid evolution. Taken together the observations suggest that pyrenoids in the UTC clade likely evolved in response to the O_2:CO_2 maxima observed following the Carboniferous period that was a product of rapid oxygenation by the radiation of vascular plants.

2.6. Discussion

The aim of this chapter was to identify hypothesised linker sequences in unexplored pyrenoids, that have conserved physicochemical properties to those of pyrenoids that have been previously characterised (*i.e.,* Chlamydomonas and Phaeodactylum). To do this, a highly selective (0.09% ± 0.13% S.D. of the input sequences) Fast Linker Identification Pipeline for Pyrenoids (FLIPPer) was constructed and tested, which demonstrated filtering of whole genome annotated sequences to a number of sequences that could feasibly be individually interrogated to determine their likelihood as *bona fide* linker proteins. In line with the rigorous testing and development of the pipeline, FLIPPer demonstrated good specificity. The pipeline robustly identified sequences that were largely disordered, with repeating structural elements that contained likely Rubisco-interacting residues across the ~100 genomes analysed. Through this, a large catalogue of candidate pyrenoid linker proteins were outputted in a wide range of phylogenetic taxa for further analysis. From the genomes analysed, 3 novel linker proteins were interrogated as high confidence outputs, and for each sequence, preliminary validation demonstrated good likelihood that the identified sequences were indeed pyrenoid linker proteins.

The similarity of structural arrangement and physicochemical properties but almost total lack of sequence identity in the 3 high confidence linkers of species separated by at least 800 million years of evolution provided the first direct evidence for the long-standing hypothesis that linker proteins, and presumably in turn pyrenoids, have evolved independently and convergently within their respective phylogenetic taxa. The identification of the putative Rubisco-binding motif in the linkers further allowed interrogation of the hypothesis that RBM-guided assembly of the pyrenoid may have also evolved convergently as a general mechanism to arrange the ultrastructural features of a pyrenoid. For *Porphyridium* and Ulva only direct evidence could be found for a potential tether between the pyrenoid matrix and traversing thylakoid membranes, suggesting that this feature at least may be conserved. In contrast, analysis of the large number of *Chlorella* genomes allowed rigorous assessment of RBM-guided assembly in the Trebouxiophyceae, where both pyrenoid matrix starch- and thylakoid-tether (PMST and PMTT) protein candidates were observed across species, suggesting convergent co-evolution of the pyrenoid assembly machinery in *Chlorella* as proposed in Chlamydomonas. The

apparently non-conserved mechanisms of the PMTT proteins in their association with the thylakoid membranes suggested separate convergent evolution of these mechanisms following the divergence of the *Chlorella* species ~150 Mya, in contrast to the linker and PMST proteins which appear to have been inherited from the common ancestor to all extant species and subsequently divergently evolved. The identification of clear linker, PMST and PMTT proteins in the genome of *Chlorella sorokiniana* UTEX1230 allowed a model for RBM-guided pyrenoid assembly to be proposed, which demonstrated simpler morphology than that of *Chlamydomonas*, in line with the ultrastructural arrangement of the *Chlorella* pyrenoid and provides good promise for plant engineering of pyrenoids.

Finally, the mapping of the UTC clade pyrenoid linker proteins to a consensus phylogenetic chronogram allowed identification of a shared evolutionary timepoint under which linkers are likely to have evolved. This timepoint of all 3 linkers aligned well with previously proposed windows for CCM evolution, in line with the vascular plant induced O_2:CO_2 maxima and provided direct evidence for concurrent, convergent evolution at this timepoint. The similarity of the 3 linker proteins across the species of the UTC clade was much higher than between the *Porphyridium* and *Phaeodactylum* linker proteins, suggesting that the specific evolutionary constraints at this timepoint in those species led to evolution of highly similar mechanisms for Rubisco condensation.

More generally, FLIPPer could likely be easily parameterised to allow identification of phase separating proteins in non-pyrenoid contexts given that the general principles of protein phase separation rely on disorder, multivalency and residue composition. Either combined with experimental data or for *de novo* discovery this tool could provide more rapid identification of the driver proteins for liquid-liquid phase separation in a wide range of biological contexts.

2.7. Materials and methods

2.7.1. General bioinformatic methods

2.7.1.1. Multiple sequence alignment

Multiple sequence alignments (MSAs) were produced using MAFFT v7.4 run in Geneious Prime 2023 or with the standalone MAFFT v7.5 terminal tool (Katoh & Standley, 2013). In both instances a gap opening penalty of 1.53 and an offset value of 0.123 were imposed. MSA figures were produced using Jalview v2.11 (Waterhouse *et al.,* 2009)

2.7.1.2. BLASTp analysis

Unless otherwise stated, all BLASTp analysis was completed using the online NCBI blastp (protein-protein BLAST) tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Sequences were compared against the non-redundant protein sequences (nr) database and the low complexity region filter was not imposed. The settings were otherwise default (100 max target sequences, 0.5 E threshold, 5 word size).

2.7.1.3. AlphaFold modelling

All AlphaFold2 modelling was completed using the "ColabFold v1.5.2.-patch: AlphaFold2 using MMseqs2" Google Colab tool (https://colab.research.google.com/github/sokrypton/ColabFold/) (Mirdita *et al.,* 2022). The modelling was completed with default settings. The top ranked model was presented in all cases, in which the per-residue pLDDT was mapped to structure using the AlphaFold Error Plot function in UCSF ChimeraX (Pettersen *et al.,* 2021)

2.7.1.4. Transmembrane prediction with TMHMM

Transmembrane prediction was completed using the online TMHMM 2.0 online tool (Krogh *et al.,* 2001) (https://services.healthtech.dtu.dk/services/TMHMM-2.0/). The program was operated with default settings and the outputted graphics were used for figure production.

2.7.1.5. Chloroplast transit peptide prediction with ChloroP

Chloroplast transit peptide prediction was completed using the TargetP 2.0 online tool (Almagro Armenteros *et al.,* 2019) run in 'plant' mode with 'long' output format.

2.7.1.6. Consensus sequence logo production

Consensus sequence logos were produced using the WebLogo online tool (Crooks *et al.,* 2004) (https://weblogo.berkeley.edu/logo.cgi) with the outputted sequence alignments produced using MAFFT. Settings were otherwise left as default.

2.7.1.7. Primary sequence analysis

Primary sequence composition and isoelectric point analysis was completed using Biopython implementation of the ProtParam module (Cock *et al.,* 2009; Gasteiger *et al.,* 2005).

2.7.2. Repeat detection software comparison

For comparison of the efficacy of repeat detection software, the full length EPYC1 sequence (Cre10.g436550) was used. For RADAR, the online tool was used with default settings (Madeira *et al.,* 2022) (https://www.ebi.ac.uk/Tools/pfa/radar/). For HHrepID analysis, the MPI Bioinformatics Toolkit implementation was used with default settings (Biegert & Söding, 2008; Zimmermann *et al.,* 2018). For XSTREAM analysis XSTREAM v1.73 was run locally from terminal with the following parameters (-i 0.4, -I 0.4, -e 2, -g 50, -m 20, -x 120, -T 0.7). The same settings for XSTREAM were used in FLIPPer v0.

2.7.3. Disorder prediction with IUPred

Disorder prediction was implemented in FLIPPer v0 using the IUPred2A python code run from command prompt in 'long' mode (Dosztányi *et al.,* 2005). Graphical ouptuts were produced using the matplotlib plot python module.

2.7.4. FLIPPer v1

FLIPPer v1 code is available in a private GitHub repository (https://github.com/james-r-barrett/FLIPPer), and is included as a supplementary file (FLIPPer-main.zip) in this thesis submission. Implementation of sequence pre-filtering, XSTREAM analysis and IUPred analysis was completed as described previously with modifications as discussed in text. The repeat filtering module was built using ProtParam analysis of the consensus repeat sequences outputted from XSTREAM, that were extracted using a regex matching of the XSTREAM output html file converted using the BeautifulSoup python module (Richardson, 2007). IUPred score filtering was completed using pandas dataframe thresholding based on the average IUPred score for each sequence. The fasta validation module was implemented using the Biopython SeqIO module. Large scale implementation of FLIPPer v1 was completed using the default settings as defined in the GitHub deposited version.

2.7.5. Regular expression searches

The code used to complete regular expression searches of genomes is available in a public GitHub repository (https://github.com/james-r-barrett/RBM-Search).

2.7.6. Construction of the consensus phylogenetic chronogram for the UTC clade

As described in text, construction of the consensus phylogenetic chronogram was based on the extensive tree of (Cortona *et al.,* 2020). Positioning of each of the major branch points in the tree was determined from an average of all available literature values for each point as described in the figure legend of Fig. 2-21. In two instances values were excluded as the average branch value with their inclusion prevented construction of the downstream from the available values for those branch points. Specifically the two basal divergence values of the Ulvophyceae from Jackson *et al.,* (2018) were excluded. Several other branch points were estimated from the phylogeny of species, where definite values were not available. Specifically: *Microglena* and *Characiochloris, Characiochloris* and *Haemotococcus, Tetradesmus* and *Scenedesmus*, *Tetrastichococcus* and *Botrycococcus.* Detailed phylogenies for the groups where high confidence linker candidates were present (Chlamydomonas, Ulva and Chlorella) were taken from previous publications as described in Fig. 2-21.

3. CSLINKER IS A *BONA FIDE* PYRENOID LINKER PROTEIN

3.1. Introduction

In the previous chapter several high confidence pyrenoid linker proteins were identified using the Fast Linker Identification Pipeline for Pyrenoids (FLIPPer), for some of which preliminary evidence of their function was presented thanks to the generous contribution of collaborators. Most of the subsequent work focusses on the validation and characterisation of the candidate pyrenoid linker protein from *Chlorella* species. The *Chlorella* linker was selected for several reasons. 1) As a genus of the Trebouxiophyceae, *Chlorella* exists in the oldest diverged class of primary chlorophytes from the Chlorophyceae in which Chlamydomonas exists (ca. 810 Mya) (Fig. 2-22b). This fact provides the greatest potential to understand the apparent convergent evolution of linker proteins with respect to their general function and *modus operandi*. 2) Although the Chlorophytes split from plant-containing Streptophytes ~1 Gya (Fig. 2-21) the common green origin of their plastids, due to a hypothetical shared green flagellate ancestor (Leliaert *et al.,* 2011), provides a direct phylogenetic relationship between the species that increases the likelihood of direct compatibility of *Chlorella* pyrenoid components *in planta* (Atkinson *et al.,* 2016), relative to other non-green origin components (*e.g.,* from *Porphyridium* and *Phaeodactylum*). 3) Analysis of the putative RBM-containing proteins in the genome of *Chlorella* species indicated a simpler pyrenoid assembly framework relative to Chlamydomonas (Fig. 2-20), in line with its simpler pyrenoid morphology (Bertagnolli & Nadakavukaren, 1970; del Pino Plumed *et al.,* 1996; Ikeda & Takeda, 1995). The *Chlorella* pyrenoid therefore potentially offers an alternative, simpler system to engineer into plants.

Unlike the Chlamydomonadales order, where *Chlamydomonas reinhardtii* has been a clear model organism since the 1940s (Harris, 2001), there is no clear model species of the Chlorellales order. It was therefore required that a *Chlorella* species and strain be selected for validation of the linker protein. To date, >15 genomes representing 6 reported *Chlorella* clade species (*C. ohadii, C. dessicata, C. sorokiniana, C. vulgaris, C. variabilis, Micractinium conductrix*) exist. *Chlorella sorokiniana* is by the far the most sequenced species with the genomes of at least 8 different strains available (Arriola *et al.,* 2018; Hovde *et al.,* 2018; Wu, Li, *et al.,* 2019). *Chlorella sorokiniana* and its historical equivalent in

the literature, *Chlorella pyrenoidosa*, is also arguably the most studied of the *Chlorella* species, with a strong literature history of ultrastructural (Bertagnolli & Nadakavukaren, 1970; Guo *et al.,* 2022) and fundamental physiological (Bonaventura & Myers, 1969; del Pino Plumed *et al.,* 1996; Sun *et al.,* 2016; Treves *et al.,* 2016) studies. Of the *Chlorella sorokiniana* species genomes, that of strain UTEX1230 (SAG 211-8k, CCAP 211/8k, ATCC 22521, CCALA 259 (CCAO 259) equivalents) was determined to be of the highest quality with respect to assembly and annotation (Table 3-1). The UTEX1230 genome presented the lowest number of contigs at the full genome size, with the highest N50 and number of annotated protein sequences, representing a near chromosomal level assembly. *Chlorella sorokiniana* UTEX1230 (Chlorella hereafter) was also readily available from culture collections, and therefore used for the remainder of the characterisations.

Table 3-1: Comparison of *Chlorella sorokiniana* genome assembly quality statistics ordered by quality.

Metric	1230	1228	1412	1602	BD09	BD08	Dachan	SLA04
Genome size (Mb)	58.5	61.4	57.9	59.4	54.0	58.6	60.4	38.6
# of contigs	20	64	65	158	1,614	3,167	5,036	13
N50 (Mb)	3.82	2.41	2.20	2.59	3.58	3.63	2.58	3.05
# of annotated proteins	12,871	12,166	12,611	9,587	9,668	10,240	9,821	-
Average protein length	544	524	644	631	578	553	535	-

This chapter focusses on the validation of the identified linker protein from Chlorella and had several key aims, in line with the requirements of the characterised pyrenoid linker protein EPCY1 from Chlamydomonas. In short it was aimed to demonstrate 1) comparable *in vivo* linker and Rubisco abundances (Hammel *et al.,* 2018; Mackinder *et al.,* 2016), 2) *in vivo* interaction and co-localisation (Mackinder *et al.,* 2016, 2017), 3) *in vivo* function (Mackinder *et al.,* 2016) and 4) *in vitro* demixing of Rubisco by the linker protein (Wunder *et al.,* 2018).

3.2. In vivo and ex vivo validation of Chlorella pyrenoid linker identity

Given the purely *in silico* discovery of the *Chlorella* linker proteins, it was of particular importance to validate the identity and function of the protein *in vivo*. As there was a distinct lack of genetic tools, biochemical protocols, and database information available, this section as much encapsulates validation as it does method development.

3.2.1. Confirmation of CsLinker identity in vivo

The linker identified from the *Chlorella sorokiniana* UTEX1230 genome (CsLinker hereafter) has gene ID CSI2_123000012064-RA and consists a 364 amino acid protein encoded by a gene with a genomic sequence of 2,792 bp and a predicted gene model comprising 7 exons and 6 introns (Fig. 3-1a). To confirm the genomic sequence, the locus was PCR amplified in 4 parts (Fig. 3-1b) and sequenced. 100% coverage of the exon sequences was obtained, and the sequence relative to the reference genomic sequence was correct throughout, in line with the quality of genomic assembly of the strain (Table 3-1). Peptides from later introduced proteomics experiments were mapped to the predicted protein sequence, and for 5 of the 6 exon-exon junctions, intron-spanning peptides were obtained, allowing confirmation of the predicted gene model (Fig. 3-1a). Elsewhere coverage of the protein sequence was almost complete, with most excluded regions accounted for by the ambiguous peptides produced from trypsin digestion of those repeated regions. The TargetP 2.0 predicted chloroplast transit peptide (cTP) spans exons 1 and 2 (Fig. 3-1c), and accordingly no intron-spanning peptides were obtained for this region (Fig. 3-1a), suggesting the protein is indeed imported to the chloroplast as hypothesised. Taken together, these results confirm the identity, gene model and expression of the protein *in vivo*. With the protein identity confirmed, the next aim was to determine the *in vivo* abundance of CsLinker relative to that of Rubisco, given that EPYC1 and Rubisco were determined to be equimolar in *Chlamydomonas* (Hammel *et al.,* 2018). In order to do this, purified protein standards of both CsLinker and Rubisco were required, such that absolute quantification mass spectrometry experiments could be completed.

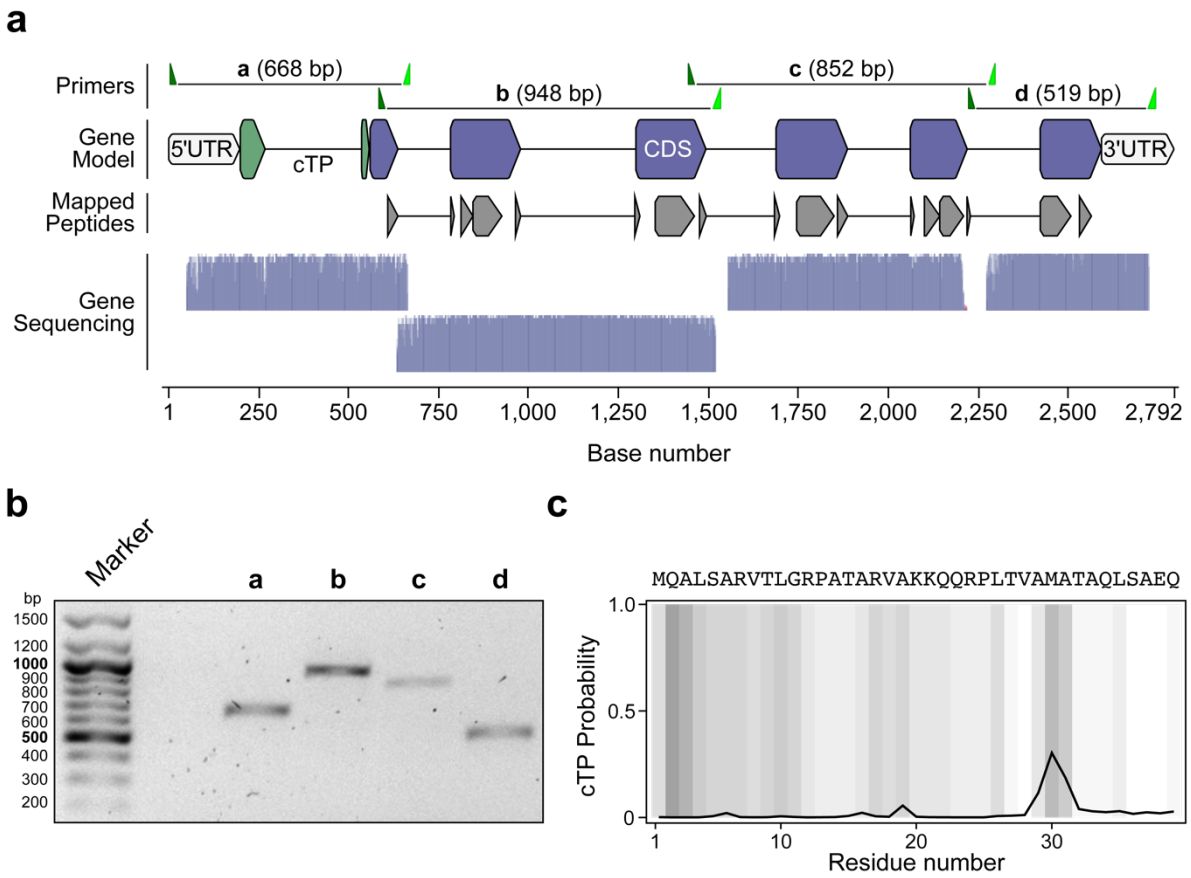


Figure 3-1 | Confirmation of CsLinker genomic and protein sequence.
(a) Schematic summary of the predicted gene model of CSI2_123000012064-RA. The primers used to amplify the sequences in Fig. 3-1b and their expected lengths (top) are shown relative to the gene model (below). The primer sequences are available in Table 3-14. In the gene model, exons are represented by the blocks, connected by a line representing the introns. The predicted chloroplast transit peptide (cTP) from Fig. 3-1c is shown in green. A selection of the mapped peptides from the later introduced absolute quantification mass spectrometry experiments are shown below the gene model where the mapped peptide sequence is represented by blocks aligned to the relative position of the corresponding amino acids in the gene sequence. Peptides that span introns are represented with a line between adjacent blocks. Sequencing results, represented as quality chromatograms, mapped to the gene are shown at the bottom of the figure. (b) DNA gel electrophoresis of PCR products corresponding to the mapped products in Fig. 3-1a. The same products were subsequently sequenced. (c) TargetP 2.0 prediction of the chloroplast transit peptide location in the first 39 residues. The predicted cleavage site is after the Met30 residue.

3.2.2. Recombinant production and purification of CsLinker

Based on literature reports of insolubility and proteolytic degradation of linker proteins during production and purification (Oh *et al.,* 2023; Wunder *et al.,* 2018), the mature coding sequence of CsLinker was fused to an N-terminal 6xHis-mEGFP in addition to a C-terminal MBP-6xHis tag. From previous attempts to purify the putative Ulva linker protein, not discussed here, the fusion of these tags to improved solubility and integrity of the fusion protein, allowing high quality purification (Fig. 3-2).

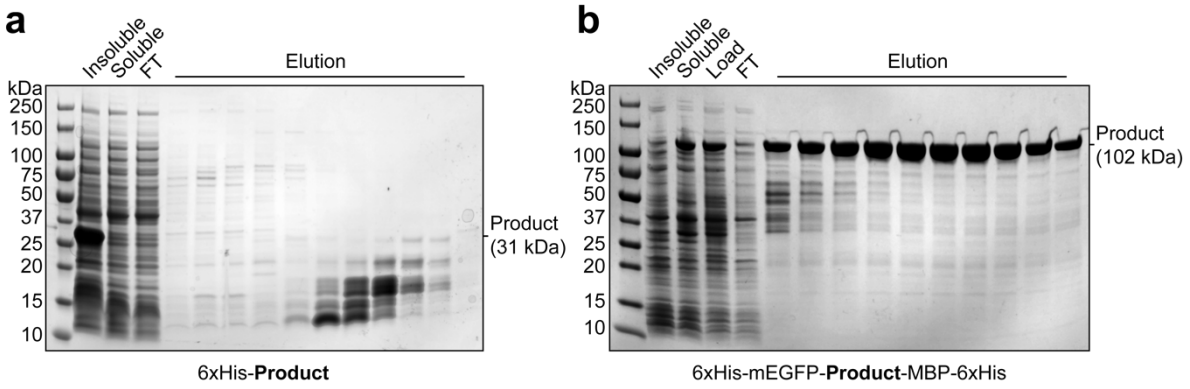


Figure 3-2 | Comparison of Ulva linker protein purification with and without protective fusion.
(a) SDS-PAGE analysis of purification of putative Ulva linker protein (not CsLinker) fused to a 6xHis tag. FT = flow through from the nickel affinity purification. (b) SDS-PAGE analysis of purification of the same Ulva linker protein fused to protective mEGFP and MBP tags. The expected product size in both instance is indicated.

The same high quality purification was achieved with the mEGFP-CsLinker-MBP fusion protein (Fig. 3-3a). Inclusion of Tobacco Etch Virus (TEV) cleavage sites between the fusion tags and the CsLinker permitted release of the mature untagged CsLinker protein from the fusion protein, which remained soluble and was purified by collecting the flowthrough of a second nickel affinity column purification (Fig. 3-3b). Notably, the cleaved CsLinker product was observed to run higher than the predicted molecular weight, relative to the cleaved MBP and mEGFP proteins, likely due to the high isoelectric point of the cleaved product (pI = 11.74). The purified protein was observed to be of good integrity with little contamination of other protein bands, when assessed by SDS-PAGE at a normalised load of ~30 µg (Fig. 3-3b, right). Importantly, due to the lack of tryptophans and consequent low extinction coefficient (11,920 M^{-1} cm^{-1}) and absorbance at 280 nm (0.33 AU) of the untagged CsLinker protein, the protein concentration could not be estimated by spectrophotometric approaches (*e.g.,* NanoDrop). The final concentration of the CsLinker was instead quantified by Bradford assay (Fig. 3-4a) and compared to a normalised 4 µg load of bovine serum albumin (BSA) (Fig. 3-4b). For reference, NanoDrop analysis of the same sample underestimated protein concentration 2.5-fold relative to the Bradford assay and normalised load.

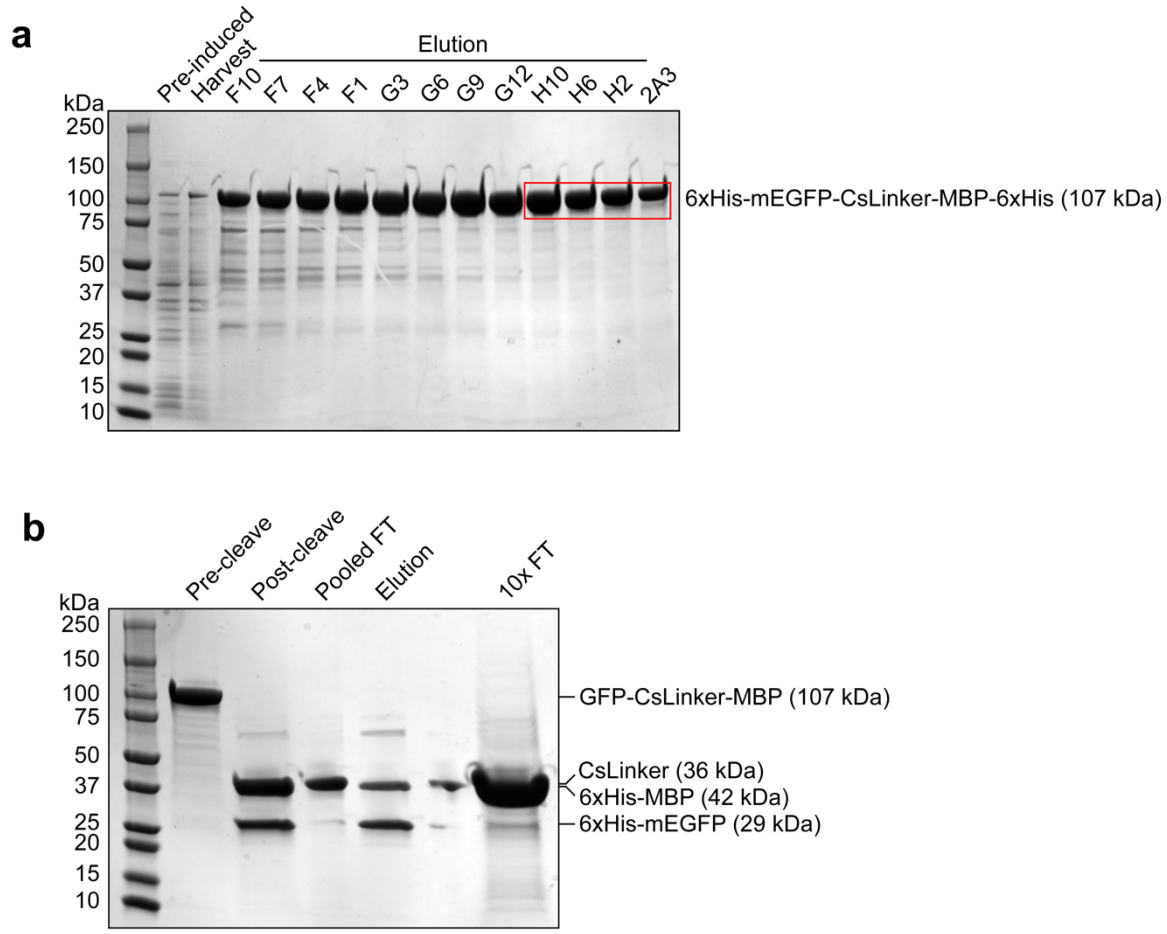


Figure 3-3 | Purification of untagged CsLinker.
(a) SDS-PAGE analysis of nickel affinity purification of full length CsLinker fused to mEGFP and MBP tags. The fractions pooled and taken for cleavage (H10-2A3) are indicated by the red box. (b) SDS-PAGE analysis of TEV cleavage products and nickel affinity column flowthrough purification of the untagged CsLinker. 10x FT indicates the normalised load was 10x that of the 'Pooled FT' lane (~30 µg), and was used to assess the purity of the sample. Due to high isoelectric point of CsLinker, the purified protein runs high on an SDS-PAGE gel and distinction between CsLinker and the MBP band is obscured in some cases.

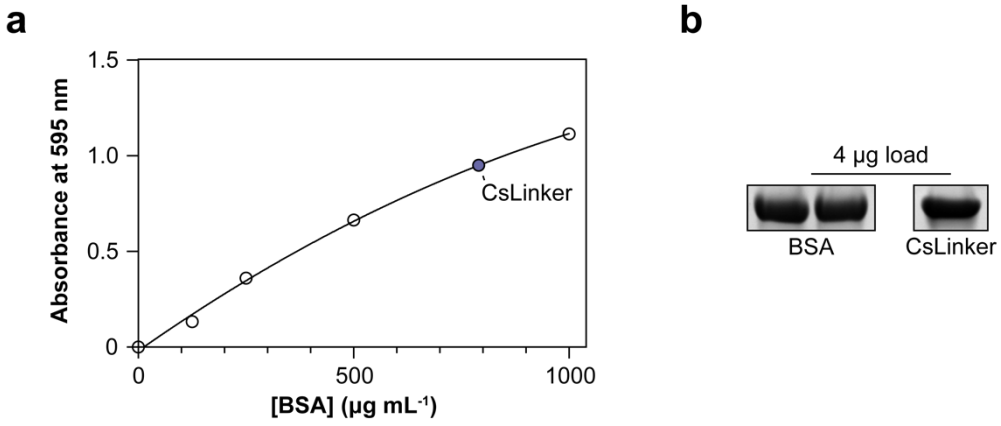


Figure 3-4 | Estimation of untagged CsLinker concentration.
(a) Bradford assay standard curve of BSA standards (open circles). A quadratic fit was used to derive the concentration of the CsLinker based on the absorbance value (filled circle). (b) Qualitative comparison of 4 µg load of BSA (weighed pure protein) and CsLinker by SDS-PAGE following quantification by Bradford assay.

3.2.3. *Purification of Chlorella Rubisco*

Although significant advancements have been made in the recombinant production and purification of Rubisco from *E. coli* (Aigner *et al.,* 2017; Buck *et al.,* 2023), the requirement for species-specific assembly factors and chaperones means purification from native material is often preferred. To this end, a Rubisco purification protocol based on that of Shivhare & Mueller-Cajar (2017) was employed (see methods). In addition to a modified ammonium sulphate precipitation which removed chlorophyll-bound proteins of the LHCII complex (data not shown – confirmed by mass spectrometry), sucrose density gradient centrifugation was introduced to remove most other soluble proteins (Fig. 3-5a). Subsequent anion exchange and size exclusion chromatography approaches polished the Rubisco sample to a high level of purity (Fig. 3-5b,c). The post-injection elution volume of Rubisco from a HiLoad superdex 16/600 200 pg column was as expected for the calculated size of the Rubisco holoenzyme (549 kDa) (Fig. 3-5c) and demonstrated the integrity of the hexadecameric complex of 8 Rubisco large subunits (RbcLs) and 8 Rubisco small subunits (RbcSs).

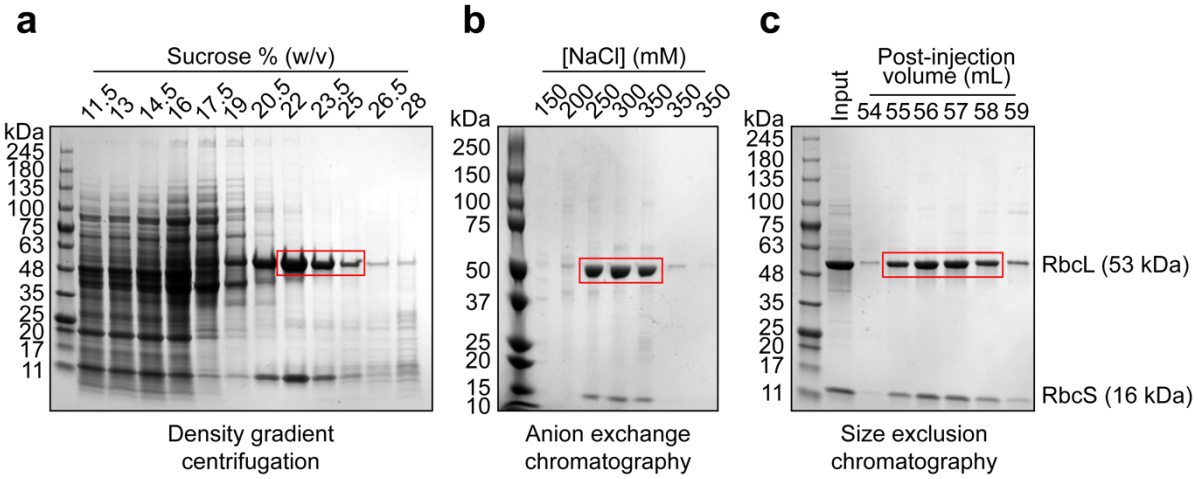


Figure 3-5 | Purification of Chlorella Rubisco.
(a) SDS-PAGE analysis of fractions from a representative sucrose density gradient following centrifugation and fractionation. The calibrated sucrose concentration based on the fractionation volume in each fraction is indicated. The red box indicates the fractions pooled for anion exchange chromatography (22-25% sucrose) (b) SDS-PAGE analysis of elution fractions from an anion exchange chromatography purification of the pooled sample from Fig. 3-5a. The approximate salt concentration based on the conductivity measurement in the elution fractions is indicated. The red box indicates the fractions that were pooled for further purification (c) SDS-PAGE analysis of elution fractions from a size exclusion chromatography purification of the pooled sample from Fig. 3-5b. The volume of the elution fraction relative to the injection is indicated. The red box indicates the pooled final fractions that were used in subsequent experiments.

3.2.4. Determining the in vivo abundance of CsLinker and Rubisco in vivo

THE WORK IN THIS SECTION WAS COMPLETED IN COLLABORATION WITH MIHRIS NADUTHODI (UNIVERSITY OF YORK) AND IS PRESENTED HERE IN THE INTERESTS OF COMPLETENESS.

To determine the *in vivo* abundance of CsLinker and Chlorella Rubisco, absolute quantification mass spectrometry using the purified proteins as concentration standards was pursued. In this methodology, the unique peptide fragments identified from the second mass spectrometer (MS2) in liquid chromatography tandem mass spectrometry (LC-MS/MS) are mapped backed to the corresponding peptide peaks from the first mass spectrometer (MS1) and the integrated, summed intensity of the MS1 peptide peaks is used to construct a calibration curve based on the indicated concentration of the purified protein standards. For the CsLinker protein standards, 14 unique peptides were identified and there was a clear linear relationship between input concentration and MS1 intensity for the 3-point calibration curve across the 2 replicates (Fig. 3-6a). For Chlorella Rubisco, only 2 unique peptides were identified for the Rubisco small subunit (RbcS), whereas 16 were identified for the Chlorella Rubisco large subunit (CsRbcL), in line with the size differences of the proteins. The RbcS was excluded from further analysis as the small number of unique peptides obscured accurate quantification of the Rubisco holoenzyme concentration due to non-linearity of the standard curve. The 3-point calibration curve for the CsRbcL demonstrated good linearity (Fig. 3-6b). The same peak integration methodology was applied to whole cell Chlorella samples in triplicate, where two technical replicates were used to determine the MS1 intensity in each replicate. For these experiments Chlorella cells were grown at ambient CO_2 under photoautotrophic conditions (under which a pyrenoid-based CCM is presumably required). Very little variance was observed between replicates, which allowed determination of the absolute concentration of the proteins in the whole cell samples (Fig. 3-6a,b). Although the data demonstrated a good linear relationship, the third calibration point would ideally be placed closer that of the lower concentrations in future experiments. Densitometric western blot confirmation of the approximate *in vivo* concentrations could also be completed relative to the protein standards to provide an independent validation of these measurements.

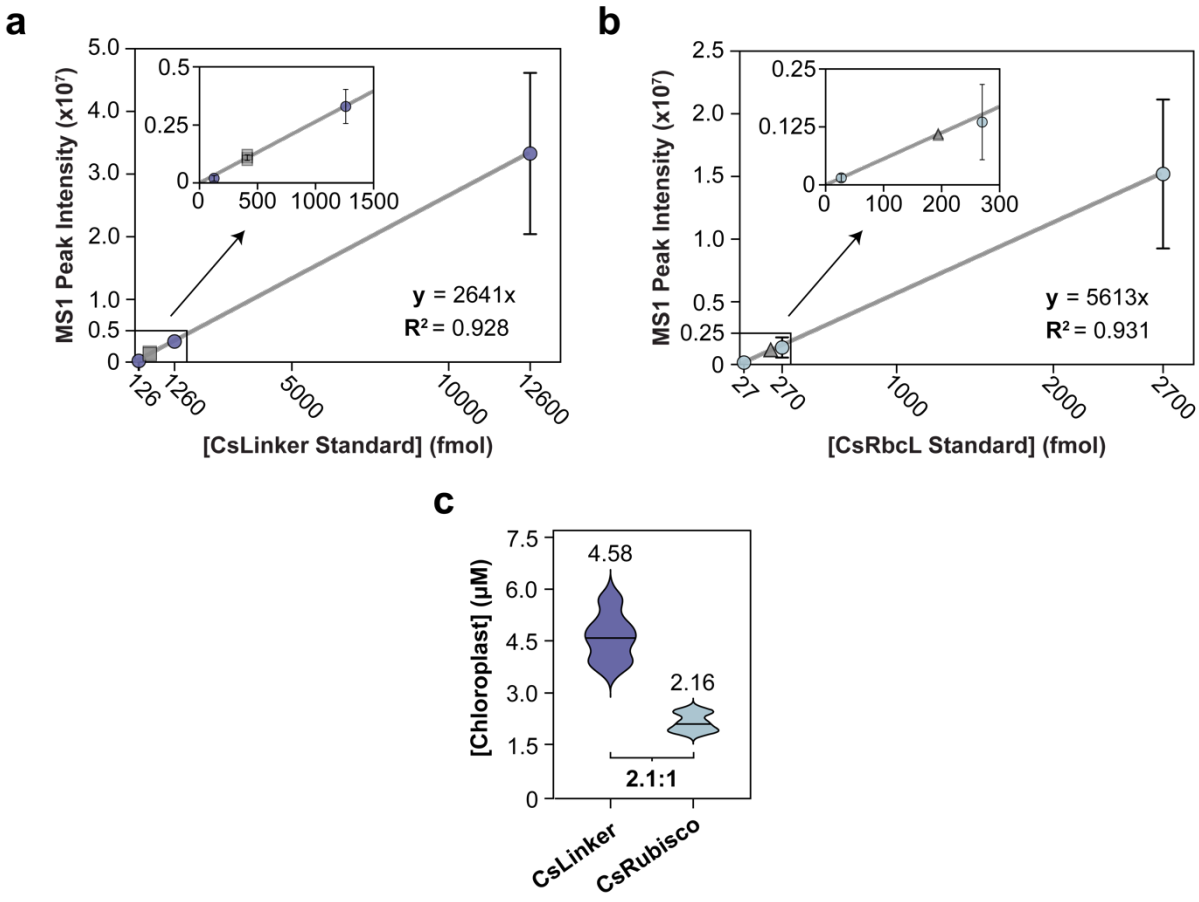


Figure 3-6 | Absolute quantification mass spectrometry determination of *in vivo* CsLinker and Rubisco concentrations.
(a) Linear calibration curve of CsLinker protein standards (purple circles). The whole cell proteomic values are represented by grey squares. Error bars where visible represent S.D. (b) Linear calibration curve of Chlorella RbcL (CsRbcL) protein standards (blue circles). The whole cell proteomic values are represented by grey triangles. Error bars where visible represent S.D. (c) Violin plot of calculated chloroplast concentration of CsLinker and Chlorella Rubisco (CsRubisco). The mean value is represented by the line with each lobe representing the average of 2 technical replicates (n = 3 biological replicates). The ratio between the two proteins is indicated.

The chloroplast volume of *Auxenochlorella pyrenoidosa* (~15 µm^3) has been measured by focused ion beam scanning electron microscopy (FIB-SEM) segmentation previously (Guo *et al.,* 2022). Based on the historic mis-classification of *Auxenochlorella pyrenoidosa* species and the similarity in genome size of the reported strain (FACHB-9) and that of *Chlorella sorokiniana* (56.8 and 58.5 Mb respectively) (Fan, Ning, *et al.,* 2015), relative to that of *bona fide Auxenochlorella* species (22.9 Mb; (Gao *et al.,* 2014)), it is likely this species was also mis-characterised and in fact represents a close relative of *Chlorella sorokiniana*, as present in historic literature. Using this reported chloroplast volume, the *in vivo* concentration of CsLinker and Rubisco in the chloroplast was calculated (Tables 3-2 , 3-3 and Fig. 3-6c). The same calculation was completed using the published values of absolute EPYC1 and RbcL *in*

vivo abundance (Hammel *et al.,* 2018) and chloroplast volume (Gaffal *et al.,* 1995; Weiss *et al.,* 2000) in Chlamydomonas (Tables 3-4, 3-5). The calculated concentration of Rubisco and the respective linker proteins was comparable between Chlamydomonas and Chlorella, though Chlorella did demonstrate ~7-fold lower chloroplast Rubisco concentration than Chlamydomonas. In contrast to Chlamydomonas, where EPYC1 and Rubisco are roughly equimolar (Tables 3-4, 3-5), CsLinker was found to be in ~2-fold excess over Chlorella Rubisco (Fig. 3-6c), with absolute concentrations of 4.58 ± 0.53 S.D. and 2.16 ± 0.04 S.D. µM respectively. Taken together these results suggest CsLinker is an abundant chloroplast component, as observed previously for EPYC1 in Chlamydomonas, but that perhaps some global concentration differences of the components in Chlorella are present.

Table 3-2: Calculation of CsLinker concentration in the chloroplast.

MS1 intensity replicate 1	1224976	fmols (/2641)	463.8
MS1 intensity replicate 2	1064847	fmols (/2641)	403.2
MS1 intensity replicate 3	975129.7	fmols (/2641)	369.2
		average fmols	412.1
		average mols	4.12E-13
		mol / cell (/6E+6)	6.87E-20
		[chloroplast] (M) (/1.5E-14)	4.58E-6
		[CsLinker chloroplast] (µM)	**4.58** ± 0.53 S.D.
		molecules / cell (av. mol*NA)	4.13E+04

Table 3-3: Calculation of Chlorella Rubisco (CsRubisco) concentration in the chloroplast.

MS1 intensity replicate 1	1110011.8	fmols (/5613)	197.8
MS1 intensity replicate 2	1095989.3	fmols (/5613)	195.3
MS1 intensity replicate 3	1067985	fmols (/5613)	190.3
		average fmols	194.4
		average mols	1.94E-13
		mol / cell (/6E+6)	3.24E-20
		[chloroplast] (M) (/1.5E-14)	2.16E-6
		[CsRubisco chloroplast] (µM)	**2.16** ± 0.04 S.D.
		molecules / cell (av. mol*NA)	1.95E+04

Table 3-4: Calculation of EPYC1 concentration in the Chlamydomonas chloroplast.

mol / cell (EPYC1)	2E-18
[chloroplast] (M) (/1.33E-13)	1.50E-5
[EPYC1 chloroplast] (µM)	**15.0**

Table 3-5: Calculation of Chlamydomonas Rubisco (CrRubisco) concentration in the Chlamydomonas chloroplast.

mol / cell (CsRubisco) (16.1E-8/2)	2.05E-18
[chloroplast] (M) (/1.33E-13)	1.54E-5
[CrRubisco chloroplast] (µM)	**15.4**

The absolute quantification measurements of the whole cell samples were completed on cells grown under ambient CO_2 conditions (0.04% CO_2). In Chlamydomonas the abundance of EPYC1 increases when cells are switched from CO_2-replete (>1% CO_2) to ambient conditions (Mackinder *et al.,* 2016; Miura *et al.,* 2004), coincident with the accompanying biogenesis of the pyrenoid (Borkhsenious *et al.,* 1998; Mackinder *et al.,* 2016) and induction of the CCM (Badger *et al.,* 1980; Manuel & Moroney, 1988). It was hypothesised the same response exists in Chlorella species, despite mixed reports of pyrenoid presence under CO_2-replete (del Pino Plumed *et al.,* 1996; Gergis, 1972; Gorelova *et al.,* 2019; Miyachi *et al.,* 1986; Villarejo *et al.,* 1997) and dark (Budd *et al.,* 1969) growth conditions. To assess the relative abundance of CsLinker under ambient and replete CO_2 conditions *in vivo*, an antibody that was raised to a unique region of the protein ('PTPVSNSGVRSAMSSG') was used. An antibody raised to a C-terminal region of the Rubisco large subunit ('EVWKEIKFEFETIDTL'), which is common to most green lineage Rubiscos was used for coincident quantification of Chlorella Rubisco relative abundance. Western blot analysis indicated that CsLinker abundance was ~9-fold higher in cells grown under ambient CO_2 relative to CO_2-replete (3%) conditions (Fig. 3-7a); a response ~5-fold stronger than that observed for EPYC1 which has 2-fold greater abundance under ambient conditions (Mackinder *et al.,* 2016). Contrastingly, the abundance of CsRbcL, and presumably that of the Chlorella Rubisco holoenzyme, was largely consistent (0.8-fold ambient to replete abundance change; Fig. 3-7b), in line with reports in Chlamydomonas (Mackinder *et al.,* 2016). The strong CO_2 response of CsLinker appears in contrast to the literature reports of pyrenoid presence under CO_2-replete conditions but consistent

with the expected low CO_2 induction of the CCM (Fan, Xu, *et al.,* 2015; Villarejo *et al.,* 1997, 1998). Further investigation into pyrenoid presence under different CO_2 conditions should certainly be further investigated in Chlorella. Taken together the results presented in this section confirm that CsLinker is an abundant cellular component with a calculated chloroplast concentration that would permit phase separation of Rubisco if an appropriate interaction between the two proteins exists under ambient CO_2 conditions.

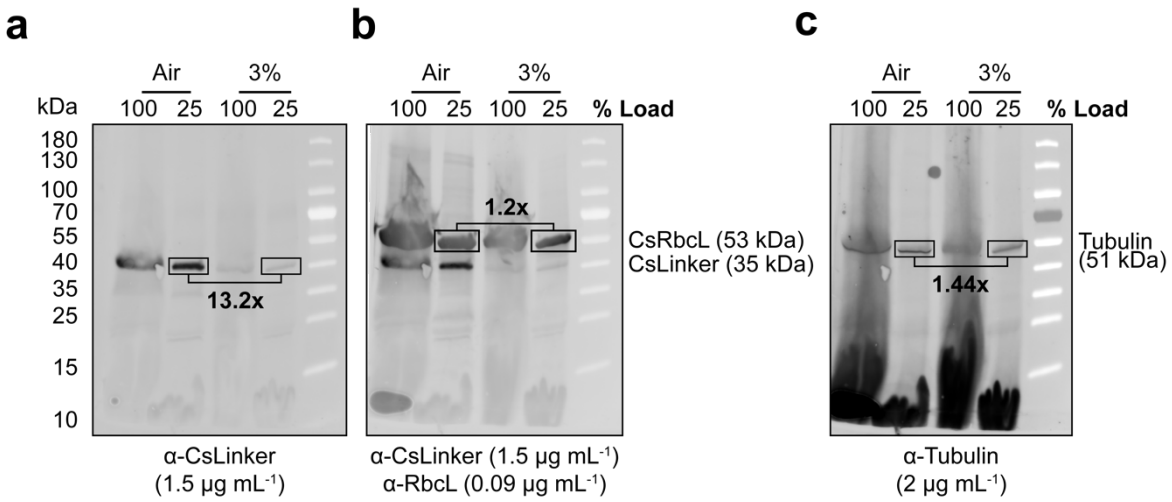


Figure 3-7 | Relative abundance of CsLinker in response to CO_2 availability.
(a) Western blot analysis of whole cell lysates of cells grown under the indicated CO_2 conditions using the CsLinker antibody at the indicated concentration. The bands used for densitometric relative quantification, and the relative intensity difference between them, are indicated. '% Load' indicates the relative loading of the same sample with 100% representing 40 µL of lysate (see methods). (b) The same membrane from Fig. 3-6a was subsequently probed with the RbcL antibody and the same analysis completed as indicated. (c) The same membrane separately probed with a Tubulin antibody that was detected with a different secondary antibody as in Fig 3-6a and Fig. 3-6b. The loading error correction applied to the abundance difference in a and b is indicated.

3.2.5. *Characterising ex vivo interaction between CsLinker and Rubisco by co-immunoprecipitation coupled mass spectrometry (co-IP)*

THE WORK IN THIS SECTION WAS COMPLETED IN COLLABORATION WITH MIHRIS NADUTHODI (UNIVERSITY OF YORK) AND IS PRESENTED HERE IN THE INTERESTS OF COMPLETENESS.

Despite widespread reports of nuclear transformation of Chlorella species by electroporation, biolistic bombardment, and peptide co-transformation, none of the published methodologies were replicable using a wide range of constructs and parameters (Table 3-6) (see methods for further details). There are no reports of immunofluorescence localisation in Chlorella species, and accordingly extensive attempts to localise Chlorella Rubisco and CsLinker were also unsuccessful (Table 3-7), suggesting poor permeability of the cells in line with the lack of successful transformation. Although immunogold localisation of Rubisco in the pyrenoid using the RbcL antibody was successful and consistent with previous reports (Villarejo *et al.,* 1998), no labelling was observed with the CsLinker antibody (Fig. 3-8). Clearly, these difficulties obfuscated the *in vivo* validation of the CsLinker-Rubisco interaction or co-localisation and instead *ex vivo* approaches were pursued.

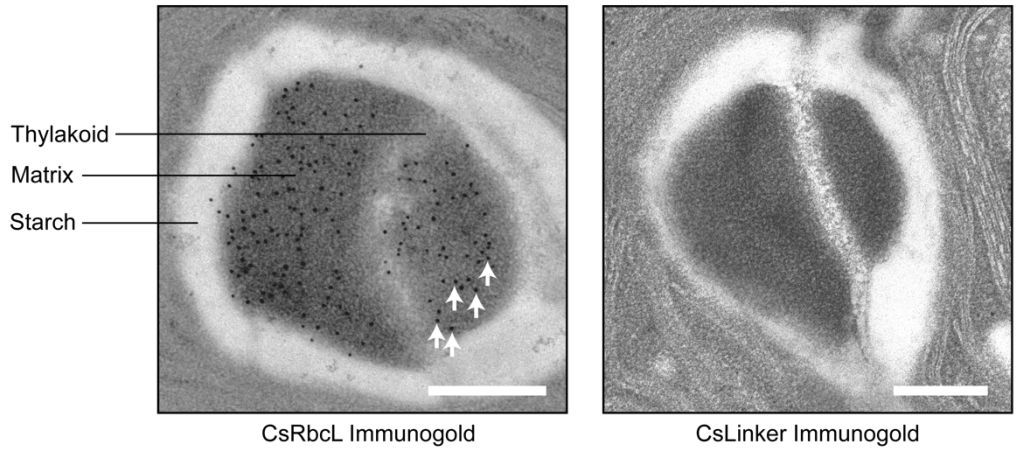


Figure 3-8 | Immunoelectron microscopy of Chlorella pyrenoid components.
TEM images of Chlorella pyrenoids incubated with RbcL (left) or CsLinker (right) primary antibodies. Examples of the immunogold particles are indicated by the white arrows. Scale bar = 200 nm.

Table 3-6: Summary of protocols used for attempted transformation of Chlorella.

No.	Construct	Method	Parameters
1	pLM017 (5 µg, linear)	Electroporation (Gene Pulser II)	$5x10^7$ cells, 800V, 25 µF
2	pLM845 (5 µg, linear)	Electroporation (Gene Pulser II)	$5x10^7$ cells, 800V, 25 µF
3	pLM847 (5 µg, linear)	Electroporation (Gene Pulser II)	$5x10^7$ cells, 800V, 25 µF
4	pLM849 (5 µg, linear)	Electroporation (Gene Pulser II)	$5x10^7$ cells, 800V, 25 µF
5	pLM630 (5 µg, linear)	Electroporation (Gene Pulser II)	$5x10^7$ cells, 800V, 25 µF
6	pCAMBIA1301 (3 µg, circular)	Electroporation (Gene Pulser II)	$1x10^8$ cells, 500, 750, 1000, 1500, 2000V, 25 µF, 30 µg salmon sperm
7	pCAMBIA1301 (3 µg, circular)	Electroporation (Gene Pulser II)	$1x10^6$ cells, 500, 750, 1000, 1500, 2000V, 25 µF, 30 µg salmon sperm
8	pCAMBIA1301 (1 µg, linear)	Electroporation (Gene Pulser II)	$1x10^8$ cells, 1000, 1500, 2000V, 25 µF, 10 µg salmon sperm
9	pCAMBIA1301 (1 µg, linear)	Electroporation (Gene Pulser II)	$1x10^6$ cells, 1000, 1500, 2000V, 25 µF, 10 µg salmon sperm
10	pCAMBIA1301 (1.5 µg, circular)	Electroporation (NEPA21)	$4x10^6$ cells, 200-300V, 2-5 poring pulses of 2-9 ms, 5 20V transfer pulses of 50 ms
10	pCAMBIA1301 (1.5 µg, circular)	Electroporation (NEPA21)	$2x10^8$ cells, 200-300V, 2-5 poring pulses of 2-9 ms, 5 20V transfer pulses of 50 ms
11	pCAMBIA1301 (1.5 µg, circular)	Electroporation (NEPA21)	$2x10^8$ cells, 200-300V, 2-5 poring pulses of 2-9 ms, 5 20V transfer pulses of 50 ms
12	pCAMBIA1301 (2.5 µg, circular)	Electroporation (NEPA21)	$2.5x10^8$ cells, 200-300V, 2-5 poring pulses of 2-9 ms, 5 20V transfer pulses of 50 ms
13	pCAMBIA1301 (4 µg, circular)	Electroporation (NEPA21)	$2.25x10^8$ cells, 200-300V, 2-5 poring pulses of 2-9 ms, 5 20V transfer pulses of 50 ms
14	pCAMBIA1301 (2.5 µg, circular)	Biolistic (PDS-1000/He)	$1x10^7$ cells, 0.5 mg gold
15	pCAMBIA1301 (4.5 µg, linear)	Biolistic (PDS-1000/He)	$1x10^7$ cells, 0.5 mg gold
16	pLM021 (2.5 µg, circular)	Biolistic (PDS-1000/He)	$1x10^7$ cells, 0.5 mg gold
17	pLM031 (2.5 µg, circular)	Biolistic (PDS-1000/He)	$1x10^7$ cells, 0.5 mg gold
18	pLM631 (2.5 µg, circular)	Biolistic (PDS-1000/He)	$1x10^7$ cells, 0.5 mg gold
19	pLM847 (2.5 µg, circular)	Biolistic (PDS-1000/He)	$1x10^7$ cells, 0.5 mg gold
20	pLM848 (2.5 µg, circular)	Biolistic (PDS-1000/He)	$1x10^7$ cells, 0.5 mg gold
21	pSK397 (2.5 µg, circular)	Electroporation (Pulser XCell)	$1x10^6$ protoplasts, 655V, 3.4 ms
22	pCAMBIA1301 (1 µg, linear)	Peptide (HIV-TAT)	$2x10^5$ protoplasts, 4 µg HIV-TAT peptide, 48 hours incubation

Table 3-7: Summary of attempted immunofluorescence protocols.

No.	Fixation	Permeabilization
1	3.7% PFA in PBS (30 minutes at RT)	0.5% T-X (15 minutes at RT)
2	3.7% PFA in PBS (30 minutes at RT)	0.05% DMSO (10 minutes at RT)
3	3.7% PFA in PBS (30 minutes at RT)	1% DMSO (10 minutes at RT)
4	3.7% PFA in PBS (30 minutes at RT)	100% MeOH (10 minutes at RT)
5	3.7% PFA in PBS (30 minutes at RT)	100% acetone (1 minute at RT)
6	3.7% PFA in PBS (30 minutes at RT)	0.01% SDS (10 minutes at RT)
7	3.7% PFA in PBS (30 minutes at RT)	Electroporation (NEPA21; (Yamano & Fukuzawa, 2020)
8	3.7% PFA in PBS (30 minutes at RT)	Electroporation (Gene Pulser II;(Emrich-Mills *et al.,* 2021)
9	3.7% PFA in PBS (30 minutes at RT)	Electroporation (Gene Pulser II; 2500V cm^{-1}, 50 μF
10	96% EtOH used for 1-step fixation and permeabilization (30 minutes on ice)	
11	100% MeOH used for 1-step fixation and permeabilization (30 minutes on ice)	
Control	3.7% PFA in PBS (30 minutes at RT)	0.5% T-X (15 minutes at RT)

To determine if the native CsLinker and Chlorella Rubisco proteins interact *ex vivo*, a co-immunoprecipitation mass spectrometry (co-IP) approach was applied. Similar approaches have previously demonstrated the interaction of Chlamydomonas and Phaeodactylum linker proteins with Rubisco by using Rubisco as a bait protein in the experiments. (Mackinder *et al.,* 2017; Oh *et al.,* 2023). The reciprocal experiment in Chlamydomonas, in which EPYC1 was the bait protein was also completed and demonstrated a high confidence interaction between the two proteins (Mackinder *et al.,* 2017). Accordingly, reciprocal co-IP experiments using both Rubisco and CsLinker as prey were completed by loading protein A resin with the primary antibodies raised to the RbcL and CsLinker proteins respectively, prior to incubation with whole cell lysate of cells grown under ambient CO_2 conditions where CsLinker abundance was highest (Fig. 3-7). Control co-IPs in which the protein A resin was loaded with a primary antibody raised to Chlamydomonas Bestrophin-like protein 2 (Mukherjee *et al.,* 2019), which was assumed to have no cross-reactivity in Chlorella, were used as a comparison for relative enrichment. The RbcL co-IP demonstrated strong enrichment of both the CsRbcL and CsRbcS subunits and indicated the intended loading of the Chlorella Rubisco holoenzyme onto the protein A beads (Fig. 3-9a). The top 5 enriched proteins in the CsRbcL co-IP were the previously introduced pyrenoid-matrix starch tether (PMST), RbcX, CsRbcS, CsLinker and CsRbcL

respectively, which all also had exceptionally low adjusted *P* values, indicating *bona fide ex vivo* interaction between the proteins and the Chlorella Rubisco holoenzyme (Fig. 3-9a). The previously introduced pyrenoid matrix thylakoid tether (PMTT) protein was also significantly enriched to a lesser extent, possibly due to residual integration in membranous material as no detergent was used during lysis of the cells. In the reciprocal CsLinker co-IP experiment, CsLinker was the most significantly enriched protein and co-enriched both CsRbcL and CsRbcS (Fig. 3-9b). As expected, given the lack of predicted Rubisco-like interfaces on the proteins, PMST, RbcX and PMTT were not enriched in the CsLinker co-IP, consistent with the specificity of their interaction with Rubisco. Taken together the co-IP experiments indicated a high confidence *ex vivo* interaction between CsLinker and the Chlorella Rubisco holoenzyme and affirmed the likelihood of CsLinker as a *bona fide* linker protein *in vivo*, should the interaction give rise to phase separation between the proteins.

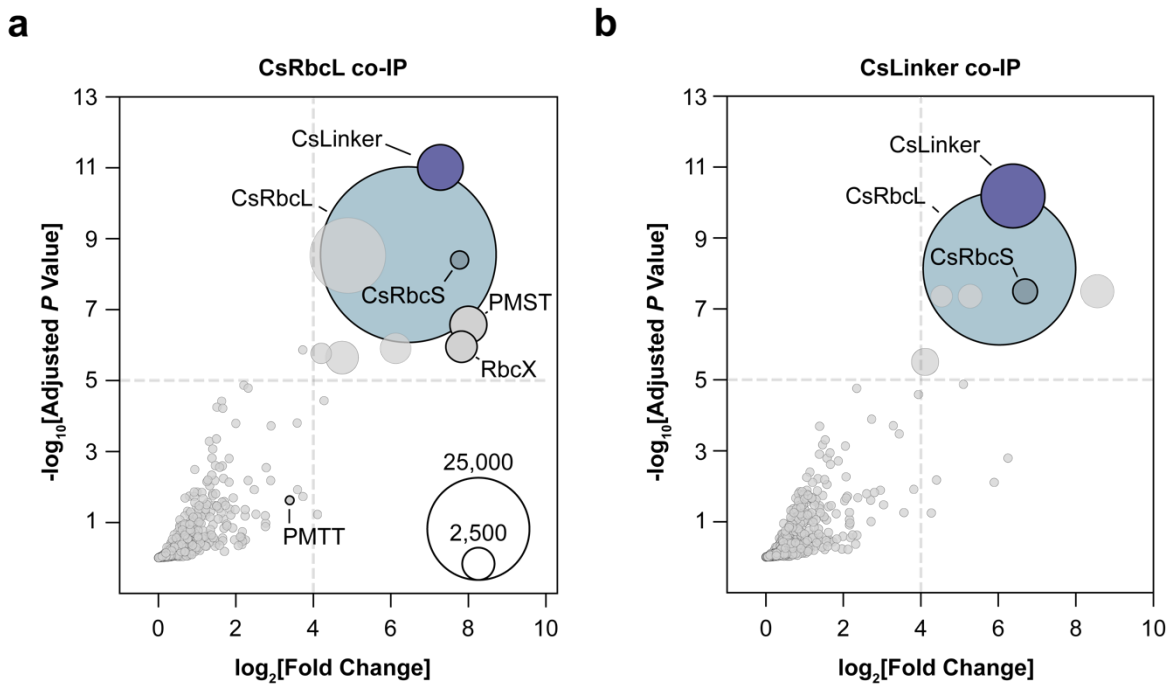


Figure 3-9 | Co-immunoprecipitation mass spectrometry (co-IP) of CsLinker and Rubisco.
(a) Scatter plot of co-IP results from experiments completed with the RbcL antibody against Chlorella RbcL (CsRbcL). Only values with a positive enrichment and adjusted *P* value < 0.05 are shown. Dashed lines indicate arbitrary significance thresholds ($-\log_{10}$[Adjusted *P* Value] > 5, $\log_2$[Fold Change] > 4), above which points are sized according to summed MS1 peak intensity from three replicates as per the inset key. (b) Scatter plot of co-IP results from experiments completed with the CsLinker antibody. Only values with a positive enrichment and adjusted *P* value < 0.05 are shown. The same significance thresholds as in a were used, above which the sizing was completed in the same way and is consistent with the key in Fig. 3-9a.

3.3. Characterising the in vitro interaction of CsLinker with Chlorella Rubisco

Having presented *in vivo* and *ex vivo* evidence that is consistent with the role of CsLinker as a pyrenoid linker protein in Chlorella, it was next required to demonstrate the interaction between the protein gives rise to the hypothesised emergent properties of pyrenoid matrices (Barrett *et al.,* 2021); namely liquid-liquid phase separation (LLPS). For Chlamydomonas and Phaeodactylum, *in vitro* reconstitutions have plainly demonstrated this fact for the respective linker proteins, with the reconstitutions presenting analogous properties to the *in vivo* condensates (Freeman Rosenzweig *et al.,* 2017; Oh *et al.,* 2023; Wunder *et al.,* 2018). Accordingly, the aim of this section was to reconstitute the Chlorella pyrenoid matrix *in vitro*, which was hypothesised to be underpinned by the interaction of CsLinker and Chlorella Rubisco, and study the properties of the resulting reconstitution as a proxy for the properties of the pyrenoid matrix *in vivo*.

3.3.1. Native PAGE validation of CsLinker-Rubisco interaction in vitro

It was initially appropriate to determine the *ex vivo* interaction demonstrated between native CsLinker and Rubisco (Fig. 3-9) was replicable with the purified proteins (Figs. 3-3 and 3-5). To this end, native polyacrylamide gel electrophoresis (native PAGE), which allows the separation of protein complexes based on size and charge due to the non-denaturing conditions under which the experiment is completed and has been used to demonstrate a range of Rubisco-linker interactions (Oh *et al.,* 2023; Oltrogge *et al.,* 2020; Wunder *et al.,* 2018), was employed. In the experiments Rubisco incubated alone was observed to migrate as a single band above the 245 kDa marker, reinforcing the purity and integrity of the purified Rubisco holoenzyme (Fig. 3-10a, left). When purified CsLinker was incubated with Rubisco, the monomeric Rubisco band demonstrated a concentration-dependent reduction in intensity (Fig. 3-10a, right). Due to the high isoelectric point (11.74) of untagged CsLinker and its consequent net positive charge in the native PAGE running buffer (pH = ~8.3), CsLinker does not enter the gel in the unbound state. Any changes in the intensity or migration of the monomeric Rubisco band are presumably therefore due to some interaction between CsLinker and Rubisco that either permits entry of the complex into the gel and retards the migration of Rubisco through size and charge effects or prevents entry of Rubisco into the gel and reduces intensity of the observed band. This hypothesised

interaction is termed the 'bound' Rubisco state. Densitometric quantification of monomeric Rubisco at each condition indicated an apparent $K_{0.5}$ for 'bound' Rubisco of ~0.8 μM and affirmed interaction between the proteins *in vitro* (Fig. 3-10b).

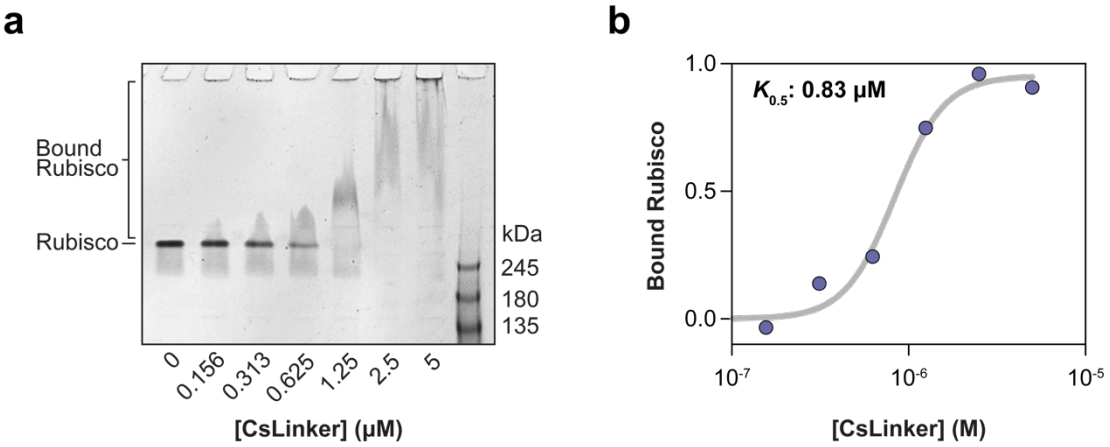


Figure 3-10 | Characterisation of *in vitro* CsLinker-CsRubisco interaction by native PAGE.
(a) Native PAGE analysis of 1 μM Chlorella Rubisco incubated with increasing indicated concentrations of CsLinker prior to separation. The intensity of the monomeric Rubisco band was used for quantification of data in Fig. 3-10b. (b) Quantification of 'bound' Rubisco derived from the inverse relative intensity of the monomeric Rubisco band in Fig. 3-10a. A hill model of binding was fitted to the data to derive the apparent $K_{0.5}$: ($Y = (Y_{max} * X^h)/(K_{0.5}{}^h + X^h)$, where h is the hill coefficient).

3.3.2. Characterising droplet formation and composition with CsLinker-CsRubisco solutions

Under the conditions at which all Rubisco was in the 'bound' state (Fig. 3-10), the solution was observed to turn cloudy seconds after mixing, relative to the control experiments in which no CsLinker was added (Fig. 3-11a). CsLinker incubated alone in solution also was not cloudy. Brightfield microscopy indicated the formation of spherical micron-scale droplets in solution when both proteins were present, in contrast to solutions of the proteins incubated alone where no droplets were observed (Fig. 3-11a). Subsequent droplet sedimentation assays completed on the centrifuged solutions indicated both Rubisco and CsLinker were present in the pelleted (P) phase of the droplet-containing solution, consistent with their formation requiring both components. Accordingly, very little Rubisco and CsLinker were present in the pelleted phase of the solutions in which they were incubated alone (Fig. 3-11b). These results were consistent with the previous observation that both EPYC1 and Chlamydomonas Rubisco were required to form droplets in the reconstitution of the Chlamydomonas pyrenoid (Wunder *et al.,* 2018). The results also suggest CsLinker does not homotypically phase separate under the quasi-physiological

conditions (pH 8.0, 50 mM NaCl, 2 µM CsLinker) used for the experiment, in contrast to PYCO1 of Phaeodactylum which homotypically phase separates into droplets under similar conditions (Oh *et al.,* 2023).

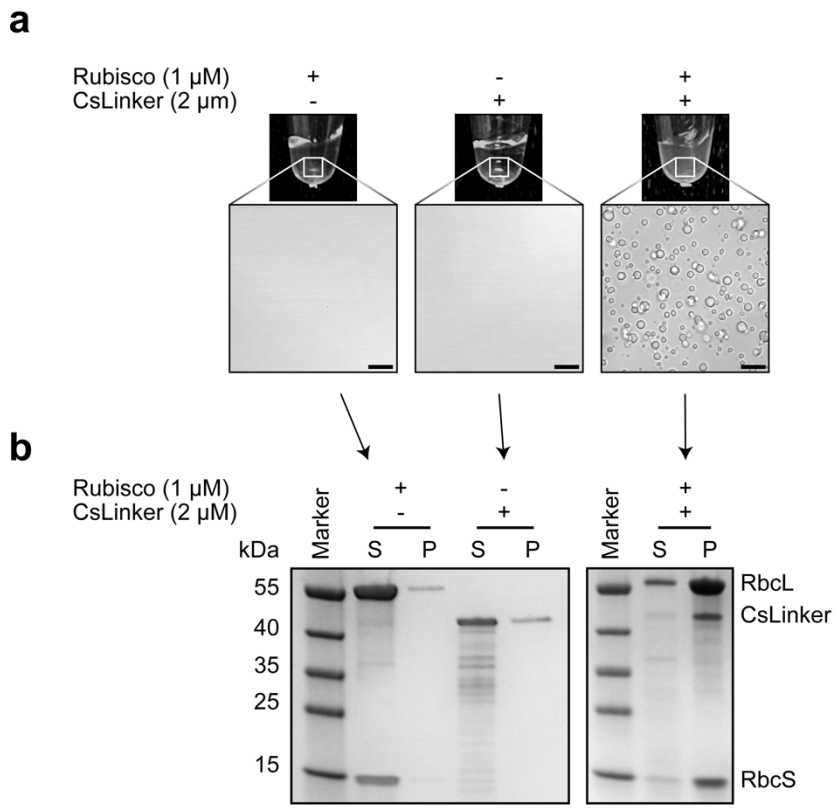


Figure 3-11 | Characterisation of *in vitro* droplet formation in solutions of CsLinker and CsRubisco. (a) Brightfield microscopy (bottom) of the solution (top) at the indicated concentrations of Rubisco (left) and CsLinker (middle) alone and together (right). Scale bar = 5 µm. (b) SDS-PAGE analysis of droplet sedimentation assays showing the composition of the supernatant (S) and pelleted (P) phases in each of the solutions formed under the same conditions as in Fig. 3-11a.

3.3.3. Quantification of droplet formation between CsLinker and CsRubisco

In Chlamydomonas, EPYC1 and Rubisco are equimolar in the chloroplast (Tables 3-4, 3-5). Titration droplet sedimentation assays with EPYC1 fixed at half the absolute *in vivo* concentrations demonstrated a two-fold excess of EPYC1 completely demixed both components (Wunder *et al.,* 2018). The resulting droplets had dynamic properties analogous to those of the *in vivo* condensates (Freeman Rosenzweig *et al.,* 2017), suggesting the reconstituted droplets in this concentration range were a good proxy for the *in vivo* condensate. To this end, the same analysis was completed with CsLinker and Chlorella Rubisco, which were both separately fixed at half the determined absolute *in vivo* concentration (2 µM and 1 µM

respectively) (Fig. 3-6c), and the other component was titrated. In both titrations, a 2-fold molar excess of CsLinker over CsRubisco completely demixed both components (Fig. 3-12), in line with ~2-fold abundance *in vivo* and consistent with the saturation concentration of 'bound' Rubisco assessed by native PAGE (Fig. 3-10). Taken together this evidence suggested droplets formed under a 2-fold excess of CsLinker are likely representative of the *in vivo* condensates, and as such most future characterisation was completed at this molar ratio.

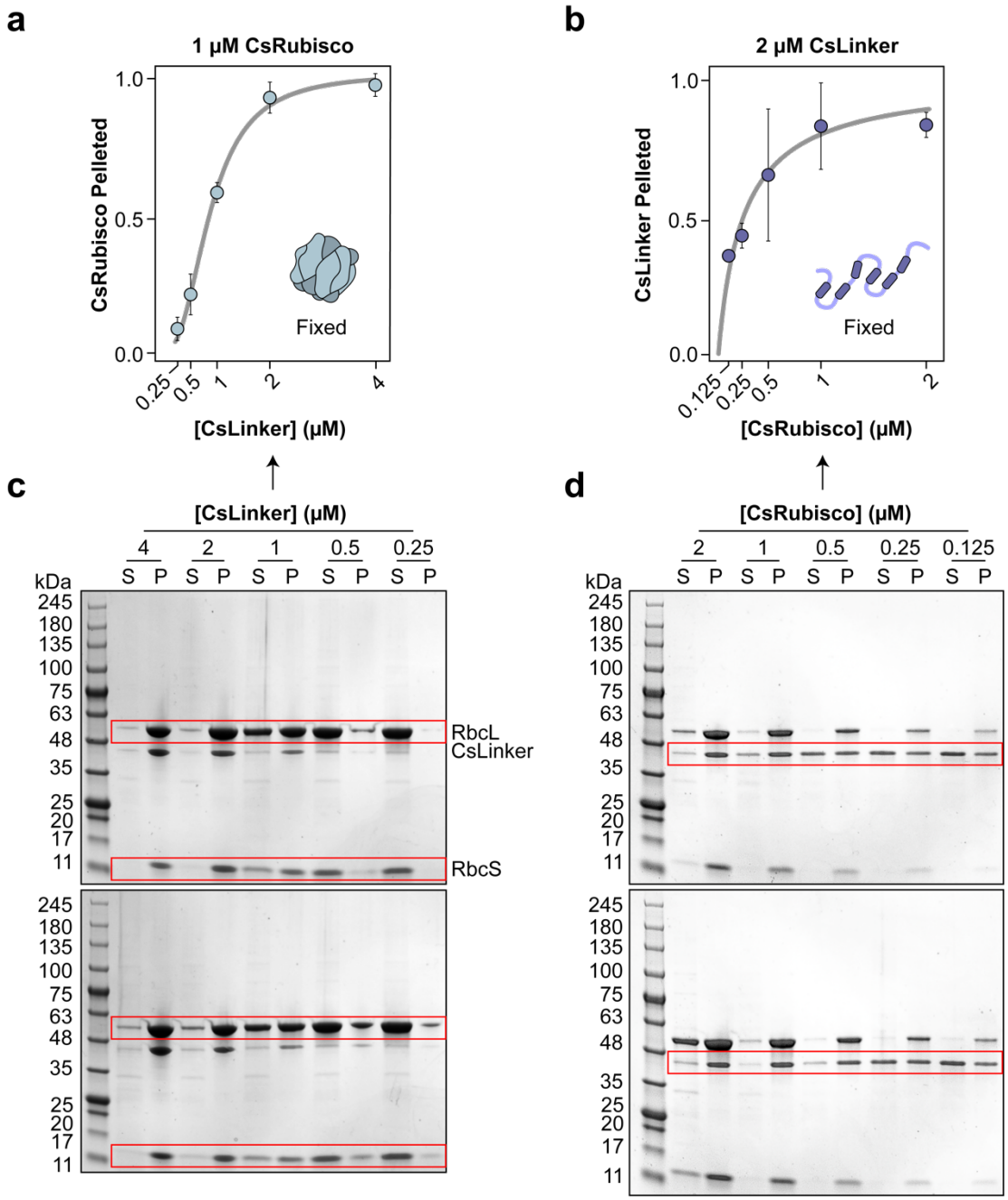


Figure legend overleaf.

Figure 3-12 | Quantification of *in vitro* CsLinker-CsRubisco droplet formation by reciprocal titration droplet sedimentation assays.
(a) Quantification of CsRubisco pelleted in droplet sedimentation assays in which CsRubisco concentration was fixed at 1 μM and CsLinker was titrated at the indicated concentrations. A hill model of binding was fitted to the data ($Y = (Y_{max} * X^h)/(K_{0.5}^h+X^h)$, where h is the hill coefficient). Error bars represent S.D. (b) Quantification of CsLinker pelleted in droplet sedimentation assays in which CsLinker concentration was fixed at 2 μM and CsRubisco was titrated at the indicated concentrations. The same analysis as in a was performed. (c) SDS-PAGE analysis of droplet sedimentation assays at fixed 1 μM CsRubisco concentration. Red box indicates the regions which were quantified by densitometry to produce Fig. 3-12a, which represents the average of the RbcL and RbcS measurements (d) SDS-PAGE analysis of droplet sedimentation assays at fixed 2 μM CsLinker concentration. The red box indicates the region which was quantified by densitometry to produce Fig. 3-12b.

As it was hypothesised the CsLinker-CsRubisco droplets were formed by liquid-liquid phase separation, their formation should only be triggered above a certain critical concentration that is dependent on solution conditions (Alberti, 2017; McSwiggen *et al.,* 2019). Analysis of solutions at increasing global protein concentrations with a fixed 2-fold excess of CsLinker to CsRubisco at pH 8.0 and 50 mM NaCl indeed indicated a critical concentration of ~0.6 μM global concentration (0.2 μM CsRubisco, 0.4 μM CsLinker) (Fig. 3-13), consistent with the traversion of a phase separation boundary.

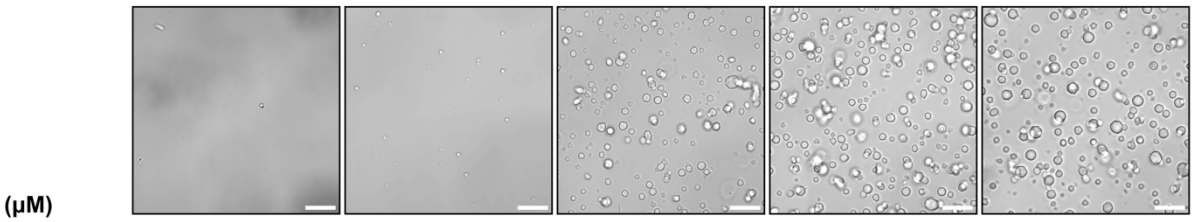


Figure 3-13 | Quantification of CsLinker-CsRubisco critical concentration for droplet formation.
Brightfield microscopy images of CsLinker-CsRubisco solutions at the indicated concentrations. Scale bar = 5 μm.

Given the analogous secondary structure of CsLinker and EPYC1 (Fig. 1-4), it was hypothesised the interaction between CsLinker and CsRubisco is also underpinned by a similar interaction of the helical regions of CsLinker, as is the case for EPYC1 (He *et al.,* 2020). Given the electrostatic nature of residues within the predicted helices of CsLinker, droplet formation should show the same salt-dependency as observed previously for the Chlamydomonas reconstitution, where 300 mM NaCl completely eradicated droplet formation presumably due to the ablation of salt bridges between EPYC1 and Chlamydomonas Rubisco (Wunder *et al.,* 2018). Droplet formation with CsLinker and CsRubisco as assessed by sedimentation assay demonstrated a similarly strong salt dependency, with a sharp transition

above 150 mM NaCl (Fig. 3-14). The apparent eradication of droplet formation at increased salt concentrations was consistent with the screening of electrostatic interactions between CsLinker and CsRubisco that shift the phase boundary to higher global protein concentrations. Taken together the results from this section demonstrated that droplets formed under quasi-physiological conditions exhibit hallmarks that begin to support their formation is due to liquid-liquid phase separation of the proteins, namely: droplet roundness (Fig. 3-13), critical concentration (Fig. 3-13) and salt-dependency (Fig. 3-14). To further demonstrate hallmarks of LLPS (*e.g.,* internal and external mixing, and fusion) it was necessary to label the components of the droplets.

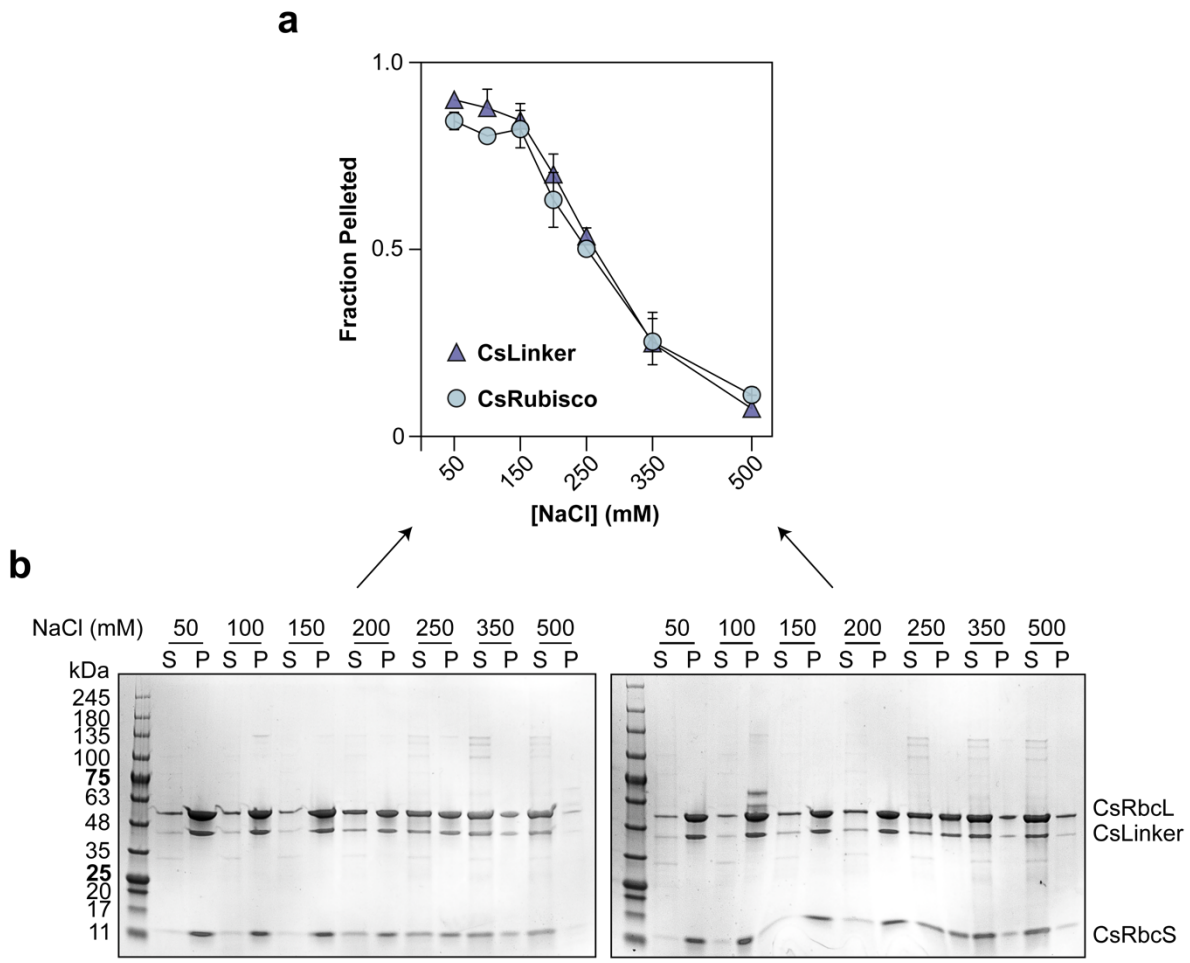


Figure 3-14 | Quantification of salt-dependency for droplet formation with CsLinker and CsRubisco.
(a) Quantification of CsLinker (purple triangles) and CsRubisco (blue circles) in droplet sedimentation assays. For CsRubisco, an average of CsRbcL and CsRbcS band intensity was taken. Error bars represent S.D. (b) SDS-PAGE analysis of droplets formed with 1 μM CsRubisco and 2 μM CsRubisco at the indicated salt concentrations.

3.3.4. Labelling CsLinker and CsRubisco

Although fluorescent tags have been shown to affect phase separation dynamics (Uebel & Phillips, 2019), when used at low concentrations relative to the untagged protein (~5% molar concentration), the effect of the tag on the droplet properties is assumed to be minimal. An N-terminally labelled mEGFP-CsLinker fusion was accordingly produced and purified. Although increased cleavage of the full length product was observed during the purification, presumably due to the lack of the C-terminal protective MBP-6xHis fusion, the final product was still of high purity (Fig. 3-15a). Due to the lack of *in vivo* tractability in Chlorella (Table 3-6), a fluorescent protein fusion of CsRubisco could not be created and purified for use in the assays. Instead, a previously employed Atto594-labelling approach was employed (Oh *et al.,* 2023; Schaefer, 2023), that labelled CsRubisco non-specifically at exposed surface amine groups of the purified CsRubisco protein. An estimate of the degree of labelling was made based on the absorption of the sample at 594 nm and 280 nm using equation 3.1 and indicated ~50 Atto-594 dyes per Rubisco holoenzyme.

Equation 3.1: Calculation of the degree of Atto594 labelling of Rubisco.

$$Degree\ of\ labelling\ = \frac{A_{max} \cdot \varepsilon_{prot}}{(A_{280} - A_{max} \cdot \mathrm{CF}_{280}) \cdot \varepsilon_{max}}$$

Where:
A_{max} is the absorbance at the absorption maxima of the dye (594 nm)
ε_{prot} is the extinction coefficient of the protein at 280 nm (848,000)
A_{280} is the absorbance at 280 nm (0.38)
CF_{280} is the correction factor determined at 280 nm due to absorbance of the dye (0.51)
ε_{max} is the extinction coefficient of the dye at the absorption maximum (120,000)

$$Degree\ of\ labelling\ = \frac{0.584 \cdot 848{,}000}{(0.38 - 0.584 \cdot 0.51) \cdot 120{,}000}$$

$$\boldsymbol{Degree\ of\ labelling\ =\ 50.2\ dyes\ /\ Rubisco}$$

The resulting labelled proteins were added to droplets formed at the empirically determined chloroplast concentrations (2 µM CsRubisco, 4 µM CsLinker) at 0.5% and 5% molar ratio relative to the untagged components respectively. As expected, signal from both proteins was present throughout the droplets (Fig. 3-15b), consistent with the results of the untagged proteins as assessed by droplet sedimentation (Fig. 3-11), indicating the droplets were appropriately labelled for further assessment.

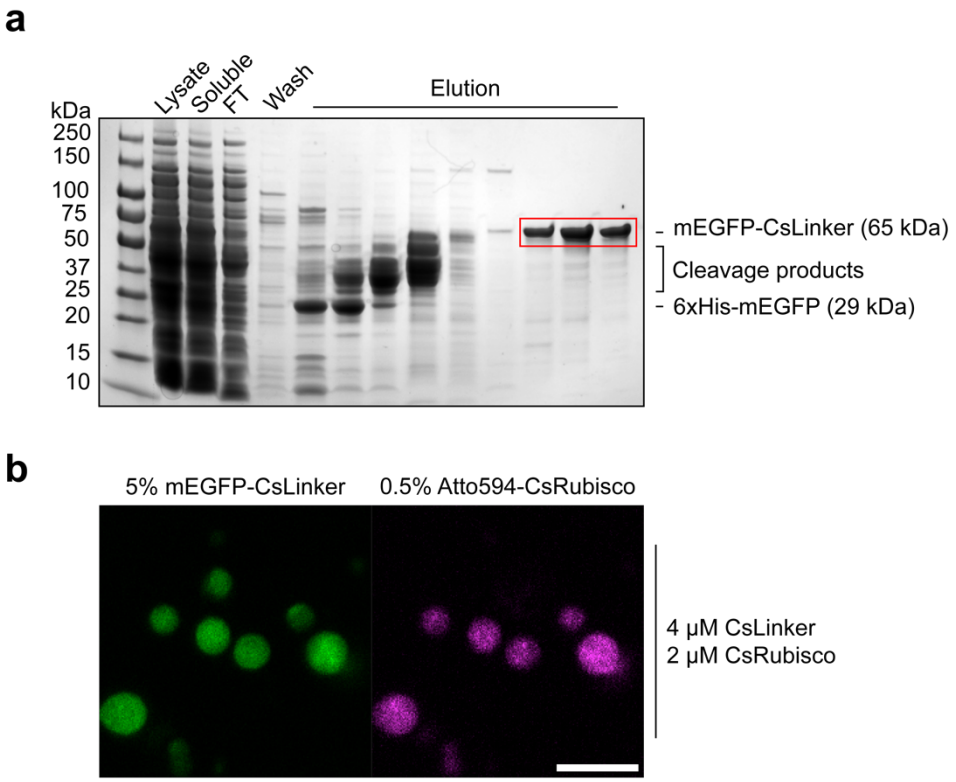


Figure 3-15 | Labelling of CsLinker-CsRubisco droplets.
(a) SDS-PAGE analysis of samples taken during purification of mEGFP-CsLinker. FT = flow though from the nickel column. The red box indicates the fractions that were pooled. The cleavage products and free mEGFP present because of cleavage are also indicated. (b) Confocal fluorescence microscopy images of droplets formed under the indicated conditions labelled with 5% molar ratio mEGFP-tagged CsLinker and 0.5% Atto594 labelled CsRubisco. Scale bar = 5 µm.

3.3.5. Assessment of the liquid-like properties of labelled CsLinker-CsRubisco droplets

Both mEGFP-CsLinker and Atto594-CsRubisco labelled droplets were observed to undergo fusion by coalescence over second timescales (Fig. 3-16a,b), though anecdotally, such events were not as commonly observed as expected, based on the observations of the Chlamydomonas reconstitution (Wunder *et al.,* 2018). Observation of droplets at the bottom of the imaging chamber in the minutes following formation demonstrated that the droplets tended to remain segregated as round droplets rather than decomposing into large amorphous condensates, suggesting a qualitatively higher barrier to fusion than observed in the Chlamydomonas reconstitution (Fig. 3-16c). Regardless, the apparent fusion by coalescence of the droplets demonstrated a definite liquid-like property of the droplets in their capability for post-fusion surface relaxation.

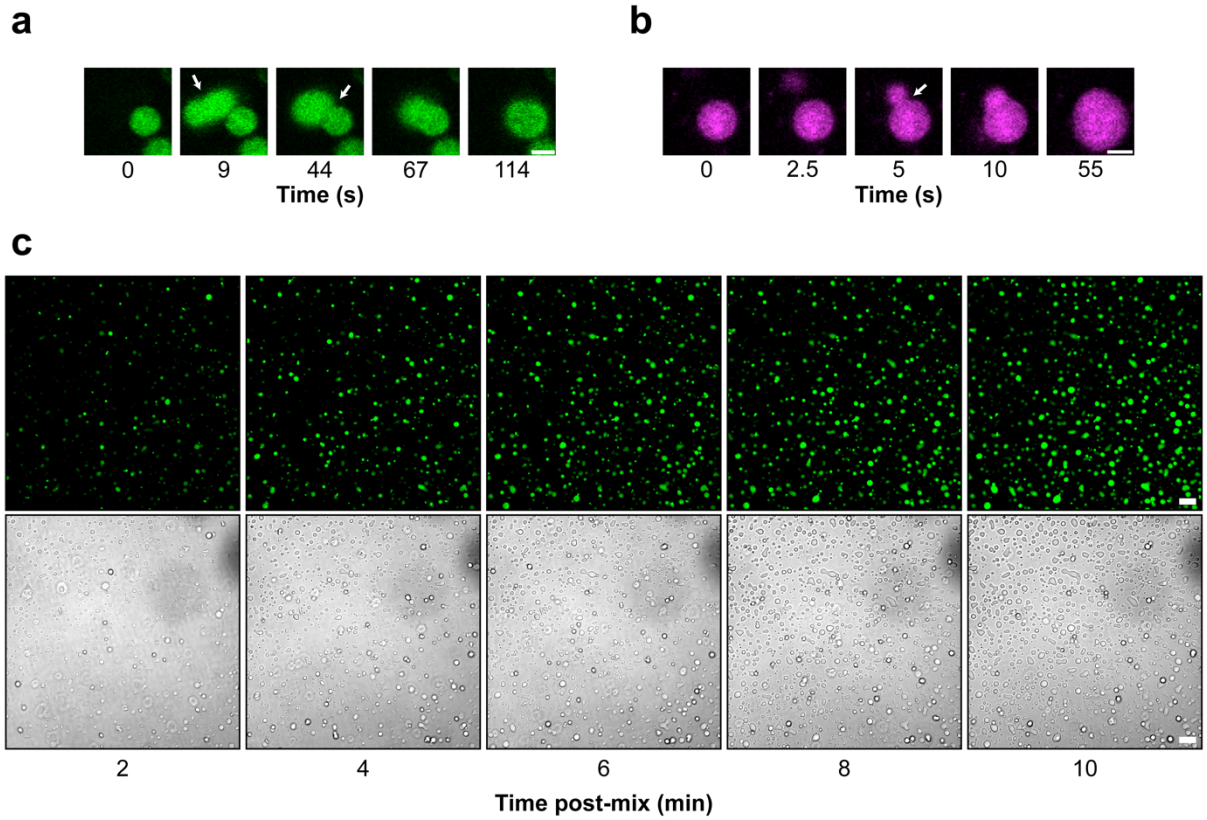


Figure 3-16 | Analysis of fusion and lack thereof in labelled CsLinker-CsRubisco droplets.
(a) Confocal microscopy images during the fusion of 5% mEGFP-CsLinker labelled droplets formed at 2 µM CsRubisco, 4 µM CsLinker. Arrows indicate imminent fusion events complete in the next frame. Scale bar = 1 µm. (b) Confocal microscopy images during the fusion of 0.5% Atto594-CsRubisco labelled droplets formed at 2 µM CsRubisco, 4 µM CsLinker. Arrows indicate imminent fusion events present in the next frame. Scale bar = 1 µm. (c) Wide confocal microscopy images taken at the bottom of the imaging chamber in the minutes following droplet formation with 5% mEGFP-CsLinker labelled droplets formed at 2 µM CsRubisco, 4 µM CsLinker. Scale bar = 5 µm.

To assess internal and external mixing of the droplet components, fluorescence recovery after photobleaching (FRAP) experiments were pursued. FRAP experiments have been used to assess the relative mobility of components in both *in vivo* and *in vitro* Rubisco-linker settings (Freeman Rosenzweig *et al.,* 2017; Oh *et al.,* 2023; Wunder *et al.,* 2018; Zang *et al.,* 2021). Half FRAP experiments, which primarily delineate internal mixing, demonstrated the mobility of mEGFP-CsLinker within droplets formed under the quasi-physiological conditions (Fig. 3-17a,b). Bleaching a region that represented approximately half the area of droplets with an ~1 µm radius (~1.6 µm^2) gave rise to an apparent $T_{0.5}$ of 131 seconds (95 confidence interval (CI_{95}): 121-142 seconds) (Fig. 3-17a,b). Although average FRAP curves are presented here, all $T_{0.5}$ values were calculated by individual fitting to each FRAP recovery curve (see methods). Whole FRAP experiments of droplets with the same

approximate radius indicated a slower apparent $T_{0.5}$ of 201 seconds (CI_{95}: 165-238 seconds), that was consistent with the increased area of bleaching (~3.2 μm^2; Fig. 3-17c,e), and plainly demonstrated exchange of CsLinker with the dilute phase. The apparent $T_{0.5}$ scaled with the bleach area between the half FRAP and whole FRAP experiments (y = 82.8x; Fig. 3-17d), suggesting fluorescence recovery was limited by internal mixing of the condensate rather than supply of CsLinker to the droplet. This was made apparent by the ring pattern of fluorescence recovery that appeared following bleaching of the whole droplet (Fig. 3-17e, arrow).

Although the FRAP results demonstrate definite mobility of CsLinker both internally and externally, the measured $T_{0.5}$ values were much slower than those observed for EPYC1 in the *in vitro* Chlamydomonas reconstitutions, where bleaching of an ~1.25 μm^2 region gave rise to $T_{0.5}$ of ~10 seconds. Similarly, *in vivo* measurement of EPYC1-Venus mobility in an unreported area (available data suggests an ~1.6 μm^2 bleach area) demonstrated a $T_{0.5}$ of ~15 seconds (Freeman Rosenzweig *et al.,* 2017). No direct comparison can be made with PYCO1 mobility in Phaeodactylum reconstitutions due to the differential mobility observed under different droplet compositions, though FRAP of PYCO1-CFP *in vivo* demonstrated little recovery (Oh *et al.,* 2023). Taken together, the FRAP results of CsLinker and the previously reported lack of decomposition of droplets (Fig. 3-16c) suggested the Chlorella reconstitution possesses analogous but distinct properties than observed for the *in vitro* Chlamydomonas reconstitution. To probe these differences further, similar FRAP experiments were completed with Atto594-CsRubisco in droplets of the same composition.

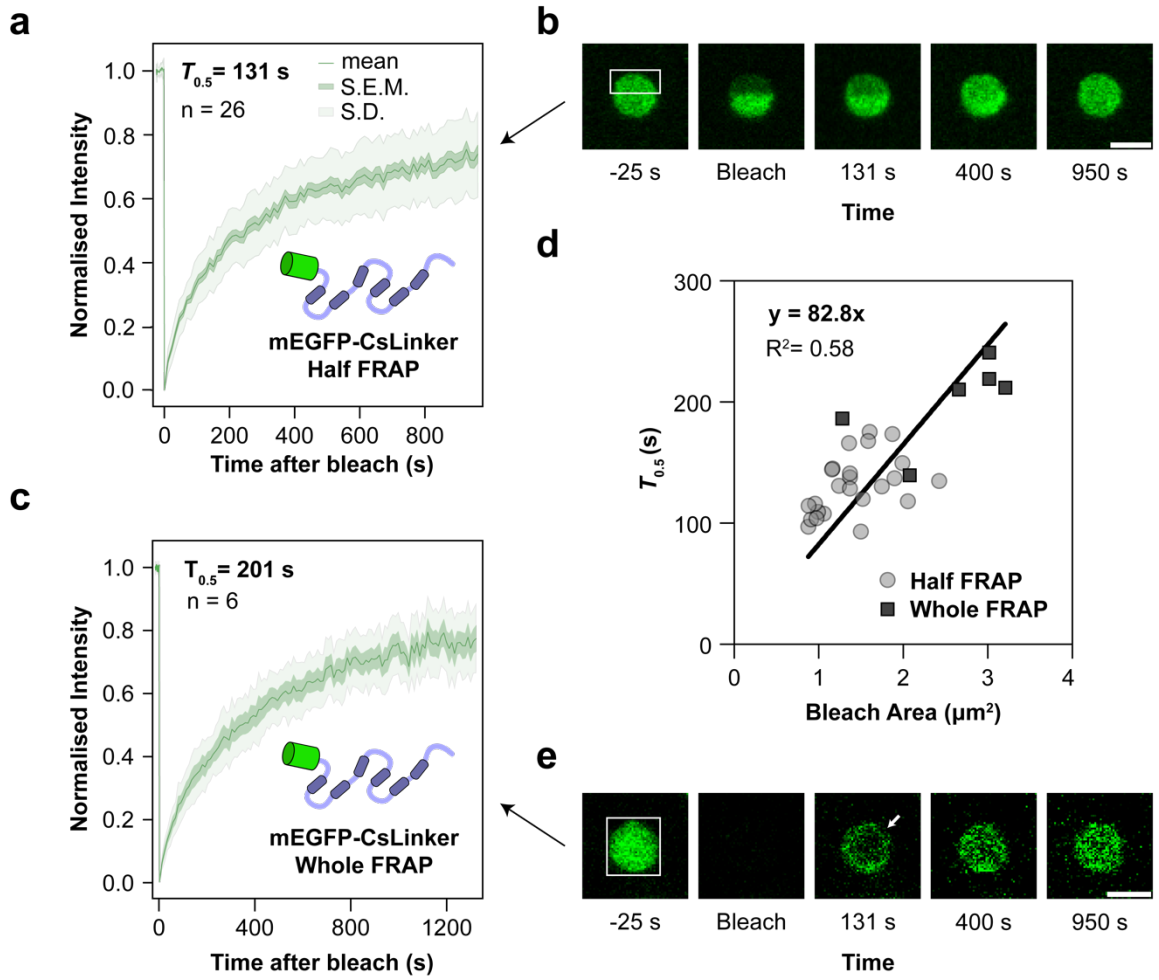


Figure 3-17 | Characterisation of mEGFP-CsLinker mobility in CsLinker-CsRubisco droplets.
(a) Average half-FRAP recovery curve of 5% mEGFP-CsLinker in CsLinker-CsRubisco droplets formed at quasi-physiological conditions (2 µM CsRubisco, 4 µM CsLinker). The mean $T_{0.5}$ from exponential fitting (see methods) of the 26 individual FRAP curves is reported. The full-scale normalised average FRAP curve is presented with the mean, S.E.M. and S.D. of each timepoint after the data was binned 5x for display. (b) Representative images from a half FRAP experiment of 5% mEGFP-CsLinker in CsLinker-CsRubisco droplets. The white box indicates the bleached region. Scale bar = 2 µm. (c) Average whole-FRAP recovery curve of 5% mEGFP-CsLinker in CsLinker-CsRubisco droplets formed at quasi-physiological conditions (2 µM CsRubisco, 4 µM CsLinker). The mean $T_{0.5}$ from exponential fitting of the 6 individual FRAP curves is reported. The full-scale normalised average FRAP curve is presented with the mean, S.E.M. and S.D. of each timepoint. (d) Scatter plot of individually calculated $T_{0.5}$ values for half- (circles) and whole- (squares) FRAP experiments. A linear regression is fitted to the data. (e) Representative images from a whole FRAP experiment of 5% mEGFP-CsLinker in CsLinker-CsRubisco droplets. The white arrow indicates the ring of fluorescence recovery following bleaching. The white box indicates the bleached region. Scale bar = 2 µm.

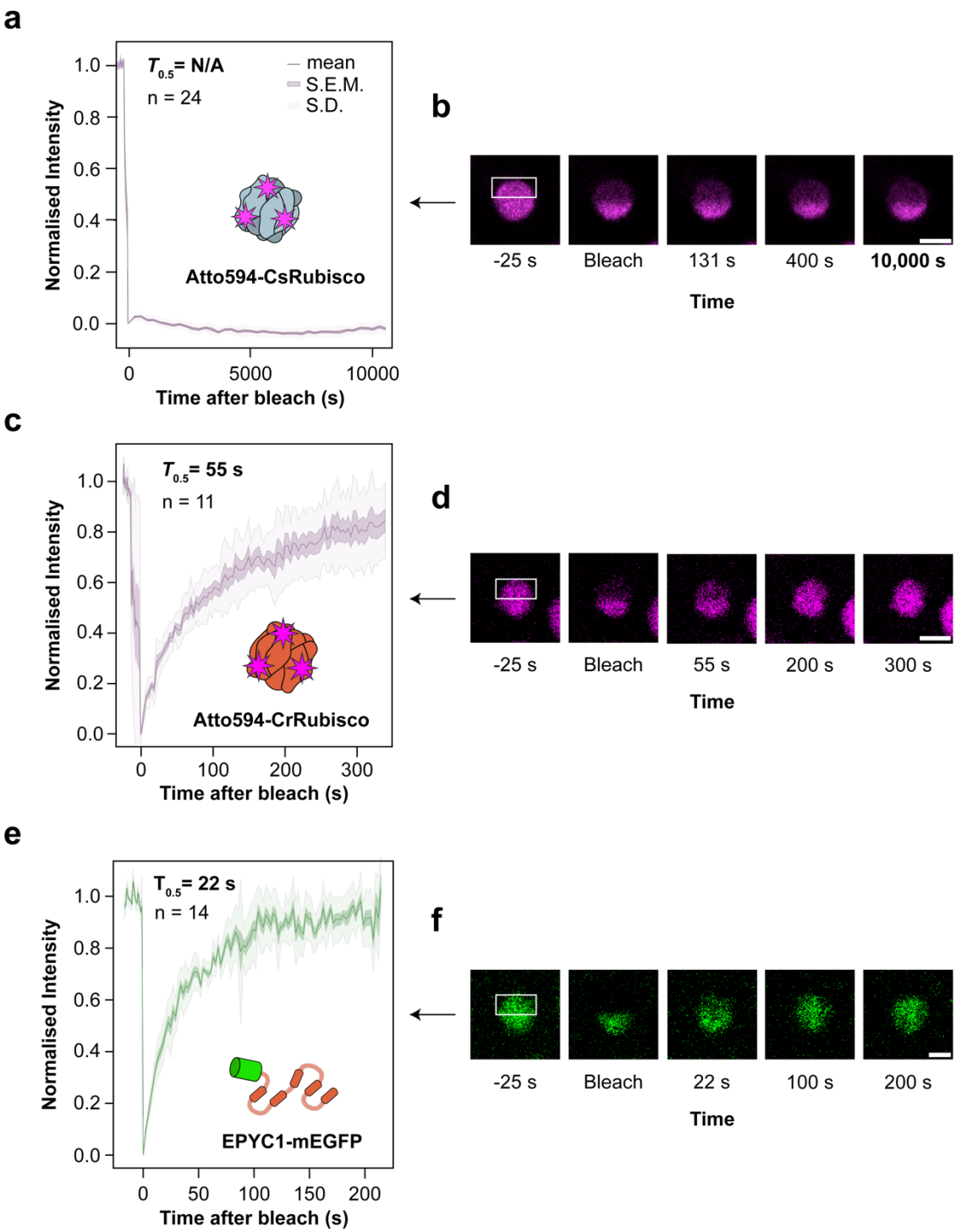


**Figure legend overleaf.*

Figure 3-18 | Comparison of Atto594-CsRubisco FRAP in CsLinker-CsRubisco droplets to *Chlamydomonas* droplets.
(a) Average half FRAP recovery curve of 0.5% Atto594-labelled Chlorella Rubisco (Atto594-CsRubisco) in CsLinker-CsRubisco droplets formed at quasi-physiological conditions (2 μM CsRubisco, 4 μM CsLinker). Clearly no $T_{0.5}$ value could be fitted to the data from the 24 individual FRAP curves. The full-scale normalised average FRAP curve is presented with the mean, S.E.M. and S.D. of each timepoint. (b) Representative images from a half FRAP experiment of Atto594-CsRubisco in CsLinker-CsRubisco droplets. The white box indicates the bleached region. Scale bar = 2 μm. (c) Average half FRAP recovery curve of 0.5% Atto594-labelled Chlamydomonas Rubisco (Atto594-CrRubisco) in EPYC1-CrRubisco droplets formed at 2 μM CrRubisco, 2 μM EPYC1. The mean $T_{0.5}$ from exponential fitting of the 11 individual FRAP curves is reported. The full-scale normalised average FRAP curve is presented with the mean, S.E.M. and S.D. of each timepoint. (d) Representative images from a half FRAP experiment of Atto594-CrRubisco in EPYC1-CrRubisco droplets. The white box indicates the bleached region. Scale bar = 2 μm. (e) Average half FRAP recovery curve of 5% EPYC1-mEGFP in EPYC1-CrRubisco droplets formed at 2 μM CrRubisco, 2 μM EPYC1. The mean $T_{0.5}$ from exponential fitting of the 14 individual FRAP curves is reported. The full-scale normalised average FRAP curve is presented with the mean, S.E.M. and S.D. of each timepoint. (f) Representative images from a half FRAP experiment of EPYC1-mEGFP in EPYC1-CrRubisco droplets. The white box indicates the bleached region. Scale bar = 2 μm.

Surprisingly, Atto594-CsRubisco showed no recovery of fluorescent signal over hour timescales in condensates, suggesting little internal or external mobility of CsRubisco (Fig. 3-18a,b). To confirm this result was not an artefact of Atto-labelling the Rubisco, a direct comparison with the Chlamydomonas reconstitution was made, where *in vivo* FRAP analysis has demonstrated the mobility of Rubisco (Freeman Rosenzweig *et al.,* 2017). The components of the Chlamydomonas reconstitution were purified according to the previous publication (see methods) (Wunder *et al.,* 2018). Droplets were formed under the same Rubisco concentration as the Chlorella droplets (2 μM), though the EPYC1 concentration was made equimolar, in line with *in vivo* abundance (Tables 3-4, 3-5). Atto594-labelled Chlamydomonas Rubisco (CrRubisco) within the droplets demonstrated an apparent $T_{0.5}$ of 55 seconds (CI_{95}: 44-67 seconds) when a region of ~1.6 μm^2 was bleached (Fig. 3-18c,d). The apparent $T_{0.5}$ of 5% GFP-tagged EPYC1 was 22 seconds (CI_{95}: 19-26 seconds) when a region of ~3 μm^2 was bleached in droplets of the same composition (Fig. 3-18e,d). Though the mobility of Rubisco in Chlamydomonas reconstitutions was not previously measured, the mobility measured here was largely consistent with the *in vivo* measured value (~30 seconds; (Freeman Rosenzweig *et al.,* 2017)). Likewise the value for EPYC1 mobility was not dissimilar from the *in vitro* (~10 seconds; (Wunder *et al.,* 2018)) and *in vivo* (~15 seconds; (Freeman Rosenzweig *et al.,* 2017)) values. Taken together the results suggest that under the quasi-physiological conditions the droplets were formed at, Chlorella Rubisco within them shows little mobility with respect to internal mixing or exchange with the dilute phase. Though there was little mobility in passively observed droplets (Fig. 3-18a,b), Atto594-CsRubisco was able to rearrange upon

surface relaxation following fusion by coalescence (Fig. 3-16b), suggesting that Rubisco is not completely immobilised within the droplets. This conclusion is similar to that of the Phaeodactylum reconstitution where Rubisco also demonstrates low mobility in FRAP experiments, but is able to rearrange during fusion events (Oh *et al.,* 2023). In this work Oh *et al.* (2023) also demonstrated that the composition of the droplets with respect to Rubisco and PYCO1 ratio had a large impact on the mobility of the condensate components. Alternative compositions were not tested here for the Chlorella reconstitution but may also have an impact on relative component mobility despite the lack of homotypic phase separation of CsLinker (Fig. 3-11).

It is not unfeasible to suggest that the observed relative mobility of the components in the Chlorella reconstitution is reflective of the *in vivo* condensate, given that the conditions used in the experiments were assumed quasi-physiological according to the measured *in vivo* concentrations. Immobility of Rubisco is readily observed in α-carboxysomes, where Rubisco is presumed to be phase separated by CsoS2 during assembly (Oltrogge *et al.,* 2020), but demonstrates immobile higher order upon assembly and still permits freedom of motion of other essential carboxysome components (Metskas *et al.,* 2022; Ni *et al.,* 2022). Similarly, Rubisco also appears to have low mobility in β-carboxysomes (Chen *et al.,* 2013), further reinforcing that internal mixing of Rubisco is not essential to its function in the condensed state. Nevertheless, further investigation of the arrangement of Rubisco in Chlorella pyrenoids *in vivo* using established methods (Engel *et al.,* 2015; Freeman Rosenzweig *et al.,* 2017; He *et al.,* 2020), as well as further assessment of component mobility *in vitro* should be pursued.

3.4. Discussion

In this chapter it was aimed to validate the role of the *in silico* identified putative linker protein of the *Chlorella* species. Although the protein was not directly localised to the pyrenoid, overwhelmingly, the evidence suggested the hypothesised role as a pyrenoid linker protein was valid. Firstly, the predicted chloroplast transit peptide of the protein was not detected in any mass spectrometry experiments despite extensive coverage of the rest of the protein (Fig. 3-1a). This evidence was consistent with the import of CsLinker into the chloroplast. Secondly, assuming correct import, CsLinker was demonstrated to be an abundant component of the chloroplast (Fig. 3-5c), and the abundance was directly linked to CO_2 availability, consistent with the hypothesised role in low CO_2 induced pyrenoid biogenesis and CCM activation. Thirdly, CsLinker and CsRubisco demonstrated a high confidence interaction *ex vivo* (Fig. 3-9), which was replicated *in vitro* (Fig. 3-10). Lastly, the purified proteins at conditions analogous to the *in vivo* scenario, demonstrated heterotypic phase separation into droplets that possessed hallmarks of liquid-liquid phase separation.

Despite the convincing *ex vivo* and *in vitro* validation of the role of CsLinker, the characterisation was not completed without resistance. Primarily the challenges encountered in this chapter were related to attempts to directly validate CsLinker's role *in vivo*. Alongside the work of Dr Mihris Naduthodi, a huge investment to replicate the reported nuclear transformation of Chlorella species was undertaken. Direct attempts to replicate published results with existing constructs as well as novel methodologies and constructs were all performed with no success. This outcome is consistent with verbal communication with several members of the CCM community at conferences, who expressed similar experiences in trying to replicate published Chlorella transformation protocols. It is therefore tentatively concluded that no *bona fide* nuclear transformation protocol for Chlorella exists, and that existing publications are 'artefactual' in their nature. Undoubtedly, should a robust protocol for the transformation of *Chlorella* species exist, there would be huge interest from a host of industrial sectors given the huge interest in *Chlorella* for a range of applications including CO_2-sequestration, bio-production and supplementation (Liu & Chen, 2016). The lack of genetic tractability precluded two pieces of primary evidence that would have provided almost total support for the hypothesised role as

a linker protein. Firstly, fluorescent tagging of the protein would have allowed localisation of CsLinker and CsRubisco within the pyrenoid and provided an important comparison for the mobility of components *in vivo*, given the surprising lack of mobility of CsRubisco demonstrated *in vitro*. Secondly, if the CsLinker protein could have been knocked down by RNA interference, or knocked out by genetic disruption (random integration or CRISPR), a direct correlation between pyrenoid presence and CsLinker expression could have been explored, as done so previously for EPYC1 in Chlamydomonas (Mackinder *et al.,* 2016) and CsoS2 in α-carboxysomes (Cai *et al.,* 2015). This second piece of evidence would also allow a direct link to functionality of the pyrenoid to be made. *In lieu* of this direct evidence in Chlorella, alternative '*in capsa*' approaches using a Chlamydomonas chassis strain are explored in Chapter 5 and demonstrate both the localisation and functionality of CsLinker by proxy. The lack of genetic tractability will pose the same challenges for the future characterisation of putative pyrenoid assembly proteins PMTT and PMST, where demonstration of the proposed functionality of the proteins *in vitro* will likely be more challenging. Proposed alternative *in capsa* approaches are further discussed in Chapter 5.

Encouragingly, co-immunoprecipitation of the PMTT and PMST proteins by CsRubisco demonstrated interaction between the proteins and was consistent with their association with the pyrenoid in Chlorella. Importantly, the co-IP experiments also showed that none of the 16 other protein sequences outputted by FLIPPer analysis of the *Chlorella sorokiniana* UTEX1230 genome (Table 2-4) were detected as potential CsRubisco interactors, lessening the likelihood that alternative linker candidates could behave analogously and strengthening the case for CsLinker as the sole linker protein in Chlorella. Likewise, no additional potential pyrenoid assembly proteins (PMTT/PMST analogues) were detected in the experiments, consistent with the hypothesis that the blueprints for pyrenoid assembly in Chlorella may be simpler than those for Chlamydomonas (Meyer, Itakura, *et al.,* 2020). This, alongside the simpler morphology of *Chlorella* pyrenoids, is of great interest with respect to engineering pyrenoid-based CCMs into C_3 crop plants (Adler *et al.,* 2022). In line with this goal, further work to determine a more complete pyrenoid proteome would greatly contribute to the understanding of pyrenoid assembly in Chlorella and allow further assessment of hypotheses outlined in Chapter 2. A precedent for pyrenoid

isolation in *Chlorella vulgaris* already exists (Villarejo *et al.,* 1998), and a mass spectrometry-based approach has previously been used to determine the Chlamydomonas pyrenoid proteome (Mackinder *et al.,* 2016; Zhan *et al.,* 2018). Both approaches should be explored in Chlorella.

In vitro characterisation of CsLinker, CsRubisco and their interaction demonstrated some key similarities and differences between the Chlorella *in vitro* reconstitution and those of Chlamydomonas and Phaeodactylum. In line with the putative role of CsLinker, and those of EPYC1 and PYCO1, all three proteins demonstrated LLPS of their respective Rubisco proteins at physiological concentrations (<5 μM). Consistent with the phylogeny and apparent concurrent convergent evolution of CsLinker and EPYC1, more parallels could be drawn between the Chlorella and Chlamydomonas reconstitutions than that of Phaeodactylum. Most strikingly, CsLinker and EPYC1 did not demonstrate homotypic phase separation, in contrast to PYCO1. It is unclear whether this behaviour of PYCO1 plays a functional role *in vivo*, and consequently whether this was a requirement during its evolution, but this behaviour was clearly not a constraint during the evolution of CsLinker and EPYC1. From the structural similarity of CsLinker and EPYC1 it was expected the droplets would also behave analogously. Although droplet formation demonstrated a similar salt-dependency and relative compositional requirement, the mesoscopic properties of the droplets formed under similar compositions were clearly different. Importantly, the Chlamydomonas reconstitution was recapitulated in this study which allowed direct comparison between the two systems. Whilst all the hallmarks of LLPS were observed for the Chlorella droplets also, there was a stark contrast in the mobility of both CsLinker and CsRubisco, suggesting differences in the *modus operandi* of the CsLinker-CsRubisco interaction which are explored in the following chapter.

3.5. Material and Methods

3.5.1. Strains and culture conditions

3.5.1.1. Escherichia coli (E. coli)

E. coli strains used in this study (Dh5α and BL21(DE3)) were maintained using 25 g L^{-1} Miller's modified Luria broth media (LB (Miller); L3522 – Sigma Aldrich). Where necessary, agar powder (A7921 – Sigma Aldrich) was added to 1.5 % weight by volume (w/v). Media was sterilised by autoclaving at 121 ºC for 15 minutes. Antibiotics were added to liquid media or melted agar media cooled to ~60 ºC according to table 3-8. Overnight cultures were of 5 mL volume and cultivated in 15 mL round-bottomed culture tubes. Plates were of ~15 mL volume in 90 mm Petri dishes. Large-scale cultures for protein purification were of 1 or 1.5 L volume and cultivated in 2.5 L half-baffle shake flasks. All cultivation was completed at 37 ºC and where relevant shaking was completed at 200 rpm.

Table 3-8: Antibiotic concentrations used in LB media.

Antibiotic.	Working Concentration	Stock
Carbenicillin	100 µg mL^{-1}	100 mg mL^{-1} dissolved in dH_2O
Chloramphenicol	25 µg mL^{-1}	25 mg mL^{-1} dissolved in 100% EtOH
Kanamycin	50 µg mL^{-1}	50 mg mL^{-1} dissolved in dH_2O

3.5.1.2. Algae

Wild type (WT) *Chlorella sorokiniana* (UTEX1230; SAG211-8k) and *Chlamydomonas reinhardtii* (CC-4533) algal strains were maintained using Tris-acetate phosphate (TAP) or Tris-HCl phosphate (TP) media with Kropat's revised trace elements (Kropat *et al.,* 2011). Where necessary, agar powder (A7921 – Sigma Aldrich) was added to 1.5 % (w/v) to maintain cultures on plates. Unless otherwise stated, algal strains were cultivated under ambient conditions in 50 mL volumes in Erlenmeyer flasks of 100 mL with shaking at 140 rpm. Light conditions were ~50 µmol photons m^{-2} s^{-1} and temperature was ~21 ºC.

3.5.2. General molecular biology methods

3.5.2.1. Polymerase chain reaction (PCR)

All PCRs were completed using Phusion™ high-fidelity DNA polymerase (F530 – ThermoFisher Scientific) with composition according to table 3-9. In all cases an extension time of 30 seconds per kilobase of target amplicon was calculated. The annealing temperature (T_m) for each reaction was calculated using the ThermoFisher Scientific online T_m calculator, after which 0.5 ºC a was subtracted. Otherwise, conditions are stated in table 3-10.

Table 3-9: Phusion™ PCR composition.

Component	Volume (µL)
dH_2O	28
5x Phusion™ GC Buffer	10
dNTP mix (R0192 – ThermoFisher)	1
Forward primer (10 µM stock)	2.5
Reverse primer (10 µM stock)	2.5
DMSO (F515 – ThermoFisher)	4.5
Template (10 ng μL^{-1} plasmid, ~50 ng μL^{-1} genomic)	1
Phusion™ DNA Polymerase (2 U μL^{-1})	0.5
Total	**50**

Table 3-10: Phusion™ PCR conditions.

Temperature (ºC)	Time (s)	Number of cycles
98	30	1
98	10	35
Calculated	30	
72	Calculated	
72	600	1

3.5.2.2. Gel Electrophoresis

Gel electrophoresis of PCR products and digested plasmids was completed using Tris-acetate EDTA (TAE) 1% w/v agarose (A9539 – Sigma Aldrich) gels run at 120 V until the desired migration was achieved. Gels were visualised using a UVITEC Essential V6 gel imaging system (UVITEC).

3.5.2.3. PCR purification, gel extraction and miniprep

PCR purification was completed using either a MinElute or QIAquick PCR purification kit (Qiagen) according to manufacturer's instructions. Gel extraction was completed using either MinElute or QIAquick gel extraction kits. Plasmids were purified using a QIAprep spin miniprep kit (Qiagen) according to manufacturer's instructions. DNA concentrations were subsequently quantified using a NanoDrop 1000 spectrophotometer (ThermoFisher Scientific).

3.5.2.4. E. coli transformation

E. coli strains were transformed by heat-shock method. 50 µL aliquots of cells were defrosted on ice before mixing with 5 µL of genetic material and incubation on ice for 30 minutes. Heat-shock was performed at 42 ºC for 30 seconds before recovery on ice for 5 minutes. 600 µL of SOC media was added and the cells were recovered for 90 minutes at 37 ºC with 200 rpm shaking. Recovered cells were collected by centrifugation for 3 minutes at 4,000 *g* prior to plating in a volume of ~50-100 µL on LB-agar plates with the relevant antibiotic selection (Table 3-8).

3.5.2.5. Golden gate assembly

Golden gate assembly was completed using composition and conditions according to tables 3-11 and 3-12. Type IIs restriction enzymes used were either BsaI-HF®v2 (R3733 – New England BioLabs [20 U μL^{-1}) or BpiI (ER1011 – ThermoFisher Scientific [10 U μL^{-1}]).

Table 3-11: General composition of Golden Gate assembly reactions.

Component	Concentration
Insert(s)	50 fmol
Acceptor vector	25 fmol
Type IIs restriction enzyme (BsaI/BpiI)	0.5 U μL^{-1}
T4 DNA Ligase (400 U μL^{-1}; M0202 – New England BioLabs)	10 U μL^{-1}
10x T4 DNA ligase buffer (New England BioLabs)	1x
dH_2O	To 10 µL

Table 3-12: General Golden Gate assembly conditions.

Temperature (ºC)	Time (m)	Cycles
37	5	35
16	5	35
37	10	1
65	20	1

3.5.2.6. Ligation-independent cloning (LIC)

2.5 µg of LIC acceptor vectors were linearised for 2 hours at 37 ºC with 1 µL of either SspI-HF (R3132 – New England BioLabs) for v1 vectors (1GFP) or HpaI (R0105 – New England BioLabs) for v2 vectors (2Cc-T) in 1x CutSmart buffer (New England BioLabs) and subsequently gel extracted. Separate LIC reactions were completed for the linearised vectors and inserts according to the composition in table 3-13. The reactions were incubated at 20 ºC for 40 minutes, followed by heat-inactivation at 75 ºC for 20 minutes. The vector and insert were subsequently mixed at a 1:3 molar ratio and incubated for 30 minutes at room temperature (~22 ºC) before transformation into Dh5α.

Table 3-13: General composition of LIC reactions for the vector.

Component	Concentration
Digested vector	~10 ng μL^{-1}
Buffer 2.1 (New England BioLabs)	1x
dGTP (for 1GFP vector and v2 inserts) or dCTP (for 2Cc-T vector and v1 inserts)	5 mM
DTT	10 mM
T4 DNA polymerase (M0203; New England BioLabs)	0.12 U μL^{-1}
dH_2O	To 10 µL

3.5.2.7. Sequencing

All sequencing reactions were completed using Eurofins LightRun Tube sequencing reactions (Eurofins Genomics) prepared according to manufacturer's instructions.

3.5.3. Sequencing of the CsLinker genomic locus

The CsLinker genomic locus was amplified in 4 parts (Fig. 3-1) using primers according to table 3-14. The products were analysed by gel electrophoresis (Fig. 3-1b) and subsequently PCR-purified and sequenced with the same primers used for their amplification.

Table 3-14: Primers used for amplification of CsLinker genomic locus.

Name	Purpose	Oligonucleotide Sequence
oLM2130	Part **a** Forward	AAGGCATTCCTGGAGCG
oLM2131	Part **a** Reverse	AAAAATGTGCTGGCACCC
oLM2132	Part **b** Forward	AAGGCATTCCTGGAGCG
oLM2133	Part **b** Reverse	AAAAATGTGCTGGCACCC
oLM2134	Part **c** Forward	AGTTCCTGGAGCGCAAG
oLM2135	Part **c** Reverse	GCTGAGTTCCACCCCAG
oLM2136	Part **d** Forward	AGACCGAGCTGCATGTTC
oLM2137	Part **d** Reverse	ACATCACACACACAGGCC

3.5.4. Cloning of CsLinker for protein purification and expression

The TargetP 2.0 predicted mature CsLinker sequence (Fig. 3-1c) was back-translated and codon-optimised using the Geneious back translation tool and *E. coli* K12 (high) codon usage table. Due to its repetitive nature, the nucleotide sequence was split and synthesised in 3 parts by TWIST Bioscience (pLM672, pLM673 and pLM674; Table 3-16). The 3 synthesised parts were re-assembled using Golden Gate Assembly into the pUAP1 universal acceptor plasmid (Addgene plasmid #63674) to create pLM701. The assembled vector was confirmed by sequencing using oLM271 and oLM272 (Table 3-14). The sequence of the mature CsLinker was PCR-amplified from pLM701 using oLM2341 and oLM2344 and subsequently gel extracted. Ligation-independent cloning (LIC) was used to clone the fragment into the pET His6 GFP TEV LIC cloning vector 1GFP (Addgene #29663) to create pLM712 which was used for the production and purification of mEGFP-CsLinker. The mEGFP-CsLinker sequence was amplified from pLM712 using oLM2487 and oLM2655 and assembled into the pET TEV MBP His6 cloning vector 2Cc-T (Addgene #37237) to create pLM872 which was used for the production and purification of the CsLinker protein. The sequence of the assembled vectors was confirmed by sequencing with oLM2025 and oLM2237.

Table 3-15: Primers used for cloning of CsLinker for protein expression and purification.

Name	Purpose	Oligonucleotide Sequence
oLM271	Forward sequencing of pLM701	AGTTGGAACCTCTTACGTGC
oLM272	Reverse sequencing of pLM701	TTTGAGTGAGCTGATACCGC
oLM2341	Forward primer for amplification of CsLinker from pLM701 for LIC cloning into pLM712	TACTTCCAATCCAATGCAATGGCCACAG CCCAATTAAGTGCTGAAC
oLM2344	Reverse primer for amplification of CsLinker from pLM701 for LIC cloning into pLM712	TTATCCACTTCCAATGTTATTACCTGCGG CGCGGGGTAGAGCTG
oLM2487	Forward primer for amplification of mEGFP-CsLinker from pLM712 for LIC cloning into pLM872	AGTTCCT TTTAAGAAGGAGATATAGTTCATGGGTT CTTCTCACCATCACCATC GGAGCGCAAG
oLM2655	Reverse primer for amplification of mEGFP-CsLinker from pLM712 for LIC cloning into pLM872	GGATTGGAAGTAGAGGTTCTCCCTGCG GCGCGGGGTAG
oLM2025	Forward sequencing of pLM712/pLM872	TAATACGACTCACTATAGGG
oLM2237	Reverse sequencing of pLM712/pLM872	GCTAGTTATTGCTCAGCGG

3.5.5. Cloning of Chlorella antibiotic resistance vectors for attempted transformation

To assemble the antibiotic resistance cassettes used for attempted transformation of Chlorella, ~500 bp regions corresponding to the predicted promoter and terminator of the Chlorella PSAD gene (pPSAD and tPSAD respectively) were amplified with primer pairs oLM2442/2443 and oLM2444/2445 respectively from Chlorella genomic DNA (Table 3-15). The resulting products were Golden Gate assembled separately into pUAP1 using BpiI to create L0 parts which were sequenced with oLM271 and oLM272. Likewise, the Rbcs2 promoter (pRbcS) and terminator (tRbcS) regions were amplified with primer pairs oLM2446/2447 and oLM2448/2449 respectively and assembled and sequenced the same way. The L0 promoter and terminator parts were assembled with L0 hygromycin (pCM0-073), paromomycin (pCM0-074) and zeocin (pCM0-077) resistance genes from the *Chlamydomonas* MoClo toolkit (Crozet *et al.,* 2018) in different combinations to create pLM845, pLM847, pLM848 and pLM849 (Table 3-16), which were used for the attempted transformations.

Table 3-16: Primers used for cloning of Chlorella antibiotic resistance cassettes.

Name	Purpose	Oligonucleotide Sequence
oLM2442	Chlorella PSAD Promoter Cloning Forward	ATCTACGAAGACATCTCAGGAGACAAGTGCGGGTAGAGC
oLM2443	Chlorella PSAD Promoter Cloning Primer Reverse	TAGTACGAAGACTACTCGCATTGTTTGGCCTCTGTCCTGG
oLM2444	Chlorella PSAD Terminator Cloning Primer Forward	ATCTACGAAGACATCTCAGCTTTCGGCCCAGCAGCTTG
oLM2445	Chlorella PSAD Terminator Cloning Primer Reverse	TAGTACGAAGACTACTCGAGCGCCAGGCTCGCAACAATCC
oLM2446	Chlorella RBCS2 Promoter Cloning Primer Forward	ATCTACGAAGACATCTCAGGAGCAATGGGTGTCGGCAGG
oLM2447	Chlorella RBCS2 Promoter Cloning Primer Reverse	TAGTACGAAGACTACTCGCATTTTGAAGGGATGTGGATGG
oLM2448	Chlorella RBCS2 Terminator Cloning Primer Forward	ATCTACGAAGACATCTCAGCTTCGGTCTGCGGCATGATG
oLM2449	Chlorella RBCS2 Terminator Cloning Primer Forward	TAGTACGAAGACTACTCGAGCGTGAATCTCTGCCGCACG

3.5.6. General protein biochemistry techniques

3.5.6.1. SDS-PAGE

Unless otherwise stated, Sodium Dodecyl Sulfate Polyacrylamide gel electrophoresis (SDS-PAGE) was completed using 10- or 15-well 4–20% Mini-PROTEAN® TGX™ Precast Protein Gels (4561094 or 4561096 – Bio-Rad Laboratories). Samples were prepared by mixing with Laemmli buffer to 1x final concentration and incubation at 100 ºC for 5-15 minutes. The running buffer for SDS-PAGE was 25 mM Tris, 192 mM Glycine, 0.1% SDS at pH 8.3 and gels were run at 200 V. Gels were stained using InstantBlue® Coomassie Protein Stain (ISB1L – Abcam) and visualised using an iBright CL750 imaging platform (ThermoFisher Scientific).

3.5.6.2. Native-PAGE

Unless otherwise stated, native PAGE experiments were completed using 15-well 4–20% Mini-PROTEAN® TGX™ gels. The 4x loading buffer consisted of 80 mM Tris-HCl pH 8.0, 200 mM NaCl, 40% glycerol and the running buffer was 25 mM Tris, 192 mM Glycine. Gels were run at 100 V for 4 hours at 4 ºC. Gels were stained and imaged as before.

3.5.6.3. Western blotting

Unless otherwise stated, an iBlot™ 2 system (IB21001 – ThermoFisher Scientific) and iBlot™ 2 mini nitrocellulose transfer stacks (IB23002 – ThermoFisher Scientific) were used for transfer of proteins from SDS-PAGE gels. The membrane was blocked with a 5% milk (w/v) solution in Tris-buffered saline (20 mM Tris, 150 mM NaCl) supplemented with 0.1% (w/v) Tween®-20 (TBST) for 1 hour at room temperature. The stated dilutions of the primary antibodies were subsequently incubated in 1% milk TBST with the membranes overnight at 4 ºC. Membranes were washed with TBST and incubated with 1:10,000 dilutions of secondary antibodies (Anti-rabbit Alexa488; A11008 or Anti-mouse Alexa555; A28180 – Invitrogen) in 1% TBST for 1 hour at room temperature prior to washing and imaging with an Amersham™ Typhoon™ 5 imaging platform (29187191 – Cytiva Life Sciences).

3.5.6.4. Densitometric analysis

Following separation by SDS-PAGE or native PAGE and in western blotting, densitometric analysis was completed using Fiji (Schindelin *et al.,* 2012) as described below. A horizontal region of interest was drawn over the bands (Fig. 3-18a), and the Analyze>Gels>Plot Lanes function was used to assess the gray value in each lane. A line was drawn at the base of each of the peaks (Fig. 3-18b) and the wand tool was used to measure the area of each of the bounded peaks. The fraction of pelleted material was derived by dividing the intensity of the pelleted lane by the sum of the supernatant and pelleted intensity for each sample.

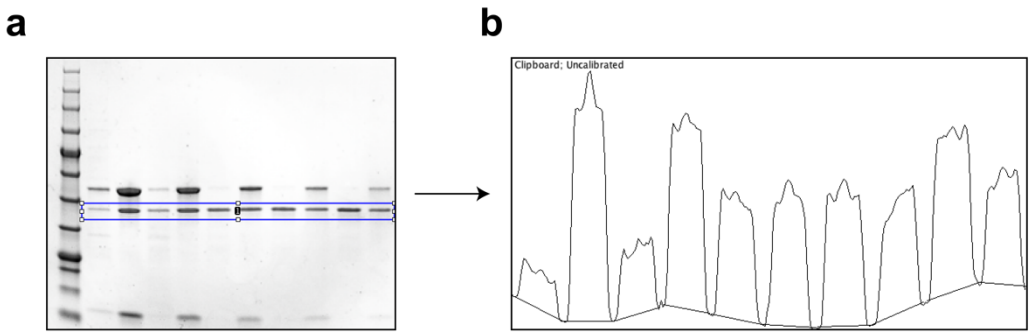


Figure 3-19 | Densitometric analysis of band intensity.
(a) Screenshot of an SDS-PAGE gel with the horizontal region of interest drawn in Fiji indicated. (b) Intensity of the lanes within the region of interest plotted with Fiji and shown with the lines drawn at the base of each of the peaks.

3.5.7. Protein expression, purification and labelling

3.5.7.1. Expression of mEGFP-CsLinker (pLM712) and CsLinker (pLM872) constructs

~50 ng of pLM712 and pLM872 were separately heat-shock transformed into BL21(DE3) cells and recovered on LB plates overnight under the relevant antibiotic selection at 37 ºC. A fresh transformation was completed each time expression was completed. A single colony from each plate was inoculated into a 50 mL volume of LB with the relevant antibiotic in 100 mL Erlenmeyer flasks and grown overnight at 37 ºC with 200 rpm shaking. All subsequent pre-cultures for other constructs were obtained in the same manner. The pre-culture was diluted 1:100 in 1 L of fresh LB with the relevant antibiotic. The cultures were grown to an OD_{600} of ~0.5-0.8 before induction with 1 mM Isopropyl-beta-D-thiogalactopyranoside (IPTG; 2938909090 – Apollo Scientific) followed by further growth for 2-3 hours at 37 ºC. At the end of induction the culture flasks were plunged into an ice bath before harvesting by centrifugation at 5,000 *g* for 10 minutes at 4 ºC in pre-cooled centrifuge and buckets. The pellets were snap-frozen in liquid nitrogen and stored at -20 ºC prior to purification.

3.5.7.2. Expression of TEV protease construct (pRK793)

A glycerol stock of BL21(DE3)-RIL harbouring the plasmid pRK793 (Addgene #8827) for the expression of TEV protease was obtained from Addgene and a published protocol (Tubb *et al.,* 2009) was mostly followed for expression and purification. The pre-culture was diluted 1:100 in 1 L of fresh LB with both antibiotics and grown under the same conditions to an OD_{600} of ~0.5 prior to induction

with 1 mM IPTG and further growth at 30 ºC for 4 hours. Cells were harvested by centrifugation at 5,000 *g* for 10 minutes prior to storage of the pellet at -20 ºC before purification.

3.5.7.3. Co-expression of pHueCrEPYC1 and pHueFLAGCrEPYC1GFP constructs with the pBadGroEL/S chaperone construct

Plasmids for the expression of EPYC1 and EPYC1-GFP as well as the chaperone GroEL/S were kindly provided by Tobias Wunder and Oliver Mueller-Cajar and as such the expression and purification of EPCY1 and EPYC1-GFP was almost solely based on the methods of Wunder *et al.,* (2018). ~100 ng of pBadGroEL/S and ~100 ng of the pHue plasmids containing either the EPYC1 of EPYC1-GFP sequence were co-transformed into BL21(DE3) by heat-shock transformation and recovered on plates containing antibiotics for the selection of both plasmids overnight at 37 ºC. The pre-culture was diluted 1:100 in 1 L of fresh LB with both antibiotics and the cultures were grown to an OD_{600} of ~0.5 before induction of expression of the GroEL/S chaperone with 0.2% (w/v) L-arabinose (104980250 – ThermoFisher Scientific). 20 minutes later, induction of the EPYC1 constructs was induced with 1 mM IPTG and growth was continued for 2-3 hours at 37 ºC with 200 rpm shaking. Cells were harvested by centrifugation at 5,000 *g* for 10 minutes prior to storage of the pellet at -20 ºC before purification.

3.5.7.4. Expression of the USP2 protease construct (pHUSP2-p2cc)

The pHUSP2-p2cc construct was kindly provided by Tobias Wunder and Oliver Mueller-Cajar and expression and purification was largely completed according to previous publication (Baker *et al.,* 2005). The pre-culture was diluted 1:100 in 1 L of fresh LB with both antibiotics and the cultures were grown to an OD_{600} of ~0.8 before induction of expression with 0.4 mM IPTG overnight at 23 ºC.

3.5.7.5. Purification of mEGFP-CsLinker

All steps during the purification of all proteins were fastidiously completed at 4 ºC or on ice. The pellet from 1 L of culture was resuspended in 50 mL of lysis buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 25 mM Imidazole), with a cOmplete EDTA-free protease inhibitor table (11873580001 – Roche). Phenylmethylsulfonyl fluoride (PMSF; 10837091001- Merck) was added to 2 mM immediately prior

to lysis which was completed by ultrasonication with 3 minutes total processing time. 3 second pulses were interspersed by 7 second breaks to allow sample cooling. The lysate was clarified by centrifugation at 47,850 *g* for 30 minutes at 4 ºC and subsequent 0.2 µm filtration of the supernatant. The clarified supernatant was loaded onto a 5 mL HisTrap FF crude column at 5 mL min^{-1} (17525501 – Cytiva Life Sciences) using an ÄKTA go purification platform (29383015 – Cytiva Life Sciences) and the column was washed with ~10 column volumes of lysis buffer until the A_{280} value reached baseline. The bound protein was eluted with a linear gradient over 10 column volumes to 100% of the elution buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 500 mM Imidazole). The eluted fractions were analysed by SDS-PAGE (Fig. 3-14) and the purest fractions were pooled. The pooled fractions were subjected to size-exclusion chromatography (SEC) using a HiLoad 16/600 Superdex 75 pg column (28989333 – Cytiva Life Sciences) equilibrated with gel filtration buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl). Eluted fractions were pooled and concentrated using Amicon® Ultra-15 Ultracel-10 regenerated cellulose membrane centrifugal filter units (UFC901024 – Merck Millipore). 5% glycerol (v/v) was added prior to snap-freezing in liquid nitrogen and storage at -80 ºC

3.5.7.6. Purification of TEV Protease

The cell pellet was resuspended in TEV lysis buffer (50 mM sodium phosphate (pH 8.0), 200 mM NaCl, 10% glycerol (v/v), 25 mM Imidazole) with a cOmplete tablet. PMSF was added to 1 mM prior to lysis as for mEGFP-CsLinker. The remainder of the purification was completed as for the mEGFP-CsLinker also except that the elution buffer was TEV elution buffer (50 mM sodium phosphate (pH 8.0), 200 mM NaCl, 10% glycerol (v/v), 500 mM Imidazole). The elution was pooled and snap-frozen for storage at -80 ºC.

3.5.7.7. Purification of untagged CsLinker

The full length mEGFP-CsLinker-MBP fusion protein was purified in the same way as the mEGFP-CsLinker except the pooled eluate from the HisTrap column was not subjected to SEC. To cleave the fusion tags, purified TEV protease was added in a 1:100 ratio relative to the concentration of the fusion protein and cleaved overnight on ice. The cleaved solution was buffer exchanged by ~20-fold

concentration in Amicon® Ultra-15 Ultracel-10 filter units and dilution in 50 mM Tris-HCl pH 8.0, 500 mM NaCl to the original volume. The buffer exchanged solution was subsequently passed over a 5 mL HisTrap FF column and the flow through was collected, pooled, concentrated and subjected to SEC as for the mEGFP-CsLinker protein before concentration and storage in the same manner.

3.5.7.8. Purification of USP2 protease

The cell pellet was resuspended in 50 mL of USP2 lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM Imidazole, 30% glycerol). The cells were lysed, and the protein was purified using the same methodology as the TEV protease, except a one-step elution was completed with USP2 elution buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 200 mM Imidazole, 30% glycerol). The protein was not concentrated following elution.

3.5.7.9. Purification of EPYC1 and EPYC1-GFP

EPYC1 and EPYC1-GFP were purified as for mEGFP-CsLinker except the HisTrap elution was not subjected to SEC. Instead the pooled eluates were incubated with a 1:100 dilution of USP2 protease after addition of glycerol to 3.5% (v/v) and β-mercaptoethanol to 4.7 mM overnight on ice to cleave the His-tagged ubiquitin moiety. The cleaved solution was buffer exchanged and re-purified as for the untagged CsLinker.

3.5.7.10. Purification of Rubisco

Purification of Chlorella and *Chlamydomonas* Rubisco was performed using an identical protocol. Cells were grown to late exponential phase ($1x10^7$ cells mL^{-1} for *Chlamydomonas* and $5x10^7$ cells mL^{-1} for *Chlorella*) in 10 L of TP media in Carboys (2251-0020 – Nalgene) sparged with 3% CO_2 under light at ~50 μmol photons m^{-2} s^{-1} and temperature at ~21 ºC. The cells were harvested by centrifugation at 5,000 *g* for 15 minutes at 4 ºC and resuspended in Rubisco lysis buffer (100 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.5 mM EDTA pH 8.0, 1 mM $MgCl_2$) supplemented with 5 mM DTT added fresh. Lysis was completed by ultrasonication with a processing time of 5 minutes. 3 second pulses were interspersed by 7 second breaks to allow sample cooling. The lysate was clarified by centrifugation at 47,850 *g* for 30

minutes at 4 ºC and subsequent 0.2 µm filtration of the supernatant. To the supernatant ammonium sulphate (89363 – Alfa Aesar) was added to 190 mg mL^{-1} final concentration and the solution was stirred at 4 ºC for 1-2 hours. Precipitate was removed by centrifugation at 47,850 *g* for 10 minutes at 4 ºC. An additional 190 mg mL^{-1} of ammonium sulphate was added to the supernatant and the solution was stirred for a further 1-2 hours. The precipitate was collected as before and resuspended in ~15 mL of Rubisco IEX buffer A (50 mM Tris-HCl, 50 mM NaCl). 1.5 mL of the solution was applied to the top of 12 mL 10-30% (w/v) sucrose gradients that were formed in Rubisco IEX buffer A using a Gradient Master 108 (Biocomp). The gradients were centrifuged at 37,000 rpm in an SW 41 Ti rotor (Beckman Coulter) for 16.5 hours at 4 ºC. The gradients were fractionated and analysed by SDS-PAGE (Fig. 3-5a) and the pure fractions were pooled and diluted 3-fold in Rubisco IEX buffer A. This solution was applied to a 5 mL HiTrap Q XL 5 mL column (Cytiva Life Sciences) equilibrated with Rubisco IEX buffer A. The column was washed with ~10 column volumes of the same buffer until the A_{280} value reached baseline. The bound protein was eluted with a linear gradient to 100% Rubisco IEX buffer B (50 mM Tris-HCl pH 8.0, 500 mM NaCl) over 10 column volumes. The eluted fractions were analysed by SDS-PAGE (Fig. 3-5b) and the purest fractions were pooled and concentrated using the filtration units used previously. The concentrated fraction was subjected to size-exclusion chromatography (SEC) using a HiLoad 16/600 Superdex 200 pg column (28989335 – Cytiva Life Sciences) equilibrated with Rubisco gel filtration buffer (50 mM Tris-HCl pH 8.0, 50 mM NaCl, 5% glycerol (v/v), 1 mM DTT). The eluted fractions were analysed by SDS-PAGE (Fig. 3-5c) and the purest fractions were pooled, concentrated and snap-frozen for storage at -80 ºC.

3.5.7.11. Atto594 labelling of Rubisco

Both Chlorella and *Chlamydomonas* Rubisco were labelled using the same methodology. An Atto 594 protein labelling kit (68616-1KT – Merck) was used largely according to manufacturer's instructions. 0.175 mg of Rubisco was buffer exchanged into sodium bicarbonate buffer solution B (0.2 M Sodium bicarbonate, pH 8.3) using Amicon Ultra-0.5 centrifugal filter units (UFC501024 – Merck). A 0.2 mg vial of Atto 594 dye was resuspended in 20 µL of sodium bicarbonate buffer solution B. Rubisco was labelled using 3.5 µL of the dye solution in a total volume of 100 µL for 2 hours at room temperature

with agitation. The labelled protein was applied to the supplied PD-10 desalting column that had been equilibrated with 5 volumes of PBS and 5 mL of PBS was applied. The labelled protein band (blue) was collected and separated from the free dye (second, pink band). The protein was subsequently concentrated using Amicon Ultra-0.5 filter units and buffer exchanged into 50 mM Tris-HCl pH 8.0, 50 mM NaCl buffer before snap-freezing and storage at -80 ºC.

3.5.7.12. Estimation of protein concentrations

For mEGFP-CsLinker, Rubisco, EPYC1, EPYC1-GFP, TEV protease, USP2 protease and mEGFP-CsLinker-MBP proteins, concentrations were estimated using the calculated molecular weights and extinction coefficients and a NanoDrop ND-1000 spectrophotometer. For untagged CsLinker, a Bradford assay (23236 – ThermoFisher Scientific) using a 4-point BSA calibration curve was completed (Fig. 3-4a) to estimate the protein concentration. The results of the Bradford assay were validated by comparison to a 4 µg load of BSA on an SDS-PAGE gel (Fig. 3-4b).

3.5.8. Attempted transformations of Chlorella (completed with Mihris Naduthodi)

In all cases, the cells that were used in attempted transformations were in the exponential phase (~$1x10^6$-$1x10^7$ cells mL^{-1}) after growth in TAP media under ~50 µmol photons m^{-2} s^{-1} and a temperature of ~21 ºC. For all transformations completed with the Bio-Rad Gene Pulser® II electroporation system, 0.4 cm cuvettes (EP-104 – Cell Projects) were used with an electroporation volume of 250 µL containing the indicated number of cells. The genetic material was mixed with the cells in the electroporation volume and pulsed immediately with the indicated parameters. For all transformations completed with the NEPA21 (Nepa Gene), 0.2 cm cuvettes (EP-102 – Cell Projects) were used with an electroporation volume adjusted such that the measured resistance was ~0.4. In all electroporation-based transformations, the cells were recovered for ~24 hours in TAP supplemented with 40 mM sucrose, then plated on TAP plates supplemented with the relevant antibiotics. For all biolistic transformations completed with the Bio-Rad PDS-1000/He particle bombardment system, 1100 psi rupture disks were used, and the transformation plate was positioned ~9 cm from the rupture disk. 550 nm gold nanoparticles (S550d – Seashell Technologies) were washed with dH_2O and bound with the indicated

genetic material using precipitation buffer as per the manufacturer's instructions. The indicated number of cells were plated TAP-agar plates ~2-4 hours prior to transformation. Following transformation, the cells were recovered on the same plates for ~24 hours prior to transfer to TAP selection plates containing the relevant antibiotics.

For the transformation of protoplasts, cells were protoplasted by incubation overnight in protoplast buffer (1 mg mL^{-1} lysozyme, 1 mg mL^{-1} sulfatase, 0.25 mg mL^{-1} chitinase) in the dark at room temperature. Protoplasted cells were subsequently either electroporated in 0.2 cm cuvettes using a Bio-Rad Pulser XCell transformation system at the indicated parameters or transformed using a HIV-TAT peptide. Following incubation with the HIV-TAT peptide and DNA solution, the cells were plated on TAP plates supplemented with the relevant antibiotics.

3.5.9. Attempted immunofluorescence localisation protocols

For all the separate fixation and permeabilization immunofluorescence methods, TP-grown cells were fixed in 3.7% paraformaldehyde made up in PBS for 30 minutes at room temperature with agitation. The cells were washed twice in PBS prior to treatment with the indicated conditions. All immunofluorescence trials. The one-step fixation and permeabilization methods were completed as indicated. Following permeabilization all samples were washed with PBS and incubated with a 1:100 dilution of the RbcL antibody (YZ8186) for 1 hour at room temperature followed by washing with PBS and incubation with a 1:1000 dilution of the goat anti-rabbit secondary antibody for 45 minutes at room temperature. The cells were imaged as for the *in vitro* droplets except that 10 µL of cells were overlaid with 30 µL of 1% low-melting point agarose made up in TP.

3.5.10. Absolute quantification mass spectrometry (completed with Mihris Naduthodi)

$6x10^6$ cells were harvested from triplicate Chlorella cultures grown to mid-exponential phase ($6x10^6$ cells mL^{-1}) in TP under ambient CO_2 conditions and at ~50 µmol photons m^{-2} s^{-1} and a temperature of ~21 ºC. The harvested cells were boiled in 50 µL of Laemmli buffer for 15 minutes. The protein standards were separately boiled in duplicate in the same volume. The whole sample volumes were

loaded into 10-well 4–20% Mini-PROTEAN® TGX™ gels and run for ~10 minutes at 120 V until the entirety of the samples had entered the gel. The samples and standards were run on separate gels. The gels were stained with InstantBlue and the sample bands were excised and submitted for mass spectrometry analysis. Samples were trypsin-digested from the gel slices, post reduced with 1,4-Dithioerythritol (DTE) and alkylated with iodoacetamide prior to separation of the extracted peptides by ultra-performance liquid chromatography (UPLC) with a 25 cm PepMap column (ThermoFisher Scientific) and 1 hour acquisition by data-dependent acquisition (DDA) using an Orbitrap Fusion Lumos Tribrid (ThermoFisher Scientific). High resolution MS1 and top speed 1s cycle linear ion trap (LIT) MS2 post higher-energy C-trap dissociation (HCD) fragmentation were used during collection. The biological triplicate whole cell samples were acquired as technical duplicates (two injections). Liquid chromatography mass spectrometry (LC-MS) chromatograms were aligned using Progenesis QI (Waters) and the MS2 spectra were exported and searched using Mascot (Matrix Science) with a 1% false discovery rate. Peptide matches were reimported and mapped to MS1 peak intensity using Progenesis QI, integrated and summed at protein level. The calibration curves from the protein standards were used to calculate protein abundance in the biological samples (Fig. 3-5). All proteomics data were deposited in MassIVE, with ProteomeXchange identifier PXD044179.

3.5.11. Antibodies

Affinity purified rabbit primary antibodies were raised to regions specific to CsLinker (C-Ahx-PTPVSNSGVRSAMSSG-amide) and Rubisco (C-Ahx-EVWKEIKFEFETIDTL-cooh) by YenZym antibodies LLC in duplicate. For this study YenZym antibodies YZ8175 (CsLinker) and YZ8186 (Rubisco) were used.

3.5.12. Western blot for relative abundance of CsLinker and Chlorella Rubisco

$1x10^8$ cells were harvested from Chlorella cultures grown in 5 mL of TP in 6-well plates and at ~50 µmol photons m^{-2} s^{-1} and a temperature of ~21 ºC either under ambient CO_2 conditions or in gallon Ziploc bags (351126 – Ziploc) supplemented with 3% CO_2. Harvested cells were resuspended in 50 µL of 1x Laemmli buffer and boiled for 15 minutes prior to loading.

3.5.13. Co-immunoprecipitation mass spectrometry (co-IP) (completed with Mihris Naduthodi)

$1x10^9$ cells were harvested from a Chlorella culture grown to mid-exponential phase ($3x10^6$ cells mL^{-1}) in TP under ambient CO_2 conditions and at ~50 µmol photons m^{-2} s^{-1} and a temperature of ~21 ºC. Cells were lysed by ultrasonication as before, in 10 mL of IP buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl, 0.1 mM EDTA, 12.5% glycerol, 5 mM DTT) with 1 cOmplete EDTA-free protease inhibitor tablet. The lysate was clarified by centrifugation at 20,000 *g* for 30 minutes at 4 ºC. All subsequent steps were also completed at 4 ºC. For each replicate, 200 µL of Dynabeads™ Protein A in 2 mL Eppendorf tubes were washed twice with 500 µL of IP buffer by mixing and removal using a magnrack. 32 µg of each of the primary antibodies (YZ8175 and Y78186) was dissolved in 500 µL of IP buffer and incubated with the beads for 2 hours. The control antibody used for comparison of enrichment was previously raised to *Chlamydomonas* BST2 (Mukherjee *et al.,* 2019). The loaded beads were washed twice with IP buffer and blocked with 500 µL of a 2 mg mL^{-1} BSA solution in IP buffer for 1 hour. The blocked beads were washed with IP buffer and incubated with 1 mL of lysate for 3 hours, followed by 3 washes and resuspension in 500 µL IP buffer. 250 µL of beads were transferred to a new tube and the IP buffer was removed. The beads were boiled in 60 µL of Laemmli buffer and the supernatant was loaded onto 10-well 4–20% Mini-PROTEAN® TGX™ gels. The samples were run, stained, excised, and digested as for the absolute quantification mass spectrometry approaches. UPLC separation was performed by 60SDP EvoSep One method (EvoSep). Acquisition was performed using a parallel accumulation-serial fragmentation data-independent acquisition (PASEF-DIA) approach with a TimsTOF HT mass spectrometer (Bruker Instruments). Data were searched using DIA neural network (DIA-NN) (Demichev *et al.,* 2020) and filtered to 1% FDR and a minimum of two peptides. Statistical comparison was completed using FragPipe-Analyst (http://fragpipe-analyst.nesvilab.org/).

3.5.14. Immunoelectron microscopy

Cells were harvested from a Chlorella culture grown to mid-exponential phase ($3x10^6$ cells mL^{-1}) in TP under ambient CO_2 conditions and at ~50 µmol photons m^{-2} s^{-1} and a temperature of ~21 ºC. The harvested cells were fixed in 1% Glutaraldehyde, 2% Formaldehyde in 0.1 M Sodium cacodylate (pH 7.4) for 1 hour with agitation at room temperature, then washed twice in the same buffer. The fixed cells

were resuspended in 2% low melting point agarose (16520050 – Invitrogen). Cubes of cells in agar blocks were dehydrated through a graded ethanol series at progressive low temperature (10%, 20% and 30% at ambient temperature followed by 50% at 4C and 70%, 90% and 100% at -20C), based on previous methods (Vandenbosch, 1990). Dehydrated specimens were infiltrated with medium hardness LRWhite resin (AGR1281 – Agar Scientific) containing 0.5% Benzoin methyl ether for 48 hours with rotation. The infiltrated samples were polymerised in gelatin capsules using UV light for 24 hours at -20 ºC, followed by a further 24 hours at -10 ºC. 70 nm sections were cut using a Leica UCT7 ultramicrotome with a Diatome knife and mounted on Nickel grids. For immunolabelling, the grids were blocked with 3% BSA in phosphate buffered saline (PBS) for 30 minutes at room temperature. The blocked grids were incubated with 1:700 dilutions of the primary antibodies (YZ8175 or YZ8186) in 1% BSA-PBS for 1 hour at 30 ºC in a humidity chamber. Grids were washed 3 times in PBS and incubated with a 1 :40 dilution of the secondary Goat anti-rabbit IgG 10 nm gold conjugate antibody (G7402 – Merck) in 1% BSA-PBS. The grids were finally washed 3 times in PBS and 3 times in dH_2O. The primed grids were then incubated in 2% uranyl acetate (w/v) for 5 minutes followed by 5 minutes in lead citrate as per previous methods (Reynolds, 1963). Imaging was completed by Dr Clare Steele-King using an FEI Tecnai T12 BioTWIN TEM operating at 120 kV with a Ceta CCD camera.

3.5.15. Droplet sedimentation assays

All droplet sedimentation assays in this study were completed using the same following methodology. The reaction volume was 5 µL in all cases, to which the reaction buffer (50 mM Tris-HCl pH 8.0, 50 mM NaCl) was added first followed by a high salt buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl) where appropriate (Fig. 3-13). Of the protein components, Rubisco was consistently added prior to the linker protein, after which the samples were immediately vortexed for ~2 seconds and incubated for ~15 minutes at room temperature. The samples were subsequently centrifuged at 17,900 *g* for 5 minutes at room temperature to pellet the droplet phase. The supernatant was removed and incubated with Laemmli buffer at a 1x final concentration. To the empty tube, 6 µL of 1x Laemmli buffer made up in reaction buffer was added to dissolve the pelleted phase. Both the supernatant and pellet fractions were

boiled for 5 minutes prior to loading on SDS-PAGE gels as described previously. Densitometric analysis of the supernatant and pelleted fractions was completed as described previously.

3.5.16. General microscopy methods

3.5.16.1. In vitro droplet formation

For all brightfield and confocal microscopy imaging of *in vitro* droplet solutions in this study, iBidi µ-Slide 15-well 3D slides (81506 – iBidi) were used. The formation of droplets was completed in the same stepwise fashion as used for the droplet sedimentation assays. The reaction volume was either 5 or 10 µL.

3.5.16.2. Brightfield and confocal microscopy

For all brightfield and non-FRAP confocal microscopy, a Zeiss LSM 880 Airyscan microscope was used in confocal mode without use of the Airyscan module. In all cases a 63x 1.4 numerical aperture (NA) Plan-Apo oil-immersion lens (Carl Zeiss) was used with Carl Zeiss™ Immersol™ immersion oil (518F – Carl Zeiss). For brightfield microscopy, the 488 nm excitation laser was used at 1% power with a pinhole of ~80 and the transmitted photomultiplier tube (T-PMT) detector was activated. The master gain was set to ~250, with a digital gain of 1.0. No offset was applied. For confocal fluorescence microscopy either the 488 nm or 561 nm laser were used at 1% power for visualisation of EGFP or Atto594 respectively. The master and digital gain settings were varied between imaging conditions but were ~750 and 1.5 respectively in most cases. A 488/561 main beam splitter (MBS) was used and the mirror was inserted. For 2-colour experiments, separate channels were used for the 488 nm and 561 nm excitation lasers and were imaged sequentially.

3.5.16.3. Fluorescence recovery after photobleaching (FRAP) microscopy

For all FRAP experiments, a Zeiss LSM980 MP Airyscan 2 microscope with a Definite Focus 3 module was used without use of the Airyscan 2 module. In all cases a 63x 1.4 numerical aperture (NA) Plan-Apo oil-immersion lens (Carl Zeiss) was used with Carl Zeiss™ Immersol™ immersion oil (518F – Carl Zeiss). Droplets were formed in 10 µL volumes as previously described, except the aqueous

solution was overlaid with 30 µL of iBidi Anti-Evaporation oil (50051 – iBidi) to prevent evaporation of the solution during imaging. Both half- and whole-FRAP experiments were completed with 15 pre-bleach frames. For mEGFP, 1 bleach cycle at 100% 488 nm laser intensity was used. For Atto594, 5 bleach cycles using the 488 nm and 561 nm lasers at 100% intensity were used for bleaching. The bleach regions were primarily drawn horizontally, parallel to the laser scanning direction, to allow more accurate positional bleaching. The bleach parameters resulted in largely consistent bleaching to ~60-75% of the mean gray intensity compared to the pre-bleach intensity. Between each of the subsequent imaging frames, which were consistent with the settings used for non-FRAP confocal microscopy images, the Definite Focus 3 module was used to retain the Z-focus of the imaging platform. Zen Blue was used for microscope control.

3.5.16.4. Fluorescence recovery after photobleaching (FRAP) image analysis

All FRAP images were initially analysed using Fiji. The brightfield (T-PMT) channel from the imported images was used to translationally stabilise the image using the Image Stabilizer plugin with 4 pyramid levels, a template update coefficient of 0.99 in translation mode and otherwise default settings. The resulting translation log was subsequently applied to the fluorescence channel prior to analysis. At least 2 regions of interest for the background (non-droplet) and reference (unbleached droplets) regions were drawn using the Fiji selection tool in each image. Using the multi-measure function, the mean gray intensity in each of the ROIs (including the bleached ROIs imported from Zen Blue) was measured and imported to Microsoft Excel.

Initially, the intensity of the ROIs in each of the images was plotted and any ROIs that showed fluorescence spikes or drops were excluded from analysis. Subsequently an average background intensity was taken from the multiple background ROIs and the average value was subtracted from all of the raw bleached and unbleached reference ROI values. To correct for photobleaching the photobleaching rate of the unbleached reference ROIs was calculated as follows. The 15 pre-bleach images in the reference unbleached ROIs were plotted and any erroneous data points were excluded before an average pre-bleach value from the remaining values was calculated. The relative intensity at

each of the subsequent timepoints in the unbleached reference ROIs was calculated relative to the average pre-bleach value. An average of the relative bleach values in each of the reference unbleached ROIs was then taken. The same correction relative to the average pre-bleach value was completed with each of the bleached ROIs, to which photobleach correction was applied by dividing the resulting value by the average relative reference unbleached value at the same timepoint. Finally, the data was full-scale normalised by subtracting the minima in each of the bleached ROIs from each of the values and subsequently dividing each value by 1 – the minima.

For each of the resulting FRAP curves, an exponential model ($y(t) = A * (1-e^{-kt})$, where A is the plateau, k is fitted and t is post-bleach time), was fitted to the post-bleach data using the SciPy curve_fit module in python3. The fit was used to derive the $T_{0.5}$ value for each of the FRAP curves, from which an average $T_{0.5}$ value with error was calculated in GraphPad Prism 9. For the average FRAP curve plots, the mean, standard error of the mean (S.E.M.) and standard deviation (S.D.) were calculated for each of the timepoints from the individual curves and plotted using python3.

Table 3-17: Plasmids used in Chapter 3.

Name	Description	Use	Origin
pLM017	EPYC1-Venus (Paromomycin)	Attempted Chlorella transformation	(Mackinder *et al.,* 2016)
pLM021	CrRbcS2-Venus (Paromomycin)	Attempted Chlorella transformation	(Mackinder *et al.,* 2016)
pLM031	CrRbcS2-mCherry (Hygromycin)	Attempted Chlorella transformation	(Mackinder *et al.,* 2016)
pLM630	*Chlamydomonas* Paromomycin cassette	Attempted Chlorella transformation	(Crozet *et al.,* 2018)
pLM631	*Chlamydomonas* Hygromycin cassette	Attempted Chlorella transformation	(Crozet *et al.,* 2018)
pICH47732	L1 Golden Gate acceptor vector	Cloning of pLM845-849	(Weber *et al.,* 2011)
pCM0-073	L0 Hygromycin resistance vector	Cloning of pLM845	(Crozet *et al.,* 2018)
pCM0-074	L0 Paromomycin resistance vector	Cloning of pLM847 and pLM848	(Crozet *et al.,* 2018)
pCM0-077	L0 Zeocin resistance vector	Cloning of pLM849	(Crozet *et al.,* 2018)
pLM845	Chlorella Hygromycin cassette (pPSAD/tPSAD)	Attempted Chlorella transformation	This study
pLM847	Chlorella Paromomycin cassette (pPSAD/tPSAD)	Attempted Chlorella transformation	This study
pLM848	Chlorella Paromomycin cassette (pRbcS2/tPSAD)	Attempted Chlorella transformation	This study
pLM849	Chlorella Zeocin cassette (pPSAD/tPSAD)	Attempted Chlorella transformation	This study
pCAMBIA1301	Plant Hygromycin cassette (pCaMV35s)	Attempted Chlorella transformation	(Sharma *et al.,* 2021)
pSK397	Chlorella Geneticin cassette (pHSP70)	Attempted Chlorella transformation	(Kumar *et al.,* 2018)
pLM672	CsLinker *E. coli*-optimised part 1	Assembly of *E. coli* optimised CsLinker for protein expression	TWIST bioscience
pLM673	CsLinker *E. coli*-optimised part 2	Assembly of *E. coli* optimised CsLinker for protein expression	TWIST bioscience
pLM674	CsLinker *E. coli*-optimised part 3	Assembly of *E. coli* optimised CsLinker for protein expression	TWIST bioscience
pUAP1	L0 acceptor plasmid	Assembly of *E. coli* optimised CsLinker for protein expression	(Patron *et al.,* 2015)
pLM701	L0 CsLinker *E. coli*-optimised	Amplification of CsLinker for assembly in expression vectors	This study
1-GFP	N-terminal 6xHis-mEGFP-TEV LIC vector	Cloning of CsLinker for mEGFP-tagging and purification	Addgene #29663
pLM712	6xHis-mEGFP-TEV-CsLinker	Purification of mEGFP-CsLinker and assembly of pLM872	This study
2-Cc-T	C-terminal MBP-6xHis LIC vector	Cloning of CsLinker for untagged purification	Addgene #37237

pLM872	6xHis-mEGFP-TEV-CsLinker-TEV-MBP-6xHis	Purification of untagged CsLinker	This study
pRK793	MBP-6xHis-TEV protease	Purification of TEV protease	(Kapust *et al.,* 2001)
pHueCrEPYC1	6xHis-Ubiqutin-EPYC1	Purification of untagged EPYC1	(Wunder *et al.,* 2018)
pHueCrEPYC1-GFP	6xHis-Ubiqutin-EPYC1-GFP	Purification of GFP-tagged EPYC1	(Wunder *et al.,* 2018)
pBADGroEL/S	GroEL/ES chaperone	Co-expression of GroEL/S chaperone with EPYC1 proteins	(Wunder *et al.,* 2018)
pHUSP2-p2cc	USP2 protease	Purification of USP2 protease	(Wunder *et al.,* 2018)

4. CSLINKER INTERACTS WITH THE LARGE SUBUNIT OF RUBISCO

4.1. Introduction

In the previous chapter it was demonstrated that CsLinker is a *bona fide* pyrenoid linker protein, and that the interaction between CsLinker and Chlorella Rubisco gives rise to largely analogous properties to those observed for EPYC1 and *Chlamydomonas* Rubisco. However, key differences in the behaviour of the *in vitro* reconstitutions, as well as the sequence disparate nature of the two sequences suggested differences in the *modus operandi* of CsLinker and EPYC1 that warrant further investigation. To this end, this chapter aimed to biochemically and structurally characterise the interaction between CsLinker and Chlorella Rubisco in an attempt to explain these differences by comparison with the previously well-characterised interaction of EPYC1 and Chlamydomonas Rubisco (Atkinson *et al.,* 2019; He *et al.,* 2020; Wunder *et al.,* 2018).

4.2. Biochemical characterisation of the interaction between CsLinker and CsRubisco

In the two previously structurally characterised eukaryotic Rubisco-linker interactions (He *et al.,* 2020; Oh *et al.,* 2023), and the prokaryotic α-carboxysome Rubisco-linker interaction (Oltrogge *et al.,* 2020), all linkers have been shown to interact with Rubisco using helical motifs (Fig. 4-1). It was therefore hypothesised that the predicted helices of CsLinker would account for the interaction interface with CsRubisco. Despite the shared structural features, previous approaches to biochemically characterise linker fragment interactions with their cognate Rubiscos have varied. In both studies completed with EPYC1 from Chlamydomonas, the fragments used were not representative of native full repeat regions of EPYC1 (He *et al.,* 2020; Wunder *et al.,* 2018), in contrast to studies of PYCO1 from Phaeodactylum, where a fragment representing a full repeat region in addition to a C-terminal extension were used (Oh *et al.,* 2023). The latter approach using a longer PYCO1 repeat fragment afforded the authors additional context in their structural study, as an additional binding region common to other pyrenoid proteins was also present in the cryo-EM density map (Fig. 4-1b, right). Likewise in the α-carboxysomal linker study, fragments of the Csos2 N-terminal domain representing full repeat regions were used for biochemical studies, though the structure of the complex was obtained by fusion of a partial consensus repeat region (CsoS2-N*) to the Rubisco large subunit prior to crystallisation (Oltrogge *et al.,* 2020). Based on these characterisations it was concluded that working with full repeat regions of sequence native to the CsLinker were likely to provide the greatest depth of data in the biochemical and structural studies of CsLinker interaction with Chlorella Rubisco.

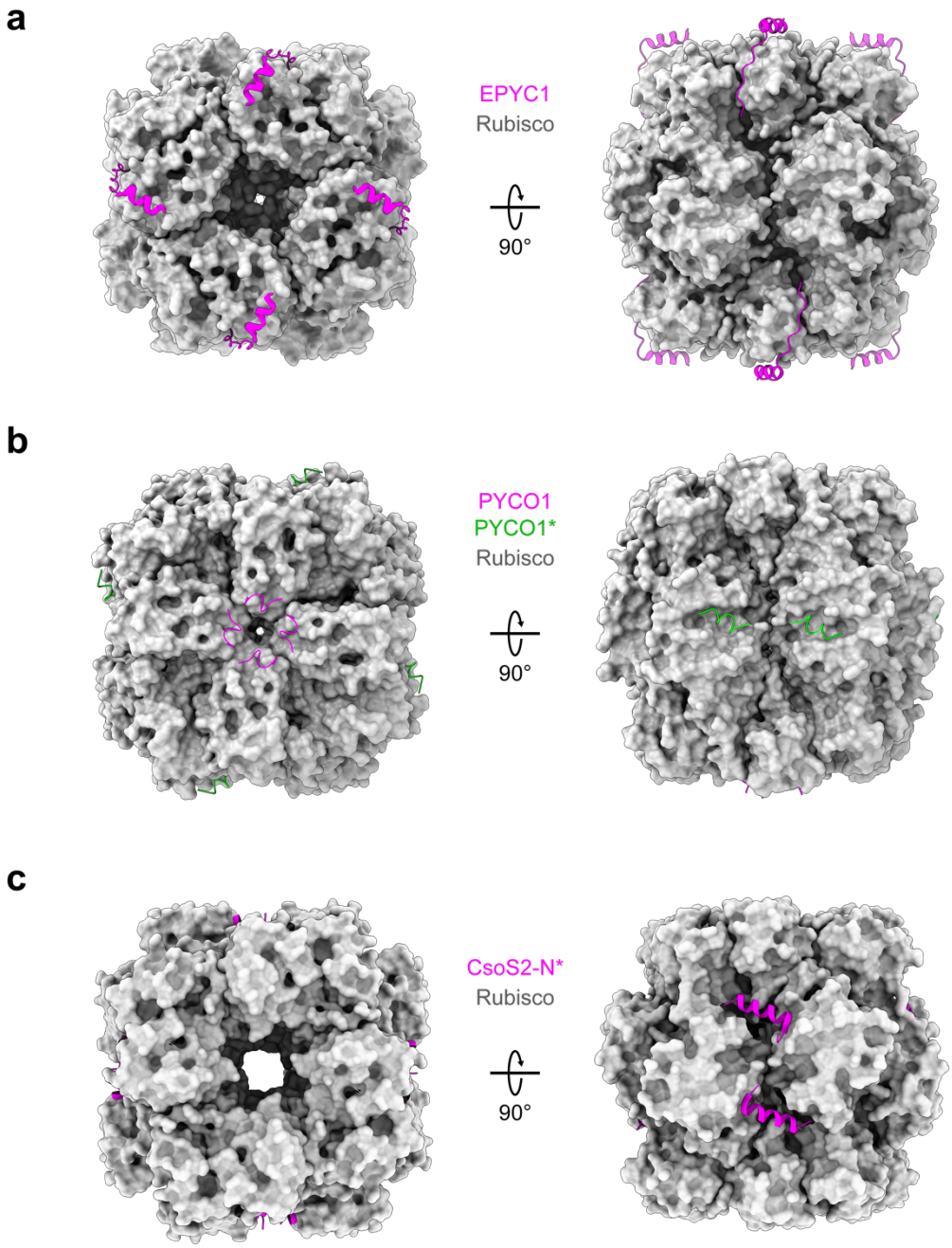


Figure 4-1 | Structurally characterised linker-Rubisco interactions.
(a) Top (left) and side (right) views of EPYC1 interaction with Chlamydomonas Rubisco. A surface representation of the Rubisco enzyme is shown in gray, with the EPYC1 fragment shown in magenta. (b) Top and side views of PYCO1 interaction with Phaeodactylum Rubisco. The additional PYCO1 density (PYCO1*) afforded by the fragment used is shown in green. (c) Top and side views of Csos2-N* interaction with *Halothiobacillus* Rubisco.

4.2.1. *Choice of CsLinker fragments for characterisation*

Based on the conclusions of the above, single and double helix fragments representing full repeat regions of CsLinker centred on the third (α3), and third and fourth (α3-α4) predicted helices were generated for biochemical characterisation. Helices three and four were chosen from the CsLinker sequence as they were deemed to be the most representative repeat regions within the full length sequence with respect to both the predicted helices and disordered spacer regions, based on sequence alignment of the repeats relative to the consensus sequence (Fig. 4-2). Both the α3 and α3-α4 fragments were purified to a high level using the same expression and purification strategy used previously for the full length CsLinker protein (Fig. 4-3).

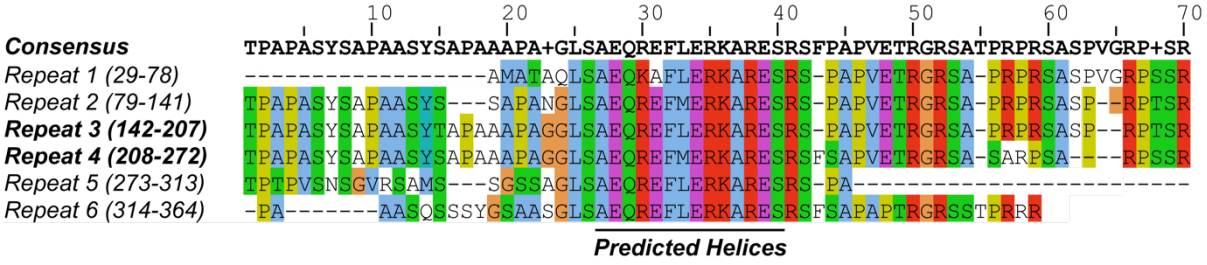


Figure 4-2 | Comparison of repeat similarity in CsLinker used to determine fragment sequences.
The six repeat regions of CsLinker aligned using MAFFT, shown relative to the consensus repeat sequence (bold). The regions used for the α3 and α3-α4 fragments (repeat 3 and repeats 3 and 4 respectively) are shown bolded. The predicted helices from AlphaFold2 are indicated. Residues are coloured by property according to the Clustal X colour scheme. Residue numbering is relative to the position in the full length CsLinker protein (left) and relative to the consensus repeat (top).

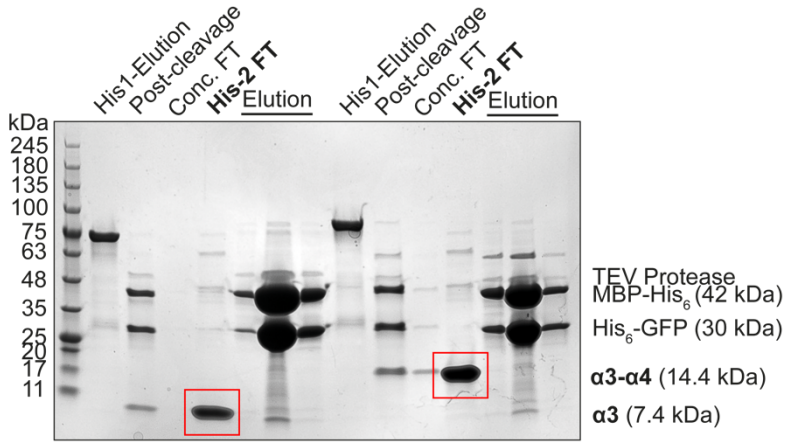


Figure 4-3 | Purification of α3 and α3-α4 fragments.
SDS-PAGE of samples taken during His1 and His2 purification of the fragments. The samples which were subjected to SEC following His2 are indicated by red boxes. FT = flow through.

4.2.2. Native PAGE characterisation of fragment-CsRubisco interaction

Due to the size, charge, and stability of the Rubisco holoenzyme and the high isoelectric point of the interacting linker proteins, native gel electrophoresis has been readily employed to assay for interaction between Rubiscos and fragments of their respective linker proteins (Oh *et al.,* 2023; Oltrogge *et al.,* 2020; Wunder *et al.,* 2018), and was used previously to characterise interaction of the full length CsLinker protein with Chlorella Rubisco (Fig. 3-9). The same approach was initially applied to the α3 fragment and CsRubisco, over a wide concentration range, guided by the low affinity interaction (K_D ~3 mM) observed for the single helix EPYC1 fragment with Chlamydomonas Rubisco (He *et al.,* 2020). Upon addition of increasing α3 fragment at a fixed concentration of Rubisco (1 μM), an apparent shift in the Rubisco band was readily observed (Fig. 4-4a). Under the same conditions which the band shift was observed, there was no apparent condensate or aggregate formation in solution as assessed by sedimentation and microscopy assays (Fig. 4-4b). This evidence indicated the concentration-dependent band shift observed by native PAGE was due to *bona fide* interaction between α3 and CsRubisco in the experiment. Clearly the observed band shift was unfit for quantification, likely due to the transiency of the interaction causing band smearing, but does appear to be approaching a qualitative half-saturation of the higher band at the 64 μM fragment concentration. The lack of phase separation observed with the single helix fragment is consistent with the multivalent requirement of complex LLPS (Harmon *et al.,* 2017), and observations with PYCO1-Rubisco interactions (Oh *et al.,* 2023).

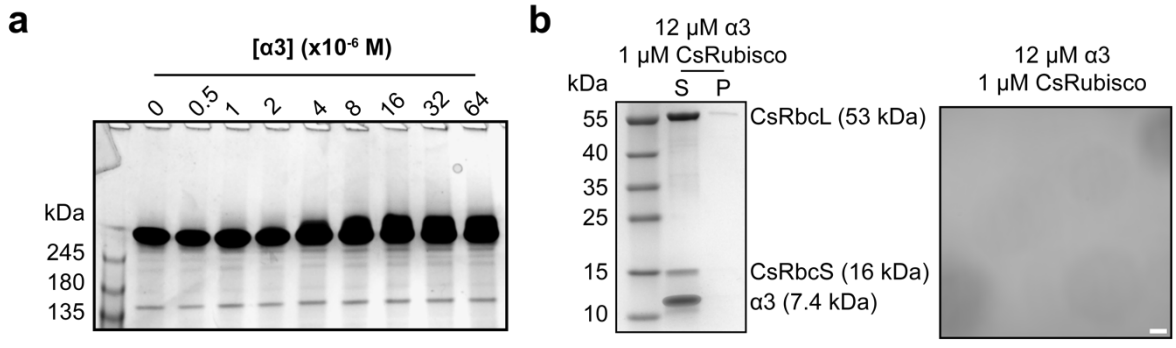


Figure 4-4 | Interaction between α3 and CsRubisco.
(a) Native PAGE band shift assay completed with CsRubisco fixed at 1 μM and increasing concentration of the single helix α3 fragment. (b) Droplet sedimentation assay (left) and brightfield microscopy (right) of α3-CsRubisco solution under the same conditions, at the indicated α3 fragment concentration (12 μM).

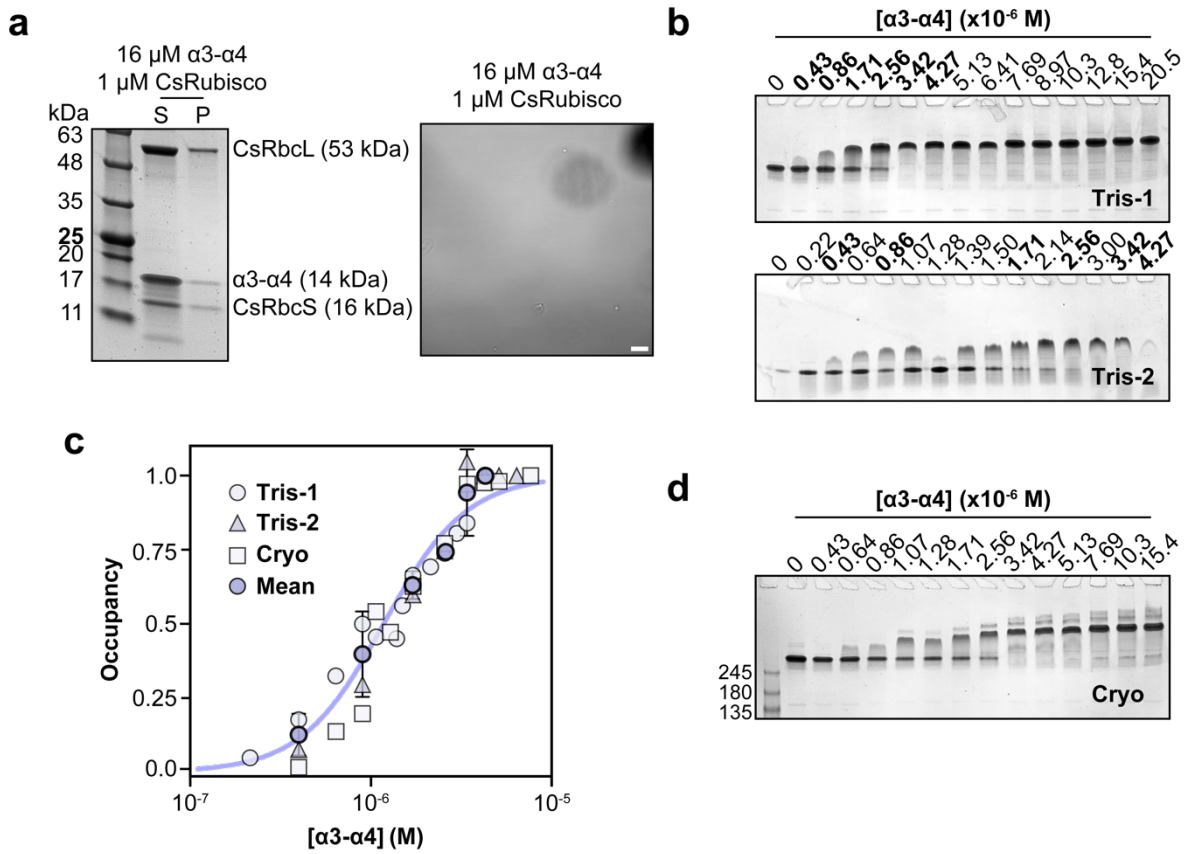


Figure 4-5 | Interaction between α3-α4 and CsRubisco.
(a) Sedimentation (left) and brightfield microscopy (right) completed under the same conditions as in Fig. 4-4b, at the indicated α3-α4 fragment concentration (16 μM). (b) Native PAGE band shift assay completed with CsRubisco fixed at 1 μM and increasing concentration of the double helix α3-α4 fragment in Tris buffer (50 mM Tris-HCl (pH8.0), 50 mM NaCl). (c) Quantification of the occupancy of the higher order α3-α4-CsRubisco complex from Fig. 4-5b and Fig. 4-5d derived from the inverse occupancy of the monomeric Rubisco band determined by densitometric analysis. The 'Mean' values are taken from the replicates in Fig. 4-5b highlighted in bold. (d) Native PAGE band shift assay completed as in Fig. 4-5b, except the buffer the proteins were incubated in prior to analysis was different (200 mM sorbitol, 50 mM HEPES, 50 mM KOAc, 2 mM $Mg(OAc)_2 \cdot 4H_2O$ and 1 mM $CaCl_2$ at pH 6.8).

Considering the multivalent requirement for LLPS, and phase separation of Rubisco with a 2 helix fragment of EPYC1 (Wunder *et al.,* 2018), it was expected that the 2 helix α3-α4 fragment would also induce phase separation of CsRubisco. By contrast, no phase separation or aggregation was observed by droplet sedimentation or microscopy assays (Fig. 4-5a), even at Rubisco and putative CsLinker binding site concentrations previously demonstrated to phase separate CsRubisco with the full length protein. When analysed by native PAGE, the α3-α4 fragment readily demonstrated a concentration-dependent band shift of CsRubisco into a stable higher order complex that had an apparent $K_{0.5}$ of ~1.3 μM (CI_{95}: 1.1-2.0 μM) in both the buffer used for condensate experiments ('Tris' in Fig. 4-5b,c), and in a buffer previously used for biochemical studies with the single helix fragment of EPYC1 ('Cryo' in

Fig. 4-5c,d). Notably the size of the band shift observed with α3-α4, relative to that of the α3, was increased consistent with the formation of a higher molecular weight, isoelectric point and/or stability complex between the proteins.

4.2.3. SPR characterisation of fragment-CsRubisco interaction

Comparison of the native PAGE band shift assay results using the α3 and α3-α4 fragments indicated that addition of the second repeat region in the α3-α4 fragment led to an increase in affinity for CsRubisco, consistent with the addition of a second equivalent putative binding region. To more accurately determine this relative affinity increase, and allow comparison with previously determined EPYC1 fragment affinity for Chlamydomonas Rubisco (He *et al.,* 2020), surface plasmon resonance (SPR) was completed with the same fragments. To allow direct comparison, the experiments were completed almost identically to those performed in He *et al.* (2020), with minor modifications (see methods). Using amine-coupled CsRubisco as bait on the sensor chip, and the α3 and α3-α4 fragments as prey in separate triplicate experiments (Fig. 4-6a), affinities for CsRubisco were determined to be 103 μM [CI_{95}: 79–135 μM] for α3 and 1.21 μM [CI_{95}: 1.10–1.34 μM] for α3-α4 (Fig. 4-6b). These values were largely consistent with the qualitative results for α3 and semi-quantitative results for α3-α4 determined from the native PAGE experiments (Fig. 4-4a and Fig. 4-5). Importantly, the ~100-fold higher affinity of α3-α4 for CsRubisco over α3 was replicated and was consistent with cooperative binding of the second putative binding region to the same Rubisco rather than to a second Rubisco in solution, supporting the lack of LLPS previously observed (Fig. 4-5a). Interestingly, the single helix fragment (α3) showed ~30-fold higher affinity for CsRubisco than the previously reported affinity of a single helix fragment of EPYC1 for Chlamydomonas Rubisco (K_D ~3 mM) (He *et al.,* 2020), but was largely consistent with affinity range estimated for a single helix fragment of CsoS2 binding to *Halothiobacillus neapolitanus* Rubisco (K_D ~100-1000 μM) (Oltrogge *et al.,* 2020). The α3-α4 affinity was also largely in agreement with the reported value for a two helix fragment of CsoS2 at 60 mM NaCl (K_D ~0.5 μM), which also did not demonstrate functional phase separation of Rubisco *in vivo*. Importantly, the experiments performed in this work, and that of Oltrogge *et al.,* (2020) were completed with full length repeats, in contrast to the partial repeat used in He *et al.,* (2020). Hence, it is possible

an increased affinity for Chlamydomonas Rubisco could be observed if regions representing total repeats were observed. Equally, assessment of a native 2 repeat fragment of EPYC1 may demonstrate more similar properties to those observed here for CsLinker, in contrast to the previous results with a non-native EPYC1 2 repeat fragment (Wunder *et al.,* 2018). The affinity of PYCO1 binding regions to Rubisco was not previously characterised, so no comparison could be made. Taken together the SPR results indicated the CsLinker fragments demonstrated affinity for Rubisco in the order observed for other Rubisco linker fragment interactions, but that there were no differences sufficient to explain the lack of emergent properties of 2 helix fragments and their interaction with the respective Rubiscos. As such, further work to structurally characterise the *modus operandi* was pursued.

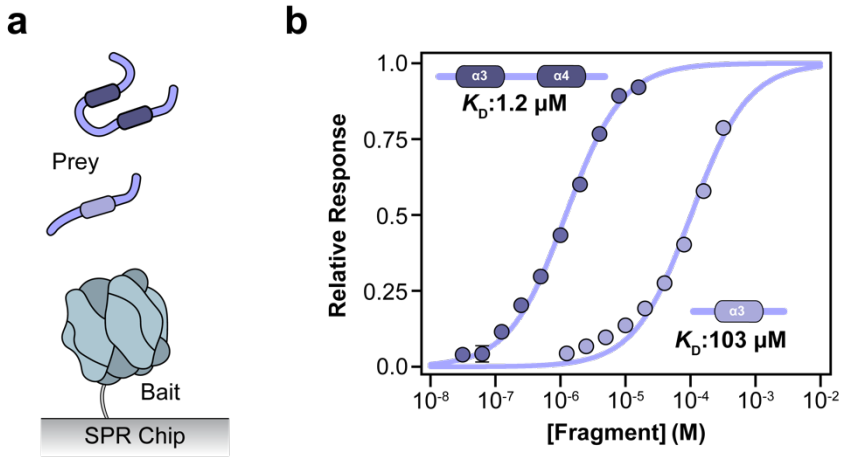


Figure 4-6 | SPR analysis of α3 and α3-α4 interaction with CsRubisco.
(a) Schematic representation of SPR experiment format of α3 or α3-α4 fragments binding immobilised CsRubisco (left). (b) SPR response normalised to Y_{max} value obtained from fit of raw data (right); n=3, error bars = S.D.

4.3. Structural characterisation of CsLinker-CsRubisco interaction

Having validated the interaction of the repeat regions of the CsLinker protein with CsRubisco, it was next desirable to understand the molecular details that underpinned this interaction. Cryogenic Electron Microscopy (cryo-EM) has rapidly evolved over the last 10 years (Callaway, 2020), and as such single-particle cryo-EM has been successfully utilised to determine the high resolution 3D structures of numerous Rubiscos with (Flecken *et al.,* 2020; He *et al.,* 2020; Huang *et al.,* 2020; Li *et al.,* 2022; Oh *et al.,* 2023; Wang *et al.,* 2019; Xia *et al.,* 2020), and without (Bhat *et al.,* 2017; Evans *et al.,* 2023; He *et al.,* 2020), interacting partners. Rubisco is particularly well suited to cryo-EM given its high degree of internal symmetry and large overall size and was therefore selected as the primary technique to determine the 3D structure of the CsLinker-CsRubisco complex.

4.3.1. Grid preparation for cryo-EM

Inherently, single particle cryo-EM requires dispersion of the sample across the grid such that picking of individual particles from the micrograph can be achieved. As it was previously demonstrated that full length CsLinker can phase separate CsRubisco even at low global concentrations (Fig. 3-12), it was implied that either α3 or α3-α4 fragments should be used for grid preparation to allow dispersion of the putative complex on the grids. To allow the best chance for capturing additional interaction outside of the predicted helical interaction interface, it was decided that the α3-α4 fragment representing two complete repeat regions should be used, though grids were prepared with the α3 fragment in case the α3-α4 fragment appeared unsuitable. Accordingly, cryo-EM grids were prepared at a concentration of CsRubisco that would allow dispersion of the Rubisco particles across the grid (0.5 μM) with a concentration of α3-α4 that was previously shown to saturate CsRubisco in the α3-α4-CsRubisco complex by native PAGE and SPR in the same buffer (16 μM) (Fig. 4-5d and Fig. 4-6b). The concentration of α3-α4 was such that the putative CsLinker binding site concentration was comparable to that empirically determined *in vivo* (32 μM here, 27 μM *in vivo* – Fig. 3-5c) The two components were complexed in the same buffer used for both the previous SPR experiments, and used by He *et al.* (2020) for the determination of the EPYC1-Chlamydomonas Rubisco complex. Under these conditions,

Rubiscos were observed to be well-distributed across the grid with an appropriate ice thickness for data collection (Fig. 4-7).

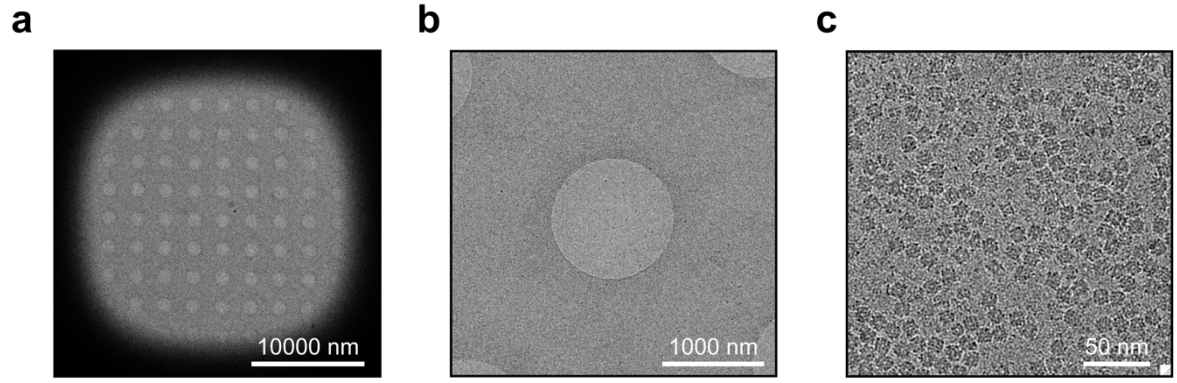


Figure 4-7 | Grid preparation used for data collection of α3-α4-CsRubisco complex.
(a) Representative micrograph of a grid square with the appropriate thickness of ice used for data collection. (b) Zoomed view of a foil hole within the same grid square showing the lack of aggregation or contamination on the grid. (c) Micrograph following motion correction showing dispersion of Rubisco particles in the foil hole.

4.3.2. Data collection

In contrast to most previously published Rubisco complex structure datasets that were collected using 300 kV electron microscopes, the dataset for the α3-α4-CsRubisco complex was collected with a 200 kV Glacios microscope at the University of York. EPU, operated in AFIS mode was used for data collection over an ~14 hour data collection period in which 4,447 micrographs were collected.

4.3.3. Particle picking, 2D classification and Initial Modelling

RELION3.1 was used for all data processing (Zivanov *et al.,* 2018). Following motion correction with MotionCor1.2 (Zheng *et al.,* 2017), and Contrast Transfer Function (CTF) estimation with CTFFIND4.1 (Rohou & Grigorieff, 2015), 465 particles were manually picked from a range of micrographs across the dataset. The manually picked particles were extracted with 2x binning and subjected to 2D classification. Representative top and side view classes of Rubisco were then used as 2D references in auto-picking (Fig. 4-8a), which resulted in 237,035 particles being picked with high accuracy (Fig. 4-8b), as confirmed by the high percentage of particles retained in the subsequently selected 2D classes (Fig. 4-8c) (224,593 particles; 94.8% of auto-picked particles). A *de novo* 3D model was generated from

the selected 2D classes in which the general shape and low resolution features of Rubisco were observed (Fig 4-8d). Notably, regions of additional density, not accounted for by the holoenzyme model of CsRubisco, were tentatively observed at the equatorial region of Rubisco at this stage.

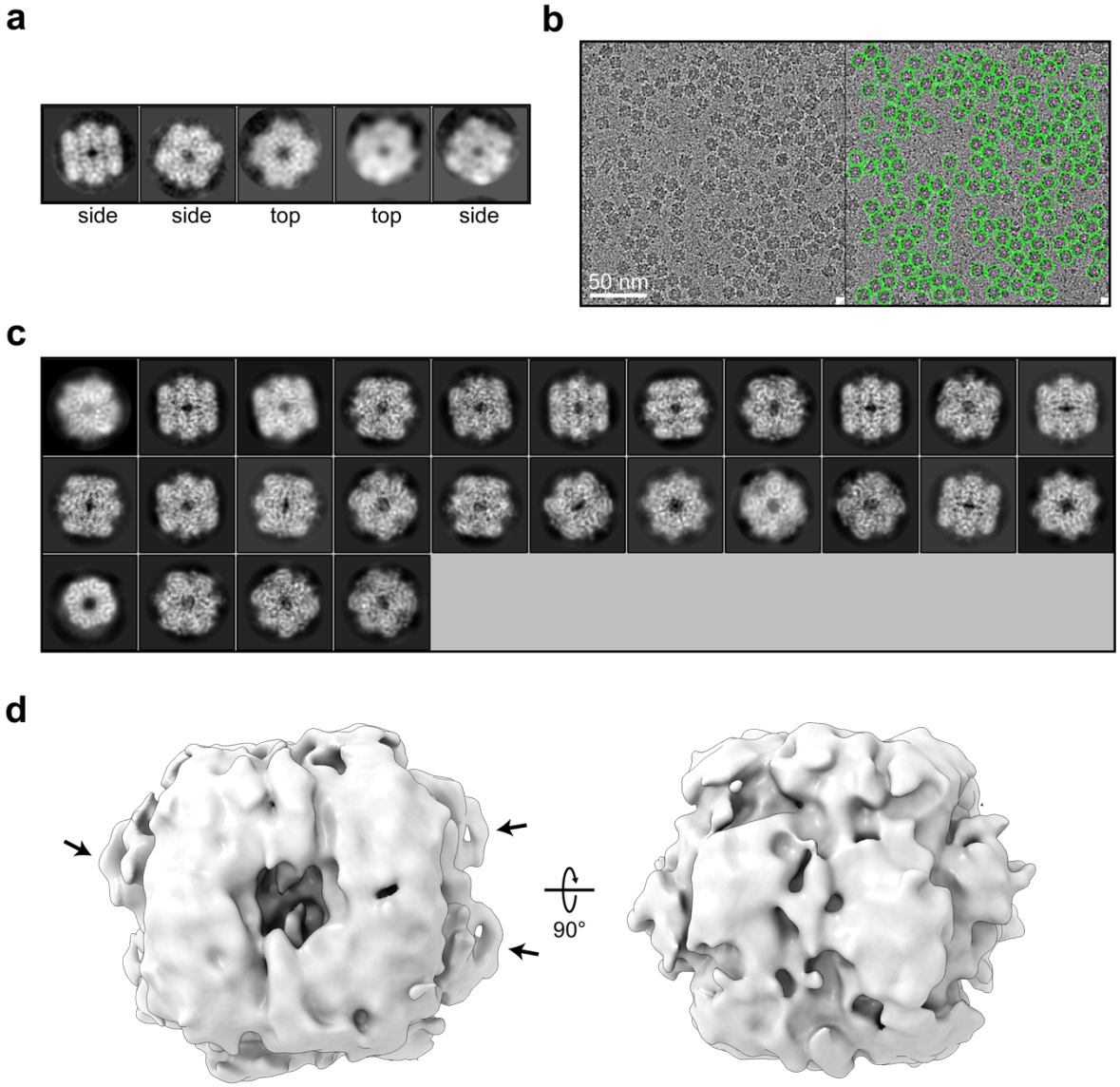


Figure 4-8 | 2D classification, autopicking of particles and *de novo* model generation.
(a) Selected side and top view 2D classes of the 465 manually picked particles used as 2D references for subsequent auto-picking. (b) Representative motion-corrected micrograph shown without (left) and with (right) the green auto-picked particle overlay. (c) Selected 2D classes of the auto-picked particles used for subsequent 3D classification. (d) *De novo* 3D Rubisco model generated from 2D classes. Arrows indicate regions of additional density present in the map, that are not accounted for by Rubisco holoenzyme structural features.

4.3.4. 3D classification

Using the *de novo* 3D model as a reference map lowpass filtered to 50 Å, 3D classification of the selected 2D particles was completed with C1 symmetry into 5 classes with a regularisation parameter

(T) of 4 (Fig. 4-9a). Of the resulting classes, only 1 demonstrated high resolution structural features consistent with the Rubisco holoenzyme. The resultant class possessed 33% of the input particles from 2D classification and consisted of 73,962 particles, suggesting some heterogeneity in the quality of the sample. The additional density that was identified in the *de novo* 3D model (Fig. 4-8d) was clearly present in each of the axes of this 3D class and was of the size scale expected for the predicted helical regions of the α3-α4 fragment. This class was subsequently subjected to 3D auto-refinement and post-processing with C1 and D4 symmetry, which demonstrated very little difference between the maps (Fig. 4-9b), confirming that no consistent asymmetric features were present in the particle dataset. Following auto-refinement and post-processing the additional density present in the 3D class was much less visible in both the D4 and C1 refinements, indicating much lower resolution information was present in this region of the reconstruction.

4.3.5. CTF refinement, polishing and post-processing of the D4 map

The D4 symmetry refined map (of 3D refinement resolution 2.86 Å) was processed through iterative rounds of CTF refinement and 3D refinement to correct for beam tilt and 4th order aberrations (2.81 Å), anisotropic magnification (2.81 Å), per-particle defocus and per-micrograph astigmatism (2.73 Å) sequentially (see methods). Bayesian polishing was completed on the resulting particles, which increased the resolution to 2.57 Å prior to post-processing (2.39 Å) (Fig. 4-9c,d). The FSC curve indicated that no additional features were introduced from masking during refinement. In the resultant map, the resolution of the additional density was too low to allow fitting of any candidate structural features. Several possible explanations for the lower resolution of this density exist. If the structural feature responsible for the density binds with heterogenous conformations in each subunit or is bound sub-stoichiometrically with respect to the number of subunits, or indeed a combination of both, the density across the subunits would be averaged leading to a lower resolution reconstruction. Given that in both scenarios the heterogeneity would presumably not show preference for certain subunits, it is also unlikely they would be isolated during C1 classification in which the particle is considered as one unit. As such, an alternative approach in which binding at each subunit is considered individually was pursued.

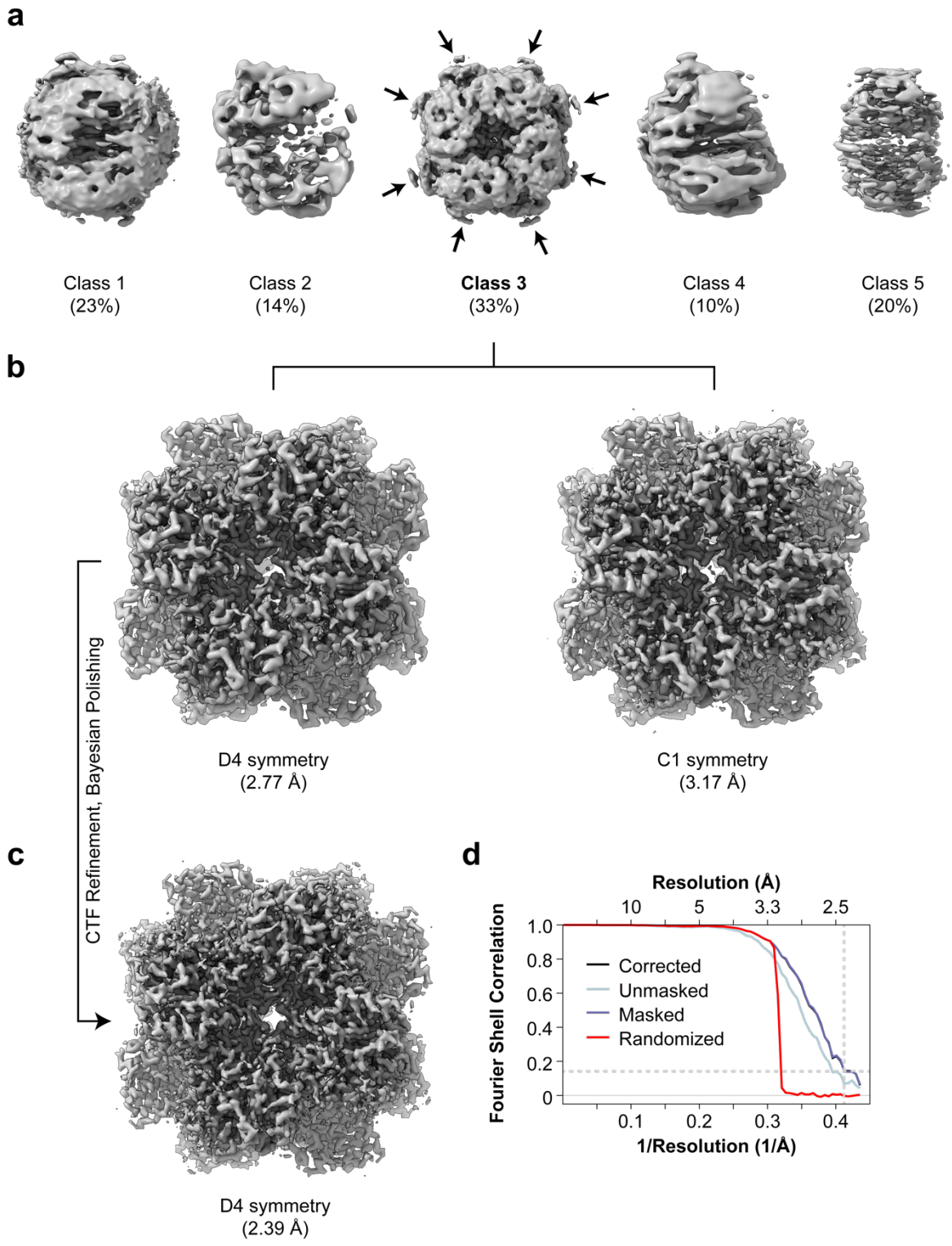


Figure 4-9 | 3D classification and symmetry comparisons.
(a) Maps of the 5 3D classes and the distribution of particles within them. Particles in class 3 (bold) were selected for further processing. Regions of additional density were present in the 3D class (arrows). All maps are shown at a contour level of 0.01. (b) Post-processed maps of class 3 from Fig. 4-9a with C1 or D4 symmetry imposed during refinement. Both maps are shown at a contour level of 0.032, at which the low-resolution additional density was not visible. (c) Post-processed map following CTF refinement and Bayesian polishing with D4 refinement. The map is presented at a contour level of 0.0553. (d) Fourier shell corelation curve showing the resolution estimate for the D4 refined map with an FSC cut-off of 0.143.

4.3.6. Symmetry expansion, 3D classification and refinement of the α3-α4-CsRubisco complex

Assuming the lower resolution of the putative α3-α4 density was due to sub-stoichiometric binding in the complex, a potential approach to increase the resolution of the reconstruction was to separately classify subunits in which the fragment is bound, and complete refinements of the resultant classes separately. The nature of the interaction between α3-α4 and CsRubisco, which was presumed to give rise to a distribution of occupancies on each Rubisco (from 0-8 occupied binding sites), precluded the use of whole particle 3D classification to do this. Instead, the polished particle dataset (73,962 particles) which gave rise to the 2.39 Å D4 reconstruction (Fig. 4-9) was D4 symmetry expanded to separate the 8 Rubisco large subunits of CsRubisco for classification separately (591,696 effective particles). A soft, featureless mask encapsulating the region of additional density present in the 3D class (Fig. 4-9a) was created in UCSF Chimera by mapping the predicted helix of α3 into the density and lowpass filtering to 20 Å (Fig. 4-10a). The resulting mask was used in 2 rounds of 3D classification, in which bound and unbound symmetry axes were clearly separated (Fig. 4-10b,c). After the second round of 3D classification, a single class with high resolution features was obtained, suggesting there was little heterogeneity in the binding pose of the α3-α4 fragment in the complex. The particles from the resulting class were refined with C1 symmetry and post-processed to 2.76 Å (Fig. 4-10d,g). During the reconstruction the symmetry axes containing the additional density were aligned, leading to reconstruction of the additional density at one of the subunits and removing the deprecating effect of averaging across the 8 symmetry axes encountered previously (Fig. 4-9c). The lack of density at the other subunits indicated good selectivity of the 3D classification. In line with this, the final non-sharpened map demonstrated improved resolution for the additional density (Fig. 4-10e) over the density in the non-sharpened post-processed D4 map (Fig. 4-10f). In this reconstruction, the density also presented clearly helical nature for the first time, in line with the hypothesis that the predicted helices of CsLinker underpin the interaction with Chlorella Rubisco. Based on the origin of the symmetry expanded particles present in the 3D class used for reconstruction (Fig. 4-10c), it was also possible to determine the distribution of the number of symmetry axes bound in each Rubisco. On average, per Rubisco ~23% of symmetry axes possessed the additional density indicating that only 2 of the 8 available binding sites for α3-α4 on CsRubisco were occupied under the conditions the dataset

was collected. In agreement, there was a clear distribution of the number of axes occupied centred on the mean (Fig. 4-10h). This clearly indicated that despite the incubation of α3-α4 at concentrations found to saturate Rubisco by SPR and native PAGE, the α3-α4 fragment was indeed sub-stoichiometrically bound.

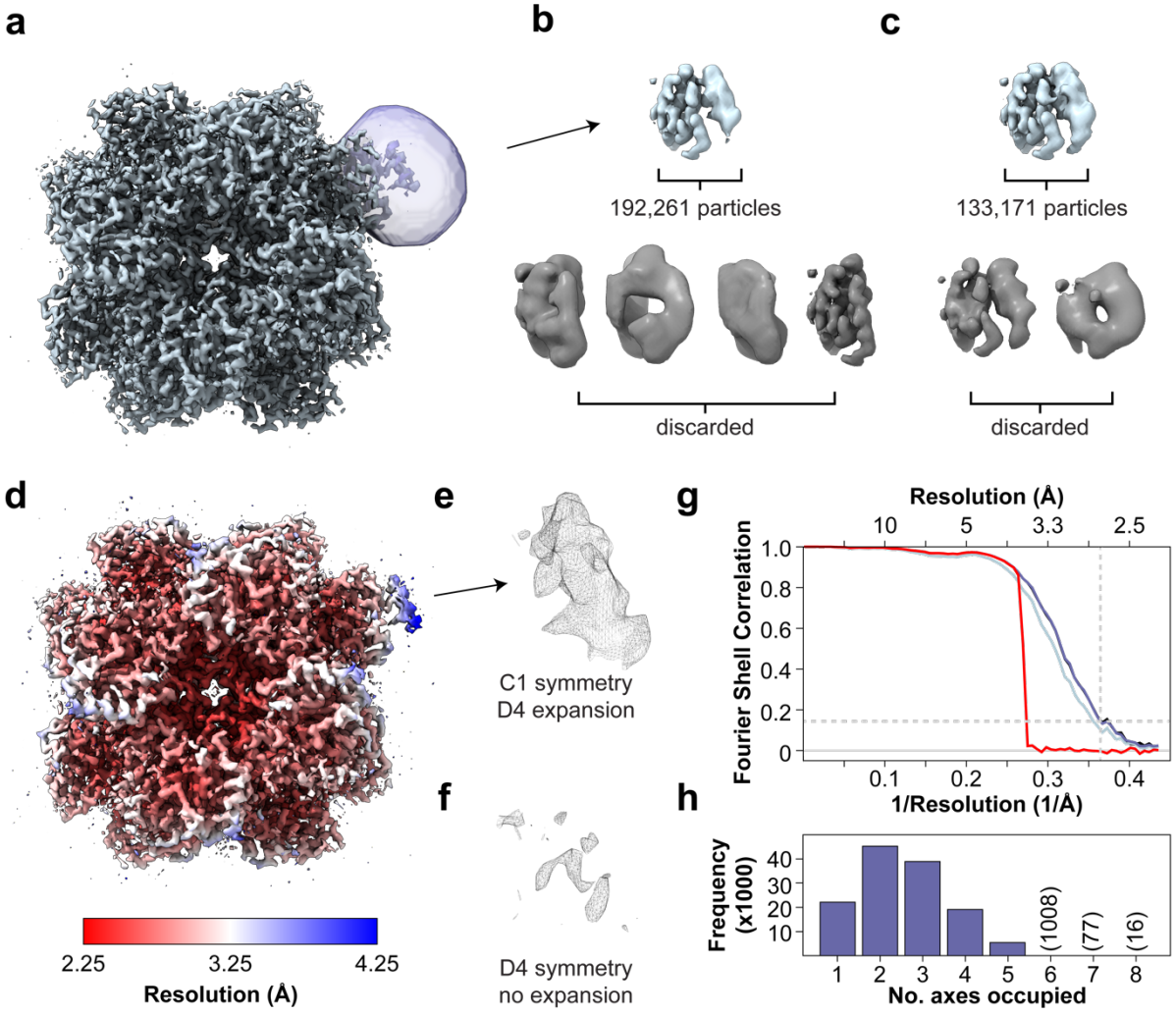


Figure 4-10 | Symmetry expansion, 3D classification and refinement of the α3-α4-CsRubisco complex.
(a) Sharpened (*B*-factor: 45) post-processed map of the C1 symmetry-refined α3-α4-CsRubisco complex using the D4 symmetry expanded particle dataset. The soft featureless mask used for the classification is shown in purple, over the region of additional density, into which the predicted helix of α3 is built. Shown at a contour level of 0.0496. (b) The 5 3D classes from the first round of C1 classification with the featureless mask. The selected class is shown in blue. (c) The 3 3D classes from the second round of classification, with the selected class used for refinement shown in blue. (d) Local resolution estimate from Phenix shown on the map as in a. (e) Focus on the additional density present in the C1 refined, unsharpened post-processed map derived from the symmetry expanded particles in Fig. 4-10c. Shown at contour level 0.0293. Density is carved within a radius of 2 around the predicted helical region. (f) Focus of the equivalent region in the D4 refined, unsharpened post-process map derived the non-symmetry-expanded dataset in Fig. 4-9c. Shown at contour level 0.0174, which showed equivalent density for the Rubisco holoenzyme as the contour level used to produce Fig. 4-10e. Carved to the same level. (g) Fourier shell corelation curve showing the resolution estimate for the C1 refined map with an FSC cut-off of 0.143. (h) Histogram of the number of symmetry expanded axes in each particle used in the 3D class selected for refinement of the C1 map.

4.3.7. *Model building and refinement of the CsRubisco holoenzyme*

Before regions of the CsLinker could be built into additional density present in the final C1 map, it was necessary to build coordinates for the CsRubisco holoenzyme itself, to allow more accurate refinement of the complex at later stages. For model building of the CsRubisco holoenzyme, AlphaFold2 models of the CsRbcS and CsRbcL sequences were aligned to the Chlamydomonas Rubisco structure (PDB: 7JN4; (He *et al.,* 2020)) and the model was rigid body fitted into the D4 map. The coordinates for one CsRbcL and one CsRbcS chain were then flexibly fitted in Coot before real-space refinement in Phenix and application to the other 7 chains. Several regions of the RbcL did not have clear density in the map, as observed for the Chlamydomonas cryo-EM structure (He *et al.,* 2020), and were removed from the model. As expected, the coordinates showed high similarity to the Chlamydomonas Rubisco cryo-EM structure (average Cα root mean square deviation of RbcL = 1.106 Å and RbcS = 0.937 Å) (Fig. 4-11).

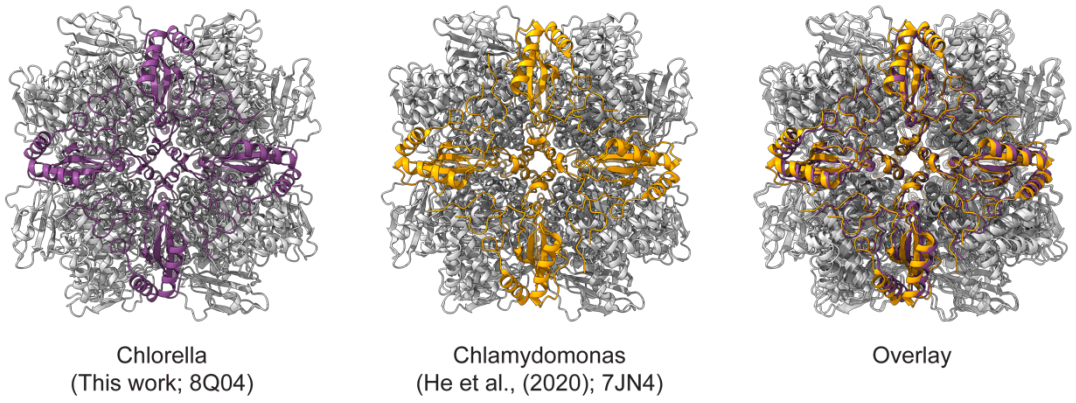


Figure 4-11 | Comparison of Chlorella and Chlamydomonas cyro-EM Rubisco holoenzyme coordinates. Ribbon representation of the D4 refined Rubisco holoenzyme coordinates built into the map from Fig. 4-9c (left). Ribbon representation of Chlamydomonas Rubisco (middle) and the overlay (right) showing high similarity of the overall structures.

4.3.8. *Building CsLinker in the complex CsRubisco-α3-α4 complex*

Given that the additional density presented clear helical nature in the higher resolution refinements (Fig. 4-10a,e) , it was assumed this corresponded to the predicted helical regions of the α3-α4 fragment that were hypothesised to underpin the interaction with CsRubisco. Given there is only a single amino acid difference between the two helical regions in the α3-α4 fragment (Fig. 4-2), it was also assumed both the α3 and α4 helices would have equal affinity for, and consequent presence in, the complex with

CsRubisco. Further to this, the α3-α4 fragment was incubated with Rubisco at 16 μM, which was well below the K_D of the α3 fragment (~100 μM), but in excess of the α3-α4 fragment (~1 μM), suggesting that for an α3-α4 fragment to be present in the complex, both helices would have to be bound and therefore should be equally abundant in the dataset. The resolution of the density was not high enough for distinction between the amino acid difference in the helices. As such, the helix of α3 was elected to build into the C1 complex map as it is the more representative helix sequence in the full length CsLinker (Fig. 4-2), though it is likely that higher resolution data could have allowed classification of the differences between the two helices. For model building, the coordinates of the holoenzyme from the D4 model (Fig. 4-11) were positioned in the C1 complex map and real space refined. In Coot, the AlphaFold2 predicted helical region of α3 was then flexibly fitted into the additional density present, using the bulky density at the centre as an anchor for the central tyrosine residue. 3 additional residues, N-terminal to the predicted helix could also be built manually into the density and represented some conserved structure in the predicted disordered region (Fig. 4-12a). The complex of one CsRbcS, one CsRbcL and the α3 helix was real space refined in Phenix and the coordinates of the CsRbcS and CsRbcL were consequently applied to remaining 7 subunits. The side chains of multiple residues on the non-interacting face of the helix were removed from the model due to their poor resolution in the map (Fig. 4-12b). These residues were all non-interacting electrostatic residues with highly flexible side chains, explaining their lack of defined conformation in the complex.

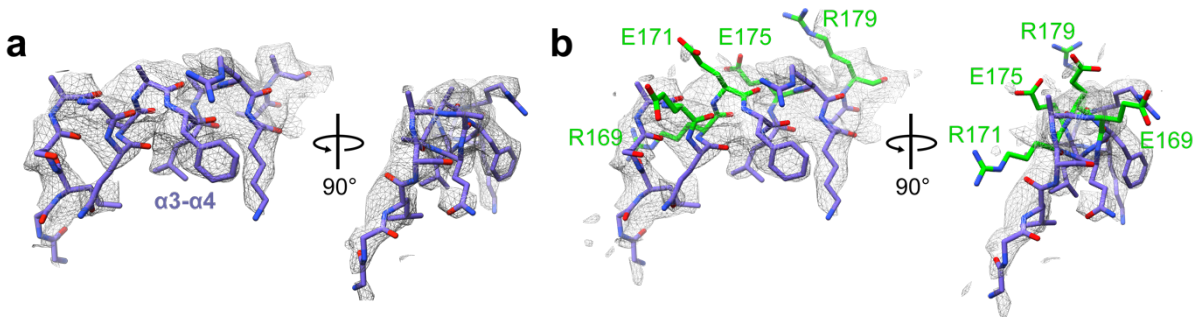


Figure 4-12 | CsLinker helical coordinates fitted into the additional density in the C1 map.
(a) Final coordinates of the α3 helix of the α3-α4 fragment mapped into the additional density present in the C1 map (Fig. 4-10d), shown with the non-fitted side chains removed at a contour level of 0.033, carved with a radius of 2 from the coordinates. (b) The same region shown with the side chains that were not modelled due to poor density in green. All the poor resolution side chains are present on the non-interacting face of the helix.

4.3.9. Collection, model quality and processing

Final coordinates of the structures were validated in MolProbity and submitted to the PDB (Table 4-1).

Table 4-1: Collection, refinement, and validation of cryo-EM models.

Data Collection and Processing		
Magnification	240,000	
Voltage (kV)	200	
Electron fluence (e-/Å^2)	50	
Pixel Size (Å^2)	0.574	
Symmetry	D4	C1
Figure for map	4-9c	4-10d
Initial Particles	237,035	591,696 (8x expanded)
Final Particles	73,962	133,171
Map resolution (Å)	2.39	2.77
FSC Threshold	0.143	0.143
Map resolution range (Å)	8.05-2.23	9.42-2.29
Refinement	**CsRubisco**	**CsRubisco-α3-α4**
Map sharpening *B* factor (Å^2)	50	45
Model Composition		
Non-H atoms	35979	35836
Residues	4424	4440
Water	1267	1011
Mean B factors (Å^2)		
Protein	44.00	44.10
Water	42.40	47.97
RMS deviations		
Bond length (Å)	0.008	0.009
Bond angles (°)	1.164	1.171
Validation		
MolProbity Score	1.30	1.30
Clashscore	2.09	2.20
Poor rotamers %	0	0
Ramachandran plot		
Favored %	95.41	95.61
Allowed %	4.59	4.39
Outliers	0	0
PDB Code	8Q04	8Q05

4.3.10. The molecular basis of CsLinker interaction with CsRubisco

Analysis of the CsRubisco-α3-α4 coordinates showed the helix of CsLinker binds to the widest region of CsRubisco (Fig. 4-10a,d), at an N-terminal interface of the CsRbcL formed by the αB helix, βC sheet, CD loop and the N-terminal region of the βD sheet (Fig. 4-13a). Given that α3 was modelled in the complex, all subsequent numbering of residues was completed according to the position of the corresponding residues of α3 in the full length CsLinker sequence. PDBePISA analysis of the complex coordinates highlighted several key interactions between the helix of CsLinker and CsRbcL (Table 4-2). The interface primarily consisted of two salt bridges on a hydrophobic interface (Fig. 4-13b). The two salt bridges were between Arg176 of CsLinker and Glu51 of αB in the CsRbcL, and between Lys177 of CsLinker and Asp86 of CsRbcL βC. Arg176 was positioned in the complex to form multiple salt bridges with Glu51 and likely adopts multiple conformations that were not resolved in the density map. Lys177 occupied a notably extended rotamer to bridge an unusually large gap (11.5 Å) to Asp86 of the CsRbcL, considering the adjacency of Arg176 and Lys177 in the helix of CsLinker. Phe173 of CsLinker dominated the hydrophobic interaction with high buried surface area in a hydrophobic pocket formed by Ile87 and Tyr97 of the CsRbcL βC and βD sheets respectively. Leu174 and Leu166 were also positioned to contribute to the hydrophobic interaction with Pro89 and Pro91 of the CsRbcL βD and CD loop respectively (Fig. 4-13b). Gln170 of CsLinker also had high buried surface area within the complex and was positioned to form hydrogen bond interactions with residues in the CD loop, though density in the C1 complex map precluded empirical conclusion of this. Taken together the model of CsRubisco-α3-α4 allowed identification of the key interacting residues between CsLinker and CsRubisco and confirmed the hypothesised interaction of the conserved residues within the helices (Fig. 2-15a). The composition of the interface residues were largely consistent with those of the EPYC1-Chlamydomonas Rubisco interface, representing a hydrophobic interface anchored with salt bridges. Importantly however, the structure of CsRubisco-α3-α4 demonstrated binding to an altogether different binding interface in the RbcL of Rubisco, in contrast to the expectation that binding would occur in the RbcS as observed previously for EPYC1 (He *et al.,* 2020). This was further evidence for the convergent evolution of the linker proteins in Chlamydomonas and Chlorella.

Table 4-2: PDBePISA analysis of α3-α4-CsRubisco complex interface.

Salt bridges			
α3-α4 residues (atom)	CsRbcL residue (atom)	Distance (Å)	Secondary structure
Arg176 (NE)	**Glu51 (OE2)**	**3.04**	**αB**
Arg176 (NE)	Glu51 (OE1)	3.93	αB
Arg176 (NH2)	Glu51 (OE2)	2.96	αB
Lys177 (NZ)	**Asp86 (OD1)**	**3.56**	**βC**
Hydrogen bonds			
Gly165 (O)	Gly92 (N)	2.66	CD loop
Gln170 (NE2)	***Glu93 (O)***	***3.29***	***CD loop***
Gln170 (NE2)	***Gln96 (O)***	***3.30***	***CD loop***
Arg176 (NE)	Glu51 (OE2)	3.04	αB
Arg176 (NH2)	Glu51 (OE2)	2.96	αB
Lys177 (NZ)	Ile87 (O)	2.68	βC
Lys177 (NZ	Asp86 (OD1)	3.56	βC
α3-α4 Interface			
α3-α4 residue	Accessible surface area (Å^2)	Buried surface area (Å^2)	% Buried
Gly164	121.72	28.64	23.5
Gly165	84.64	31.20	36.9
Leu166	**105.80**	**36.00**	**34.0**
Ser167	63.01	15.15	24.0
Ala168	87.15	0.00	0.0
Glu169	73.76	23.02	31.2
Gln170	***95.46***	***95.46***	***100.0***
Arg171	46.76	0.00	0.0
Glu172	48.95	0.00	0.0
Phe173	**90.09**	**90.09**	**100.0**
Leu174	**82.49**	**12.39**	**15.0**
Glu175	63.88	0.00	0.00
Arg176	158.68	62.48	39.4
Lys177	149.89	64.24	42.9
Ala178	89.52	0.00	0.0
Arg179	124.39	0.00	0.0

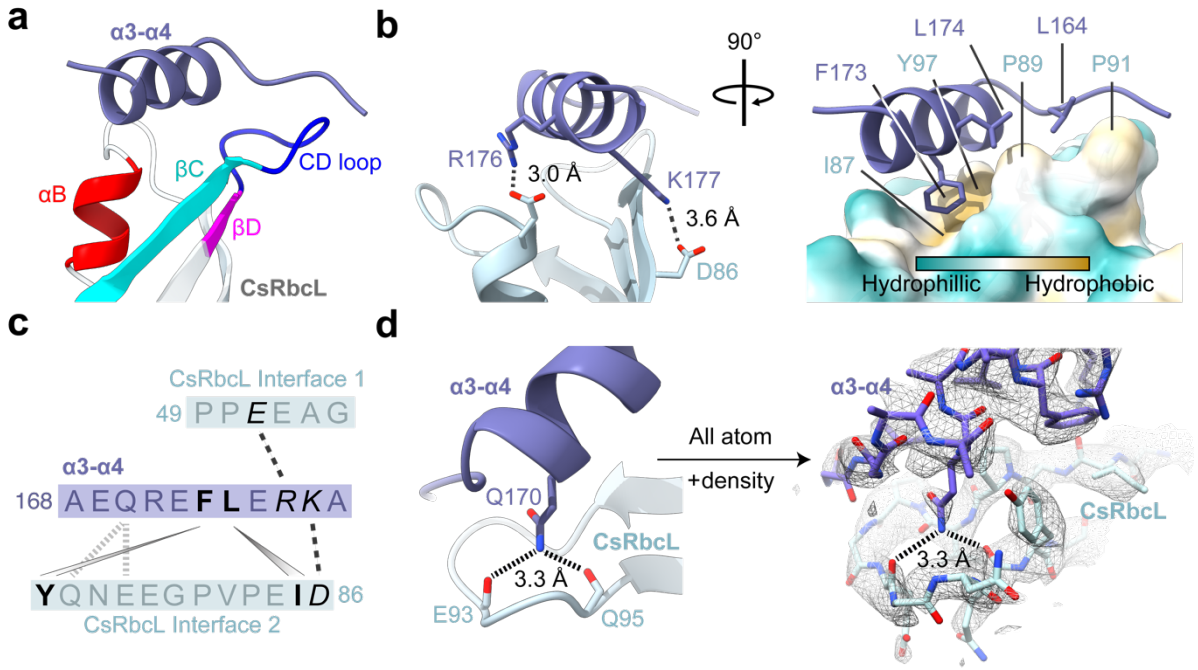


Figure 4-13 | CsLinker interaction with CsRbcL.
(a) Ribbon representation of the CsLinker(α3-α4)-CsRubisco interaction interface from the C1 complex structure. CsRbcL is coloured according to the structural features discussed in text; red is helix αB, cyan is sheet βC, blue is the CD loop, magenta is sheet βD. (b) Dominating molecular interactions at the interface determined by PDBePISA analysis of the structure. The shortest range electrostatic residues for each residue are shown (left), as well as the residues identified as contributing to the hydrophobic interface (right), where the surface of CsRbcL is coloured according to hydrophobicity according to Kyte-Doolittle scoring. (c) Map of interactions between α3-α4 and CsRubisco, where salt bridges are indicated by black dashed lines, hydrophobic interactions by wedges and possible hydrogen bond interactions by gray fine dashed lines. (d) The possible hydrogen bond network between Q170 and the CD loop of CsRbcL shown as a ribbon representation (left) and as an all atom model in the C1 complex map density (Fig. 4-10d), shown at contour level 0.0396 (right).

4.4. Biochemical validation of the CsLinker interaction interface

Having determined the 3D structure of the complex between α3-α4 of CsLinker and CsRubisco, it was next desirable to validate the identified molecular interactions at the interface. As the previously completed native PAGE experiments with the α3-α4 fragment proved to be repeatable and consistent with high resolution SPR experiments (Fig. 4-5 and Fig. 4-6), mutagenesis of the fragment for use in native PAGE experiments was identified as a primary technique to validate the binding site of the CsLinker protein. Validation of the binding site on Rubisco is completed by proxy in Chapter 5.

4.4.1. Site-directed mutagenesis (SDM) and native PAGE validation of α3-α4-CsRubisco interaction

Based on the identified interactions in the C1 complex (Fig.4-13c), two strategies to separately ablate the hydrophobic and electrostatic interactions were designed. For these strategies the initial goal was to perform the same mutation at the relevant residues in both helices in the α3-α4 fragment. To ablate the hydrophobic interactions, the central Phe (position 173 in α3 and position 239 in α4) and adjacent Leu (position 174 in α3) or Met (position 240 in α4) were mutated to Asp residues, and the resulting fragment was denoted the 'hydrophobic' mutant. Asp was chosen as it possesses vastly different properties to the origin residues, and was previously shown to ablate hydrophobic interaction between EPYC1 and Chlamydomonas Rubisco in peptide tiling arrays, yeast two-hybrid and *in vivo* experiments (He *et al.,* 2020). To ablate the electrostatic interaction, Glu (positions 169 (α3) and 235 (α4)) and Arg (positions 176 (α3) and 242 (α4)) residues were mutated to Ala based on the same reasoning as the hydrophobic mutations, in addition to the helix-promoting nature of Ala residues (Rohl *et al.,* 1999). This fragment was denoted the 'electrostatic' mutant. Mutation of the Glu (169/235) residues was based on an early refinement of the structure (not shown), which suggested this residue plays a role in the interaction, contrary to the final high-resolution structure (Fig. 4-13), as well as conservation of the residue in the consensus helix (Fig. 2-15a). Analysis of the electrostatic SDM fragments is therefore interpreted independent of changes to this residue and focusses on the impact of the Arg(176/242) mutation. Due to fidelity of the Gibson Assembly SDM approach used to create the constructs (see methods), an additional unplanned construct in which only Arg176 was mutated to Ala was also obtained, which was

denoted the ‘partial electrostatic’ mutant. Importantly, AlphaFold2 modelling of the mutated α3-α4 fragments indicated the same length predicted helix as the WT sequence would be formed in solution (Fig. 4-15a,b), allowing isolation of any interaction changes to molecular details rather than structural conformation changes.

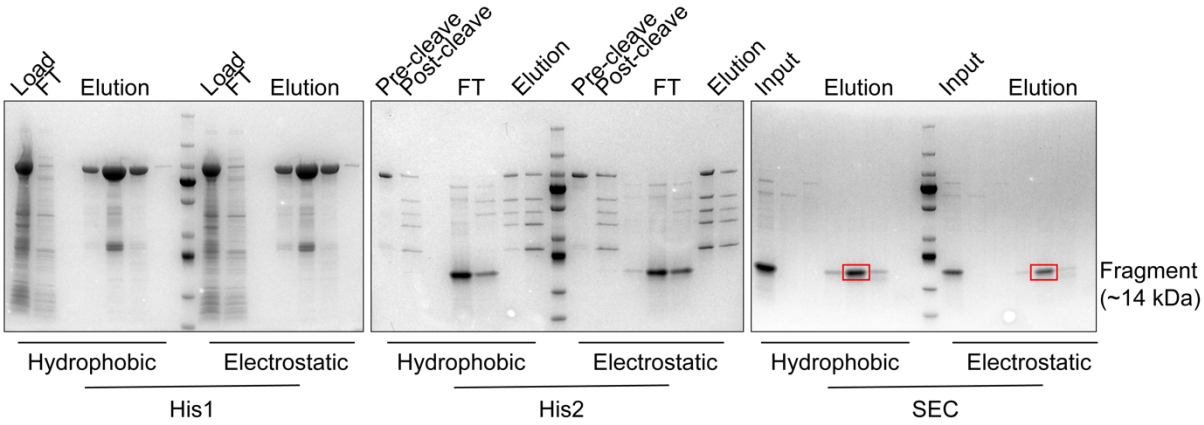


Figure 4-14 | Purification of the SDM mutants of the α3-α4 fragments.
SDS-PAGE analysis of samples taken during the first and second His affinity (His1 and His2 respectively) and size-exclusion chromatography (SEC) purifications of the hydrophobic and electrostatic mutant fragments. The pooled fractions taken for experiments are highlighted by the red boxes for each mutant. Data for the partial electrostatic mutant is not shown.

The mutant fragments were purified to a high level (Fig. 4-14), using the same strategy previously employed for the WT α3-α4 fragment (Fig. 4-3). Both the hydrophobic and electrostatic mutant fragments showed no band shift in native PAGE experiments at concentrations shown to fully shift the CsRubisco into the higher order complex with the WT fragment (Fig. 4-15c). Interestingly, the partial electrostatic mutant appeared to show an intermediate band shift. The concentration at which the partial band shift was observed was consistent with the concentrations observed to shift CsRubisco using the single helix fragment (α3) previously (Fig. 4-4a). This could suggest that by mutating only one of the Arg residues, only one interaction site has been ablated and the affinity for CsRubisco has been shifted towards that of the α3 fragment. Irrespective of this conclusion, the intermediate shift demonstrates that the Arg has a definite contribution to the interaction with CsRubisco. Taken together, the SDM native PAGE experiments were consistent with the interacting residues highlighted from the 3D structure of the α3-α4-CsRubisco complex.

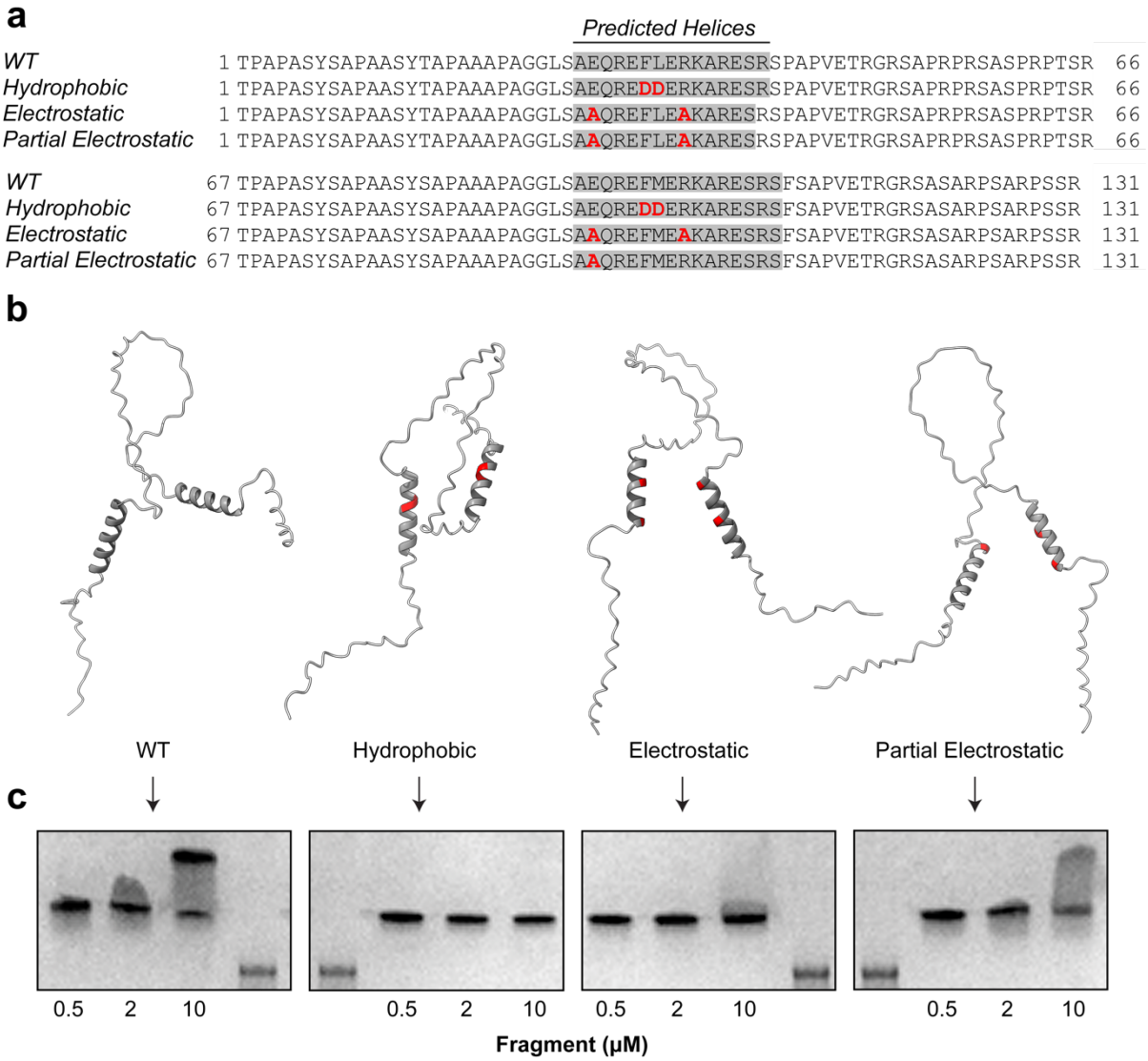


Figure 4-15 | SDM mutants of the α3-α4 fragments.
(a) Alignment of the 3 SDM mutants of the α3-α4 fragments with the AlphaFold2 predicted helical regions indicated by the gray box. The mutated residues in the respective constructs are shown in red. (b) Top ranked AlphaFold2 models of the SDM mutants. The mutated residues within the helical regions are shown in red. (c) Native PAGE band shift assays completed with the SDM mutated α3-α4 fragments according to b.

4.5. Discussion

In this body of work the aim was to characterise, at a biochemical and structural level, the interaction between CsLinker and CsRubisco. Based on the analogous structural properties of the CsLinker and analogous mesoscopic properties of the *in vitro* pyrenoid reconstitution when compared to that of Chlamydomonas, it was naively hypothesised that the molecular details of the interaction would also be largely similar to that of Chlamydomonas. Instead, the picture of convergent evolution of pyrenoid proteins outlined in Chapter 2, was reinforced by several key differences alongside functional similarities in the linker protein that emerged from the characterisation.

The first key hypothesis, and similarity with Chlamydomonas and Phaeodactylum linker proteins that was confirmed, was that the interaction of the CsLinker protein with CsRubisco was underpinned by the predicted repeating helical motif. Based on the similarity in sequence composition of the helices of CsLinker and EPYC1 (Table 4-3), it was expected that the helix would also have similar affinity and interaction mode for CsRubisco. Instead, a significantly different affinity of the single helix fragments for Rubisco (~100 μM in CsLinker vs. ~3000 μM in EPYC1) was demonstrated, alongside an equally different interaction mode of the double helix fragments (stable complex formation vs. LLPS). Although no direct comparison can be made between the experiments due to differences in the construct design, this finding does suggest that there is likely a broad parameter space for linker interaction strength and mode that has been optimised over the ~250 My evolution of the proteins to allow functional convergence.

Table 4-3: EPYC1 and CsLinker helix composition.

Residue property	EPYC1 helix	CsLinker helix
Negative (D/E)	2	2
Positive (K/R)	2	2
Aromatic (F/W/Y)	1	1
Hydrophobic (A/I/L/M/V)	2	2
Other	4	1

A second hypothesis, that EPYC1 and CsLinker may have evolved to use the same interface of Rubisco, was spectacularly disproven. This hypothesis was based on the similarity in helix size, properties and spacing in the two proteins, as well as their phylogenetic relationship, and the shared interaction interface used in cyanobacterial linker proteins (Oltrogge *et al.,* 2020; Wang *et al.,* 2019), that likely evolved in the same period. Additionally, preliminary HADDOCK molecular docking experiments, performed according to correct predictive work with EPYC1 and Chlamydomonas RbcS prior to 3D structure determination (personal communication with Oliver Caspari), showed convincing likelihood of a RbcS-mediated interaction (Fig. 4-16). In contrast, an interaction interface in the N-terminal region of the CsRbcL was uncovered that shared no similarity with previously published interaction sites of linker proteins in pyrenoids (He *et al.,* 2020; Oh *et al.,* 2023), or carboxysomes (Oltrogge *et al.,* 2020; Wang *et al.,* 2019). The increasing number of non-shared interaction interfaces characterised in pyrenoid linker proteins suggests that there is little restraint on the Rubisco binding site used for interaction, as long as the interaction can fulfil some minimal affinity criteria that can be compensated by the rest of the linker sequence and its interaction mode; a feat that seems eminently achievable over 250 My of evolution. The binding to the large subunit of Rubisco is also of huge interest with respect to plant engineering goals, as the RbcL is chloroplast-encoded and highly conserved in the green lineage. This is discussed in detail in Chapter 5.

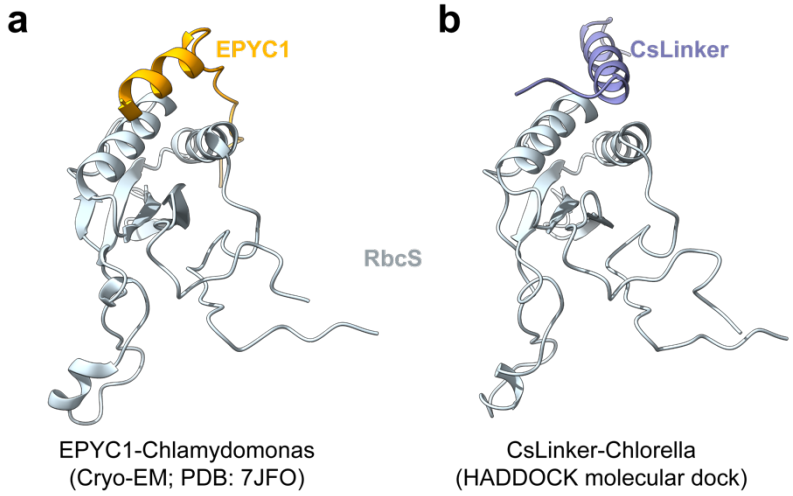


Figure 4-16 | Chlorella RbcS was a plausible interaction interface for CsLinker.
(a) EPYC1 (residues 49-72) bound to the Chlamydomonas RbcS from the indicated cryo-EM structure. (b) HADDOCK molecular dock of CsLinker to the Chlorella RbcS, the lowest energy dock is presented (see methods).

The use of the 2 helix α3-α4 fragment of CsLinker in cryo-EM studies granted the highest chance of yielding a complete picture of CsLinker binding to CsRubisco, in contrast to previous approaches using shorter fragments that did not represent full repeat regions of the respective linker proteins (He *et al.,* 2020; Oltrogge *et al.,* 2020). In contrast to the Phaeodactylum linker protein PYCO1 where two regions of additional density corresponding to different sequence elements were present in the complex map, the CsLinker-CsRubisco complex map demonstrated only a single region of additional density (Oh *et al.,* 2023). This suggested that the disordered spacer region between helices does not have a direct role in binding to CsRubisco but does not rule out a role in overall affinity through entropic effects. Further work to assess the contribution of the disordered regions to the binding affinity, either by native PAGE or SPR of shorter fragments could be completed to empirically determine this.

More generally, the high affinity and different interaction modes of CsLinker fragments for CsRubisco relative to EPYC1 fragments for Chlamydomonas Rubisco may explain the decreased mobility of pyrenoid components observed in the *in vitro* reconstitution (Fig. 3-16). Whether the affinity of the fragments explains the formation of the observed stable α3-α4-CsRubisco complex over LLPS remains to be seen, though evolving modelling approaches for Rubisco-linker interactions may go some way to answer these questions (GrandPre *et al.,* 2023; Schaefer, 2023). The sub-stoichiometric binding of the α3-α4 fragment to CsRubisco determined by symmetry expansion in cryo-EM appears especially puzzling, given the apparent saturation of the higher order complex by native PAGE and SPR analysis. Further work to understand the relative position of each of the sub-stoichiometrically bound helical regions on each CsRubisco, using the relative transformations performed during symmetry expansion, may go some way to explain this but were unfinished within the timeframe of the study. As an attempt to offer a theoretical explanation; the primary sequence of the helices in CsLinker determines an implied directionality to themselves and the disordered spacer region between them. Equally, the 3D structure of Rubisco determines a directionality of the equivalent interaction site. Based on the assembly of Rubisco, the interaction sites on adjacent RbcLs are antiparallel, meaning the corresponding interacting helices and disordered spacer region must be positioned equally antiparallel to interact with adjacent Rubiscos (Fig. 4-17a). One can imagine an alternative scenario in which binding of the two helical

regions of α3-α4 is preferred to non-adjacent CsRbcLs, such that the adjacent CsRbcL binding site is obscured, reducing the effective valency of CsRubisco from 8 to 4 (Fig. 4-17b). The preference for each of these interaction modes would presumably be dominated by the entropic penalty of confining the disordered spacer region to an extended confirmation, which in turn is determined by the innate flexibility of said spacer region (Pritišanac *et al.,* 2019). Determination of the flexibility of the spacer region, in combination with aforementioned computational models and relative RbcL occupancy would likely go some way to answering this question, and more generally the role of flexibility of the spacer region in determining the overall properties of the linker protein.

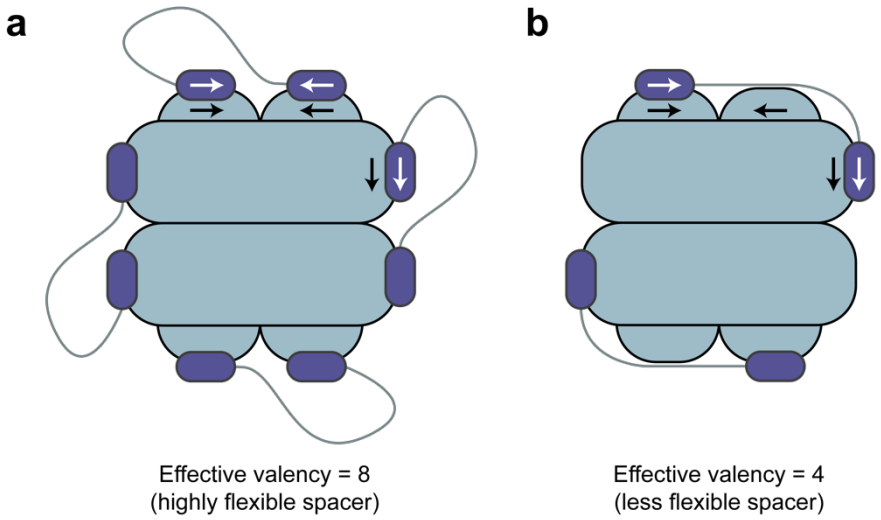


Figure 4-17 | Proposed explanation for the sub-stoichiometric binding of α3-α4 to CsRubisco.
(a) A scenario in which a high flexibility spacer regions permits the binding of adjacent, anti-parallel binding sites on CsRubisco (indicated by arrow directionality). The effective valency of CsRubisco is 8, as each CsRbcL is accessible. (b) A low flexibility spacer region scenario, in which binding to non-adjacent, parallel binding sites on CsRubisco is preferred, obscuring the intermediate anti-parallel site from binding, and reducing the effective valency of CsRubisco to 4.

In summary, drawing on the expertise of key members at the University of York and building on the technical foundations of others in the field, in this chapter a robust biochemical and structural characterisation of the CsLinker-CsRubisco interaction was completed. Through this it was demonstrated that a linker protein with analogous secondary structure to EPYC1 has evolved a different affinity, interaction site and binding mode for Rubisco, reinforcing the convergent evolution of linker proteins and their properties. Notably, the binding to the Rubisco large subunit is a cause for great excitement, as demonstrated in Chapter 5.

4.6. Materials and methods

Unless otherwise stated, all strains and culture conditions, general molecular biology and protein biochemistry methods were performed consistent with the methods outlined in Chapter 3. General methods introduced in this chapter are outlined here.

4.6.1. General molecular biology methods

4.6.1.1. Gibson assembly

For all Gibson assembly reactions, 100 ng of the longest fragment (usually the plasmid backbone) was used in each reaction with a 3x molar excess of each of the inserts calculated based on the molarity of the backbone in the reaction. All reactions were completed in 20 µL volumes with NEB Gibson Assembly master mix (E2611 – New England BioLabs). The reactions were incubated for 1 hour at 50 ºC, before 5 µL aliquots of the reaction mix were separately transformed 3 times into Dh5α and plated on the same relevant antibiotic selection plates.

4.6.1.2. Colony PCR

For all colony PCR reactions, a reaction composition as outlined in table 4-4 and cycling conditions as outlined in table 4-5 were used. Colonies were restreaked with 200 µL pipette tips to fresh selection plates prior to mixing with the 25 µL colony PCR reaction volumes by aspiration. The annealing temperature (T_m) for each reaction was calculated using the ThermoFisher Scientific online T_m calculator set to *Taq*-based polymerase, after which 2.5 ºC was subtracted. In all cases an extension time of 60 seconds per kilobase of target amplicon was used. 10 µL of the reaction mix following amplification was resolved by DNA gel electrophoresis as described previously.

Table 4-4: *Taq* colony PCR composition.

Component	Volume (µL)
dH_2O	14.4
10x CoralLoad Buffer	2.5
dNTP mix (R0192 – ThermoFisher)	0.5
Forward primer (10 µM stock)	1.25
Reverse primer (10 µM stock)	1.25
5x Q-solution	5
Aspirated colony	-
Taq DNA Polymerase (5 U μL^{-1})	0.1
Total	**25**

Table 4-5: *Taq* colony PCR conditions.

Temperature (ºC)	Time (s)	Number of cycles
94	300	1
94	30	
Calculated	45	38
72	Calculated	
72	300	1

4.6.2. Cloning of α3 and α3-α4 CsLinker fragments

The α3 and α3-α4 sequences were PCR-amplified from pLM872 using primer pairs oLM3148/3149 and oLM3150/3151 respectively (Table 4-6), then subsequently gel extracted. The backbone of pLM872, from the C-terminal TEV site to the N-terminal TEV site was separately amplified with primers oLM3144/3145 and gel extracted. The α3 and α3-α4 fragments were Gibson assembled separately into this pLM872 backbone fragment to create pLM1286 and pLM1287 respectively. Colonies were screened by colony PCR using primers oLM1622/2237 and constructs with the correct amplicon were sequenced with oLM1622.

Table 4-6: Primers used for cloning of α3 and α3-α4 fragments.

Name	Purpose	Oligonucleotide Sequence
oLM3144	pLM872 Backbone forward	GAGAACCTCTACTTCCAATCCG
oLM3145	pLM872 Backbone reverse	GGATTGGAAGTACAGGTTTTCCTC
oLM3148	α3 forward	GATCGAGGAAAACCTGTACTTCCAATCCACC CCAGCACCTGCGAGTTATTC
oLM3149	α3 reverse	CGGATTGGAAGTAGAGGTTCTCACGAGAGG TCGGACGTGGTG
oLM3150	α3-α4 forward	GATCGAGGAAAACCTGTACTTCCAATCCACC CCAGCACCTGCGAGTTATTC
oLM3151	α3-α4 reverse	CGGATTGGAAGTAGAGGTTCTCACGACTAC TTGGGCGCGCTG
oLM1622	Colony PCR/forward sequencing of pLM1286/pLM1287	CACTACCTGAGCACCCAGTC
oLM2237	Colony PCR of pLM1286/pLM1287	GCTAGTTATTGCTCAGCGG

4.6.3. Cloning of site-directed mutagenized α3-α4 fragments

For site-directed mutagenesis of the α3-α4 fragment, primers containing the appropriate nucleotide mutations were designed for the nucleotide sequences corresponding to the helices of α3-α4. The 'backbone' (forwards from α4 throughout the backbone to α3) and 'insert' (forwards from α3 to α4) were amplified separately according to the primers in table 4-7. Separate Gibson Assembly reactions for each of the introduced mutations using the respective backbone and insert fragments were performed to create pLM1454 and pLM1455. pLM1461 was not planned but was isolated as an artefact of the Gibson Assembly approach used for the assembly of pLM1455. Colonies were screened by colony PCR using oLM1622/3448 and constructs with the correct amplicon were sequenced with the same primers.

Table 4-7: Primers used for cloning of α3-α4 SDM fragments.

Name	Purpose	Oligonucleotide Sequence
oLM3424	pLM1454 'insert' forward	AGCAGCGCGAGGATGATGAGCGTAAAGCC
oLM3425	pLM1454 'insert' reverse	GCCTTACGCTCATCATCCTCACGTTGCTCG
oLM3426	pLM1454 'backbone' forward	CGAGCAACGTGAGGATGATGAGCGTAAGGC
oLM3427	pLM1454 'backbone' reverse	GGCTTTACGCTCATCATCCTCGCGCTGCT
oLM3428	pLM1455 'insert' forward	GCTGCGCAGCGCGAGTTTCTTGAGGCTAAAG
oLM3429	pLM1455 'insert' reverse	CCTTAGCCTCCATAAACTCACGTTGCGCGGC
oLM3430	pLM1455 'backbone' forward	GCCGCGCAACGTGAGTTTATGGAGGCTAAGG
oLM3431	pLM1455 'backbone' reverse	CTTTAGCCTCAAGAAACTCGCGCTGCGCAGC
oLM1622	Colony PCR and sequencing	CACTACCTGAGCACCCAGTC
oLM3448	Colony PCR and sequencing	CCGACTTCAGCGAGACCG

4.6.4. Expression and purification of α3, α3-α4 and site-directed mutagenized α3-α4 fragments

Expression and purification of all fragments was completed consistent with the methods used for mEGFP-CsLinker and untagged CsLinker used in Chapter 3 except that instead of 10 kDa Amicon filtration units, 3 kDa units were used instead (UFC900324/UFC500324 – Merck).

4.6.5. Native PAGE experiments and quantification

Native PAGE experiments and densitometric analysis were also completed consistent with the methods outlined in Chapter 3. For experiments completed in the 'cryo' buffer (200 mM sorbitol, 50 mM HEPES, 50 mM KOAc, 2 mM $Mg(OAc)_2 \cdot 4H_2O$ and 1 mM $CaCl_2$ at pH 6.8) the Rubisco and α3-α4 fragment were buffer exchanged into the same buffer using Amicon 0.5 mL 3 kDa (for α3-α4) or 10 kDa (for Rubisco) centrifugal filter units. Glycerol was added to 10% (v/v) final concentration prior to loading on gels which were run at the same conditions as used previously.

4.6.6. Surface plasmon resonance (SPR)

SPR experiments were completed largely consistent with previously published methods for the *Chlamydomonas* EPYC1 fragment (He *et al.,* 2020) with minor modifications. All experiments were completed using a Biacore™ T100 fitted with a T200 upgrade kit (28977963 – Cytiva Life Sciences) with a sensor temperature of 25 °C and Series S CM5 sensor chips (BR100530 – Cytiva Life Sciences).

The flow path mode was sequential across all 4 flow channels (1-2-3-4) where flow path 1 was used as a blank reference for subtraction in all cases. Prior to use the chip was washed with a 1M NaCl solution made up in HBS-EP for 60 seconds contact time at 30 μL min^{-1}. An Amine Coupling Kit (BR100050 – Cytiva Life Sciences) was used for the modification of the chip surface for Rubisco immobilisation. For flow path 1 control modification and immobilisation reactions were performed. A 1:1 N-Hydroxysuccinimide (NHS)/1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) solution was injected for 420 seconds at 5 μL min^{-1}, followed by 420 seconds with HBS-EP buffer and 60 seconds with the 1M NaCl solution at the same injection rate. Free amine groups were subsequently capped by injection of a 1 M Ethanolamine hydrochloride-NaOH pH 8.5 solution. For the Rubisco flow paths (2-3-4), Rubisco samples were diluted in 10 mM sodium acetate (pH 4.5) to a final concentration of 0.1 mg mL^{-1} immediately prior to immobilisation. The same modification and immobilisation procedure as used for the blank reference flow path was used here, except instead of HBS-EP injection after NHS/EDC injection, the Rubisco sample was injected until the normalised response units in each flow path was ~5,000. The Rubisco flow paths were washed and capped in the same way. All flow paths were subsequently washed for 30 seconds with the 1 M NaCl solution. for which the running buffer was a modified HBS-EP buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.05% Tween-20).

For all SPR experiments presented in this work, flow path 2 was immobilised with *Chlamydomonas reinhardtii* Rubisco, flow path 3 with Chlorella Rubisco and flow path 4 with *Ulva mutabilis* Rubisco (see Chapter 5). The running buffer for all the presented experiments was the same buffer used for subsequent cryo-EM experiments, as well as in some of the native PAGE experiments (Fig. 3-5d) (200 mM sorbitol, 50 mM HEPES, 50 mM KOAc, 2 mM $Mg(OAc)_2 \cdot 4H_2O$ and 1 mM $CaCl_2$ at pH 6.8) 2-fold serial dilutions of the analytes (α3 and α3-α4) were made in the same buffer immediately prior to the start of the experiments. The fragments were initially buffer exchanged into the running buffer using Amicon 0.5 mL 3 kDa centrifugal filter units. For each of the three replicate experiments presented, each of the serial dilution concentrations were injected and dissociated once, starting with the lowest analyte concentration. For each analyte injection the flow was 15 μL min^{-1} for 30 seconds contact time, followed by 360 seconds of dissociation with the running buffer. After a complete set of serial dilution

concentrations had been injected and dissociated, the chip was washed with a 1M NaCl solution for 30 seconds, prior to equilibration with the running buffer for a further 360 seconds before the start of injection of the next set of serial dilutions in the next replicate.

The average of the raw response unit (R.U.) values at each analyte concentration, following subtraction of binding to the reference chip, were fitted with a least squares regression hill equation (($y(x) = B_{max}*x / (K_D + x)$), where B_{max} is the maximum specific binding, K_D is the apparent dissociation constant, and x is the analyte concentration) using GraphPad Prism 9. The raw data was normalised to the calculated B_{max} value from the initial fit of the raw data to give the relative SPR response reported in the text. The 95% confidence intervals (CI_{95}) of the K_D values were calculated using GraphPad Prism 9 in asymmetrical mode.

4.6.7. Cryogenic electron microscopy

4.6.7.1. Grid preparation

Chlorella Rubisco and the α3-α4 fragment were separately buffer exchanged, as described for SPR experiments, into cryo buffer (200 mM sorbitol, 50 mM HEPES, 50 mM KOAc, 2 mM $Mg(OAc)_2 \cdot 4H_2O$ and 1 mM $CaCl_2$ at pH 6.8) immediately prior to grid preparation. Chlorella Rubisco and α3-α4 were mixed in the same cryo buffer to final concentrations of 0.5 μM and 16 μM respectively. After mixing, the reaction was incubated at room temperature (~21 °C) for ~15 minutes as completed for the native PAGE experiments. R1.2/1.3 Cu 400-mesh grids (4240C – Quantifoil) were glow-discharged for 60 seconds with a current of 15 mA in a PELCO easiGlow system with the carbon side of the grid exposed. The glow-discharged grids were held in an FEI Mark IV Vitrobot (ThermoFisher Scientific) with a chamber temperature of 4 °C and a relative humidity setting of 95%. 2.5 μL of the α3-α4/Rubisco reaction mix was applied to the carbon side of the grids and incubated for 30 seconds prior to blotting for 8 seconds with a relative blot force of -5. The blotted grids were plunge frozen in liquid ethane and stored in liquid nitrogen before use. Prior to data collection the grids were clipped by Sam Hart (University of York cryo-EM research technician).

4.6.7.2. Data collection

Grids were loaded using a Cryo-Autoloader into the University of York 200 kV Glacios Cryo-Transmission Electron Microscope (Cryo-TEM) (ThermoFisher Scientific) equipped with a Falcon IV counting direct electron detector. Initial calibrations and Atlas collection were completed by Sam Hart, after which grids were screened for appropriate ice thickness using the EPU control software. Once an appropriate grid was selected, individual grid squares were screened for ice thickness and integrity, and holes were manually selected using the 'hole selection' tool in EPU. All screening was completed with a defocus of 2.0 µm.

EPU was used for subsequent automated data collection in aberration-free image shift (AFIS) mode. A nominal magnification of 240,000 x and electron fluence of 50 $e^-/Å^2$ with a calibrated pixel size of 0.574 $Å^2$ was used during collection in which each exposure was 6.52 seconds. The 100 µm objective aperture was inserted during collection and the C2 aperture was 50 µm. A range of defocus values were used (-1.8, -1.6, -1.4, -1.2, -1.0, -0.8 µm). Data collection was completed in a single overnight session (~14 hours) over which 4,447 micrographs were captured.

4.6.7.3. Data processing of the CsRubisco map

All image processing and 3D reconstruction was completed using Relion 3.1.2 (Scheres, 2012; Zivanov *et al.,* 2018) operated on the University of York Viking supercomputer cluster, as discussed in text. Raw EER files were imported and of the 1568 EER frames, 32 were grouped to give a fluence per frame of 1.02 $e^-/Å^2$. Relion's own implementation of MotionCor2 was used for motion correction in which a Bfactor of 150 was applied and a 5x5 patch pattern was used. CTFFIND4 (Rohou & Grigorieff, 2015) was used for CTF estimation in which the dose-weighting and summed power spectra from motion correction were used and a spherical aberration of 2.7 mm was assumed.

As discussed in text, 465 particles were manually picked from a selection of the higher defocus grids using a template size of 130 Å diameter. The extracted particles were 2x binned and reference-free 2D classification was completed into 50 classes with a regularisation parameter (T) of 1, over 50 iterations

with a soft circular mask of 150 Å diameter. Selected 2D classes (Fig. 4-8a), that represented 388 particles were used as 2D references for auto-picking in which they were lowpass filtered to 20 Å. Autopicking was completed with a picking threshold of 0.9, a minimum inter-particle distance of 90 Å and a maximum standard deviation of noise of 1.1 which resulted in the peaking of 237,035 particles. Extraction and reference-free 2D classification were completed using the same parameters as used for the manually picked particle set. Selected 2D classes (Fig. 4-8c) which represented 224,593 particles were used for *de novo* 3D initial model generation, in which C1 symmetry and a 150 Å diameter mask were used. 3D classification was completed on the same particles selected from the 2D classes, in which the *de novo* 3D model was lowpass filtered to 50 Å and used as a reference map. 3D classification was performed with C1 symmetry into 5 classes with a regularisation parameter (T) of 4 over 25 iterations with a soft spherical mask of 150 Å diameter. The selected 3D class (Fig. 4-9a), representing 73,962 particles, was subjected to 3D auto-refinement in which the lowpass filtered *de novo* 3D model was used as a reference and either C1 or D4 symmetry was imposed and a soft mask of diameter 150 Å was imposed. The resulting maps were post-processed with no sharpening factor (Fig. 4-9b). The D4 symmetry imposed post-processed map was subjected to CTF refinement in which beam tilt and 4^{th} order aberrations were estimated before another round of 3D auto-refinement with D4 imposed symmetry and post-processing. The resulting particles were subjected to another CTF refinement in which anisotropic magnification was estimated, prior to another round of 3D auto-refinement and post-processing with the same parameters. In the next round of CTF refinement, per-particle defocus and per-micrograph astigmatism were fitted and beamtilt was also estimated. 3D auto-refinement was again performed on the resulting particles. Bayesian polishing (Zivanov *et al.,* 2018) was performed using s_vel, s_div and s_acc values of 0.783, 3915 and 3.765 respectively, based on a training job completed on a previous dataset. A final D4 symmetry-imposed 3D auto-refinement job was completed on the polished particles, after which a post-processing job with sharpening *B* factor ($Å^2$) of 50 was used and resulted in the final CsRubisco map (Fig. 4-9c).

4.6.7.4. Data processing of the CsRubisco-α3-α4 map

The final polished particles from the CsRubisco map were D4 symmetry expanded using the relion_symmetry_expand function, resulting in a particle dataset of 591,696 effective particles. To create a mask for 3D classification of particles containing α3-α4 density, the predicted α3 helix of CsLinker was positioned in the additional density present in the C1 symmetry-imposed 3D class from the initial 3D classification of the CsRubisco map (Fig. 4-9a) in UCSF Chimera (Pettersen *et al.,* 2004). The UCSF Chimera molmap function was used to create a 5 Å surface map of this region, which was imported to Relion and a mask, lowpass filtered to 20 Å, with a binarization threshold of 0.05, extended 6 pixels with a soft edge of 5 pixels was created (Fig. 4-10a). An initial round of C1 symmetry-imposed 3D classification in which this mask and the final 3D auto-refined CsRubisco map, lowpass filtered to 50 Å and used as a reference, was performed on the symmetry expanded particle dataset. 5 classes and a regularisation parameter (T) of 20 were imposed across 25 iterations with a soft spherical mask of 160 Å diameter. The selected 3D class (Fig. 4-10b), representing 192,261 particles, was subjected to a second round of C1 symmetry-imposed 3D classification with the same parameters except the number of classes was reduced to 3. The selected 3D class (Fig. 4-10c), representing 133,171 particles, was subjected to a round of C1 symmetry-imposed 3D auto-refinement with the --skip_align argument. In this refinement a solvent mask, created from a 20 Å lowpass filtered map of the holoenzyme map, was used. A mask was created from the resulting 3D auto-refined map and lowpass filtered to 20 Å with a binarization threshold of 0.05, extension of 10 pixels and a soft edge of 6 pixels. The resulting mask was used in post-processing of the final 3D auto-refined CsRubisco-α3-α4 map, in which the sharpening *B* factor ($Å^2$) was -45 and resulted in the final CsRubisco-α3-α4 map (Fig. 4-10d). Local resolution estimation was completed using the relevant module in Phenix 1.2 (Adams *et al.,* 2010).

4.6.7.5. Model building, fitting, and refinement of the CsRubisco structure

The 3D structures of Chlorella RbcS and RbcL were predicted using AlphaFold2 (Jumper *et al.,* 2021) in ColabFold (Mirdita *et al.,* 2022). The 3D structures were aligned to the coordinates of the *Chlamydomonas* Rubisco cryo-EM structure (PDB: 7JN4; (He *et al.,* 2020)) using the MatchMaker

function in UCSF Chimera to create a model of the Chlorella Rubisco holoenzyme. The holoenzyme model was rigid body fitted into the final CsRubisco map in Chimera. The map and coordinates for one RbcS and one RbcL were imported to Coot v0.9.8.7 (Emsley *et al.,* 2010), in which flexible fitting was performed. The residues of flexible regions, for which no density in the map was observed were removed from the model. For the RbcL, residues 1-21, 60-78 and 461-475 were not visible, and for the RbcS, residues 1 and 77-82 were not visible. The flexible fitted coordinates were real-space refined in Phenix 1.20 (Adams *et al.,* 2010) using default settings with the addition of hydrogens prior to refinement. Rotamer outliers and clashes introduced during refinement were manually corrected in Coot and the Phenix cryo-EM validation module was used to reassess the model quality. After multiple rounds of correction, the map symmetry module was used to identify the symmetry axes in the final map, and the apply NCS operators module was used to expand the RbcS/RbcL model to recreate the holoenzyme. Any clashes introduced during this process were manually corrected in Coot. Finally, the douse module in Phenix was used to add waters to the final model. The CsRubisco holoenzyme map and model were submitted to the PDB with code 8Q04.

4.6.7.6. Model building, fitting and refinement of the CsRubisco-α3-α4 structure.

For the CsRubisco-α3-α4 model, the coordinates of one RbcS and one RbcL from the CsRubisco holoenzyme structure (PDB: 8Q04) were rigid body fitted into the final CsRubisco-α3-α4 map and imported to Coot. The predicted α3 helix of CsLinker was manually built into the additional density present on one of the RbcLs (Fig. 4-10d) in Coot, using the bulky density of the phenylalanine as a central anchor point. The built coordinates for the α3 helix were subsequently flexibly fitted in Coot also. As mentioned in text, the density for several side chains could not be resolved, and these were removed from the model prior to refinement (Fig. 4-12). Following this, the same refinement, manual correction, and validation processes as completed for the CsRubisco model were completed for this model in Phenix. The same NCS application process was also used, except that the 7 additional helices present after NCS application were removed from the final model. The same douse process was also completed. The final CsRubisco-α3-α4 map and model were submitted to the PDB with code 8Q04.

4.6.7.7. Determination of occupied symmetry axes

In the symmetry expansion process, each particle is duplicated and transformed to position each of the symmetry axes in the same orientation, such that the supplied mask can be used to identify additional density in the relative position of each axis. For Rubisco with D4 symmetry, this results in each particle having 8 symmetry axes which are individually assessed in the 3D classification process. To extract the number of symmetry axes used from each particle during the reconstruction, a custom python script which utilised the starfile package (https://github.com/teamtomo/starfile), was used to extract particle data from the particle.star file of the 3D class used for the recreation of the final CsRubisco-α3-α4 map. The X (rlnCoordinateX) and Y (rlnCoordinateY) coordinates as well as the Image name (rlnImageName) were used to extract the number of symmetry axes from each particle used in the final reconstruction. A histogram of the number of axes from each particle was plotted using GraphPad Prism 9 (Fig. 4-10h).

4.6.7.8. Figures

Structural figures were made using either UCSF Chimera or UCSF ChimeraX (Pettersen *et al.,* 2021).

4.6.8. HADDOCK molecular docking

For the CsLinker -Chlorella RbcS dock the HADDOCK 2.4 web server was used (van Zundert *et al.,* 2016). AlphaFold2 predicted structures of the Chlorella RbcS and residues 164-186 of CsLinker (α3). The dock was constrained to use the helix of the CsLinker fragment (residues 168-182) and the surface helices of Chlorella RbcS (residues 23-35 and 86-99) as 'active' residues. The settings were otherwise left default. The highest scoring (lowest energy) complex is presented in text (Fig. 4-16).

Table 4-8: Plasmids used in Chapter 4.

Name	Description	Use	Origin
pLM872	6xHis-mEGFP-TEV-CsLinker-TEV-MBP-6xHis	Cloning of pLM1286/pLM1287	This study
pLM1286	6xHis-mEGFP-TEV-α3-TEV-MBP-6xHis	Purification of α3 fragment	This study
pLM1287	6xHis-mEGFP-TEV-α3-α4-TEV-MBP-6xHis	Purification of α3-α4 fragment and cloning of variants	This study
pLM1454	6xHis-mEGFP-TEV-α3-α4(hydrophobic)-TEV-MBP-6xHis	Purification of hydrophobic α3-α4 mutant	This study
pLM1455	6xHis-mEGFP-TEV-α3-α4(electrostatic)-TEV-MBP-6xHis	Purification of electrostatic α3-α4 mutant	This study
pLM1461	6xHis-mEGFP-TEV-α3-α4(partial electrostatic)-TEV-MBP-6xHis	Purification of partial electrostatic α3-α4 mutant	This study

5. CSLINKER PROMISCUOUSLY PHASE SEPARATES GREEN LINEAGE RUBISCOS

5.1. Introduction

In the previous chapter, the 3D structure of a CsLinker fragment in complex with CsRubisco demonstrated a binding site on the large subunit of Rubisco (RbcL). This was a tantalising discovery with respect to developing a plant Rubisco compatible linker protein, that could shortcut efforts to engineering pyrenoids in crop plants for several reasons (Adler *et al.,* 2022).

Firstly, the large subunit of Rubisco is universally encoded on the chloroplast genome (plastome) (Vitlin Gruber & Feiz, 2018), overwhelmingly with only a single gene copy present in each plastome. This is in contrast to the RbcS, which in plants is nuclear-encoded and present in multiple non-homologous gene copies (Mao *et al.,* 2022). This is significant as previous efforts to engineer pyrenoids in *Arabidopsis* have required the replacement of the native nuclear-encoded RbcS genes with those of Chlamydomonas, to allow interaction with EPYC1 (Atkinson *et al.,* 2017, 2019, 2020). Though this approach has been successful in allowing Rubisco condensation *in planta* into a 'proto-pyrenoid' (Atkinson *et al.,* 2020), the feasibility in crop species where large numbers of RbcS genes (*e.g.,* 25 in bread wheat) are present (Mao *et al.,* 2022) will likely be hindered by the huge engineering effort of replacing the multiple RbcS genes as well as the complexity of crossing multiple edited genomic loci into the elite crop lines (Moose & Mumm, 2008). As such, a large subunit binding linker protein offers an exciting alternative as only a single gene copy in plants may need to be edited to allow cross-reactivity with the linker, and this can be completed in a single step by homologous recombination (Kanevski *et al.,* 1999; Lin *et al.,* 2021; Whitney *et al.,* 2015; Zhou *et al.,* 2023). However, in many crop plants chloroplast genome engineering is not yet established.

Secondly, the sequence and structure of the large subunit of Rubisco is highly conserved across green lineage species (Fig. 5-14), given its phylogenetic origin in a hypothetical ancestral flagellate that was the most recent common ancestor to all present day green lineage species (Leliaert *et al.,* 2016; Shih *et al.,* 2016). Again, this contrasts the RbcS genes which demonstrate much lower sequence conservation

across species given their nuclear encoding in the green lineage (Dean *et al.,* 1989; Spreitzer, 1993; Spreitzer & Salvucci, 2002), and more recent evolution (Shih *et al.,* 2016). The sequence conservation of the RbcL therefore offers the prospect that the CsLinker could either natively, or with slight modification, itself exist as a plant-compatible linker protein. To this end, in this body of work the primary objective was to explore if the RbcL binding of CsLinker affords any increased cross-reactivity relative to the RbcS binding of EPYC1.

5.2. *Characterising in vitro species cross-compatibility of linker proteins from Chlorella and Chlamydomonas*

The most obvious place to start in the journey of characterising the relative cross-compatibility of Rubisco linker proteins was with a direct comparison between CsLinker of Chlorella and EPYC1 of Chlamydomonas. Understanding the molecular details of the interactions of both proteins with their cognate Rubiscos was essential to this process.

5.2.1. *Direct comparison of EPYC1 and CsLinker binding sites*

A comparison of the RbcS binding site of EPYC1 and RbcL binding site of CsLinker in the Rubisco sequences from Chlamydomonas and Chlorella demonstrated the expected high conservation of CsLinker interacting residues in the RbcL relative to the less conserved EPYC1 interacting residues in the RbcS (Figure 5-1a,d). Analysis of the 3D structures of Chlamydomonas and Chlorella RbcS relative to the binding site of EPYC1 indeed confirmed that the electrostatic interactions between EPYC1's Arg64 and the position 24 residue of the Chlorella RbcS would not be possible due to substitution from a Glu to an Ala residue (Fig. 5-1b). Additionally, the hydrophobicity of the binding interface was shown to be markedly reduced in the Chlorella RbcS due to the substitution of Chlamydomonas Met87 and Val94 with the less hydrophobic Gly87 and Gln94 residues respectively (Fig. 5-1c). In contrast, the CsLinker binding site in the RbcL showed total sequence conservation between the two species (Fig. 5-1d) as well as good positional conservation of the relevant residues in the 3D structures (Fig. 5-1e,f). Both the electrostatic interactions and hydrophobic interface were shown to be almost indistinguishable between the Chlorella and Chlamydomonas RbcL structures, presumably allowing equivalent interaction of both Rubiscos with CsLinker. Together these comparisons suggested a greater likelihood that the CsLinker would be cross-compatible with Chlamydomonas Rubisco than vice versa with EPYC1.

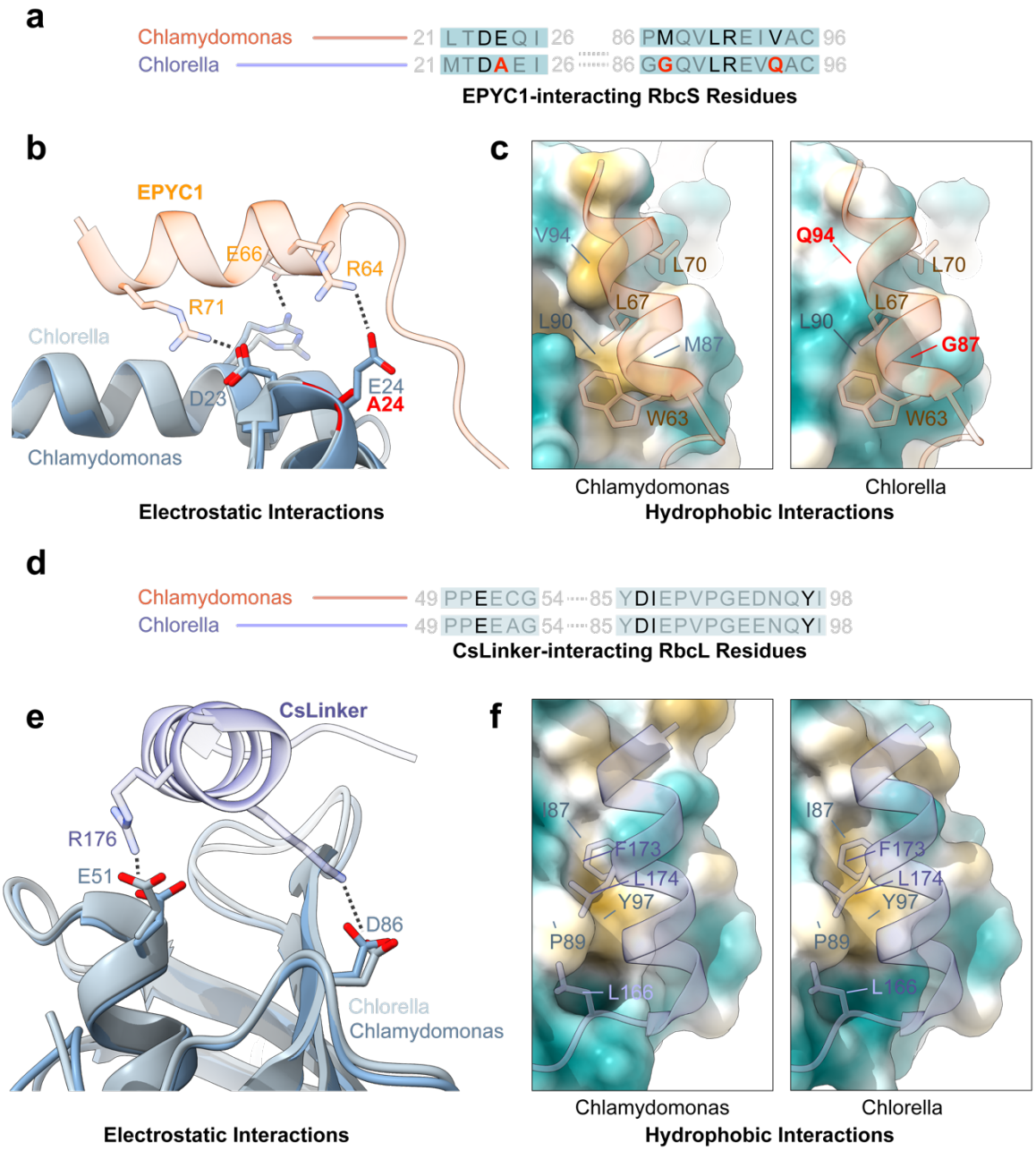


Figure 5-1 | Comparison of Rubisco interaction interfaces of EPYC1 and CsLinker.
(a) Alignment of the EPYC1 interaction interfaces of the *Chlamydomonas* RbcS and equivalent region in the Chlorella RbcS. Interacting residues in the *Chlamydomonas* sequence are shown in black, with non-conserved equivalent residues in the Chlorella sequence shown in red. (b) Electrostatic interactions at the EPYC1-RbcS interface. The structure of the *Chlamydomonas* RbcS (7JFO; (He *et al.,* 2020)) and Chlorella RbcS (8Q04; Chapter 4) are displayed overlaid. The electrostatic interactions are indicated by dashed lines. The non-conserved Ala residue is shown in red. (c) Hydrophobic interactions in the *Chlamydomonas* (left) and Chlorella (right) RbcS. The RbcS is shown as a surface representation coloured by hydrophobicity calculated according to Kyte-Doolittle scoring. Non-conserved residues in the RbcS are shown in red. (d) Alignment of the CsLinker interaction interfaces in the RbcL. Interacting residues are shown in black. (e) Electrostatic interactions at the CsLinker-RbcL interface. The structure of the *Chlamydomonas* RbcL and Chlorella RbcL are displayed overlaid. The electrostatic interactions are indicated by dashed lines. (f) Hydrophobic interactions in the Chlamydomonas (left) and Chlorella (right) RbcL. The RbcL is shown as a surface representation coloured by hydrophobicity calculated according to Kyte-Doolittle scoring.

5.2.2. *Biochemical characterisation of CsLinker fragment interaction with Chlamydomonas Rubisco*

To determine if the apparently equivalent interaction interface in the 3D structure of Chlamydomonas Rubisco (Fig. 5-1) translated to an equivalent interaction with the CsLinker, a characterisation of the interaction with the previously generated fragments of CsLinker was completed. Native PAGE band shift assays showed a similar concentration-dependent shift with both Chlamydomonas and Chlorella Rubisco albeit with less definite complex formation in the Chlamydomonas sample (Fig. 5-2a). Brightfield microscopy and droplet sedimentation assays indicated that no visible condensation or aggregation occurred in the solution (Fig. 5-2b,c), as previously observed for the Chlorella Rubisco (Fig. 4-5a,b). In agreement with these observations, the apparent K_D of the α3 and α3-α4 fragments for Chlamydomonas Rubisco was almost indistinguishable from that of Chlorella Rubisco, as determined by SPR experiments that were completed in tandem on the same sensor chip as the previously present Chlorella SPR results (108 μM [CI_{95}: 101–115 μM] for α3 and 1.30 μM [CI_{95}: 1.15–1.47 μM] for α3-α4) (Fig. 5-3 and Fig. 4-6). Together this evidence quantitively agreed with the structural comparison that suggested an equivalent interaction between CsLinker and both Chlorella and Chlamydomonas Rubiscos (Fig. 5-1d-f), and hinted cross-reactivity of the full length CsLinker protein.

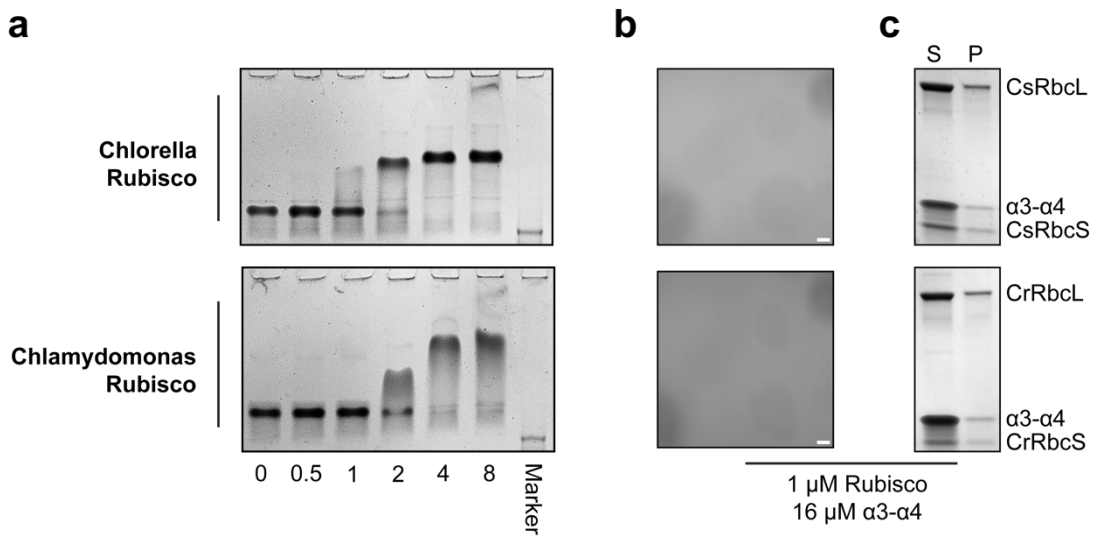


Figure 5-2 | Biochemical cross-reactivity of the α3-α4 fragment with Chlamydomonas Rubisco.
(a) Comparison of tandem native PAGE band shift assays completed with the α3-α4 CsLinker fragment and Chlorella (top) and *Chlamydomonas* (bottom) Rubiscos. (b) Brightfield microscopy of α3-α4-Rubisco solutions at the indicated concentrations with each Rubisco according to Fig. 5-2a. (c) Droplet sedimentation assays of the solutions in Fig. 5-2b at the same concentrations. Data for Chlorella Rubisco in Fig. 5-2b and Fig. 5-2c is also presented in Fig. 4-5 but are presented again here for clarity.

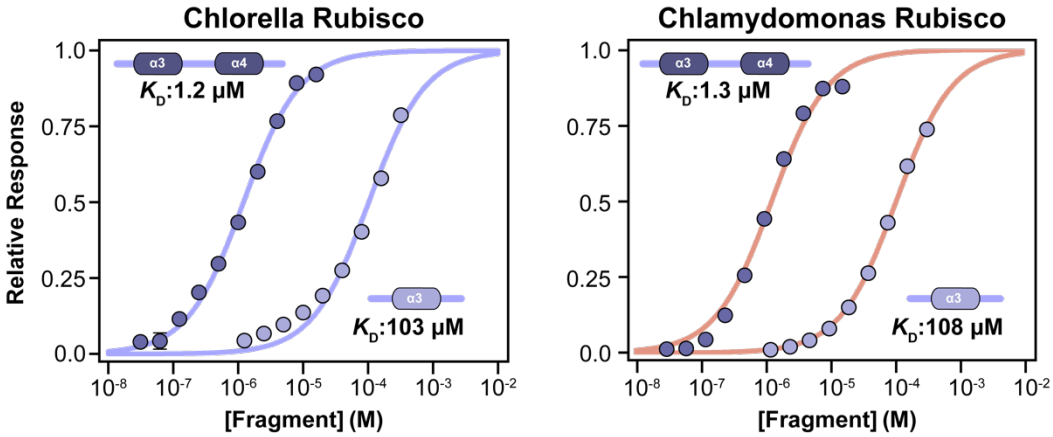


Figure 5-3 | SPR quantification of cross-reactivity of CsLinker fragments with Chlamydomonas Rubisco. SPR response normalised to Y_{max} value obtained from fit of raw data for Chlorella (left) and Chlamydomonas (right) Rubisco; n=3, error bars = S.D. (see methods). The Chlorella data was presented previously in Fig. 4-6 but is presented again here for direct comparison.

5.2.3. In vitro comparison of phase separation cross-compatibility of EPYC1 and CsLinker with Chlamydomonas and Chlorella Rubisco

Based on the cross-reactivity of the CsLinker fragments, it was expected that Chlamydomonas Rubisco would also phase separate with the full length CsLinker protein. Although no quantitative biochemical characterisation of EPYC1 fragment interaction with Chlorella Rubisco was completed, based on structural analysis it was hypothesised Chlorella Rubisco would not interact with the same affinity and consequently likely not phase separate with EPYC1 (Fig. 5-1a-c). Indeed, at the same concentrations observed to induce phase separation of Chlorella Rubisco, Chlamydomonas Rubisco was also readily phase separated (Fig. 5-4a,b). Reciprocally, EPYC1 was unable to induce phase separation of Chlorella Rubisco at conditions under which Chlamydomonas Rubisco was readily phase separated (Fig. 5-4c,d). For experiments with EPYC1, a ratio of 1:1 EPYC1:Rubisco was used based on the *in vivo* relative abundances of the proteins (Hammel *et al.,* 2018), though the Rubisco concentration remained consistent with that of the Chlorella solution.

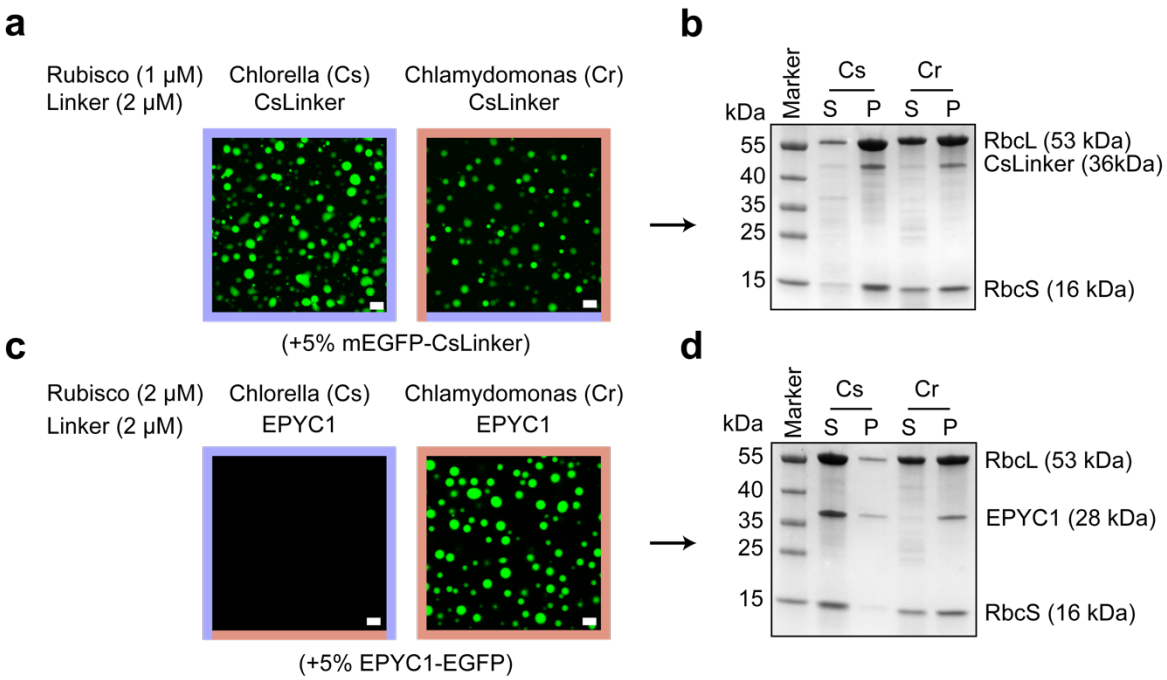


Figure 5-4 | Cross-compatibility of full length CsLinker and EPYC1 with *Chlamydomonas* and Chlorella Rubiscos.
(a) Confocal fluorescence microscopy images of droplets formed at the indicated conditions with CsLinker ~10 minutes post-formation. mEGFP-tagged CsLinker was added to 5% molar ratio to allow visualisation. Scale bar = 5 μm. (b) SDS-PAGE analysis of droplet sedimentation assays of the solutions formed under the same conditions as in Fig. 5-4a. (c) Confocal fluorescence microscopy images as in Fig. 5-4a formed with EPYC1, with 5% EPYC1-mEGFP added. (d) SDS-PAGE analysis of droplet sedimentation assays of solutions formed under the same conditions as in Fig. 5-4c.

5.3. Characterising in vivo species cross-compatibility of CsLinker in Chlamydomonas

The *in vitro* cross-reactivity of the CsLinker, presumably underpinned by its utilisation of the same CsLinker interface (Fig. 5-1d-f), appeared to show similar behaviour of CsLinker and EPYC1 despite their distinct evolutionary origins. To probe whether this *in vitro* behaviour translated to analogous functionality it was desirable to empirically validate this hypothesis *in* vivo. Previous work developed a Chlamydomonas strain which lacks the native linker protein EPYC1 (ΔEPYC1), and thus is pyrenoid-less and unable to grow in ambient CO_2 conditions due to the lack of a functional pyrenoid-based CCM (Mackinder *et al.,* 2016). This strain provided a unique opportunity to interrogate whether the *in vitro* cross-reactivity of CsLinker translates to functional *in vivo* cross-compatibility, using functionality of the Chlamydomonas CCM as a phenotypic readout.

Crucially with respect to this characterisation, the high CO_2-requiring ΔEPYC1 strain developed by Mackinder *et al.* (2016) remains unedited with respect to other essential CCM and pyrenoid-assembly components and is only ablated in formation of the Rubisco condensate. As discussed in Chapter 1, assembly of the Chlamydomonas Rubisco condensate with the pyrenoid's ultrastructural features is hypothesised to be driven by interaction with the same RbcS-interaction interface utilised by EPYC1 (Itakura *et al.,* 2019; Meyer, Itakura, *et al.,* 2020), which also remains unedited in the ΔEPYC1 strain. It was therefore hypothesised that if CsLinker could recover formation of the condensate of Chlamydomonas Rubisco in ΔEPYC1 as observed *in vitro*, retention of the Chlamydomonas RbcS would allow for the correct assembly and function of the pyrenoid-based CCM with the rest of its functional components.

5.3.1. Generation of a ΔEPYC1ΔRbcS strain

To probe this hypothesis, it was desirable to generate a strain in the same ΔEPYC1 background that also lacks the native Chlamydomonas RbcS interface and should be unable to recover a functional pyrenoid-based CCM, even if the Rubisco condensate is recapitulated. At the time the objective of generating this strain was set, CRISPR manipulation of Chlamydomonas was very much in its infancy

in comparison to recent developments (Chen, Yang, *et al.,* 2023; Nievergelt *et al.,* 2022). As such, a genetic crossing approach using previously characterised strains was pursued.

To this end, paromomycin-resistant ΔEPYC1 strain CC-5360 (mt-) was crossed with zeocin-resistant ΔRbcS strain CC-4691 (mt+), and the ΔEPYC1ΔRbcS strains were recovered by growth on dual selection plates (see methods). PCR screening indicated successful genetic crossing occurred in 7 of the 8 ΔEPYC1ΔRbcS colonies screened (Fig. 5-5). Subsequent western blot analysis of 6 of the genetically crossed strains demonstrated clear RbcS knockout, though EPYC1 could not be detected in the positive control WT (CC-4533) or ΔRbcS strains (Fig. 5-6a). It was presumed that EPYC1 was either not expressed or not detectable due to the necessity of growing the ΔRbcS and ΔEPYC1ΔRbcS strains in heterotrophic media (TAP) in the dark due to the lines not being able to grow photosynthetically in the absence of Rubisco functionality. To reconcile this, it was desirable to grow the strains under photosynthetic conditions in which EPYC1 is expressed (Turkina *et al.,* 2006). In order to recover photosynthetic growth of the ΔEPYC1ΔRbcS and ΔRbcS strains complementation with a WT copy of Rbcs2 was completed. For these transformations the strains were recovered on minimal media (TP) plates under 3% CO_2 conditions to select for successful recovery of photosynthetic growth by complementation of Rubisco function. Blotting demonstrated successful reintroduction of the RbcS protein in both the ΔEPYC1ΔRbcS and ΔRbcS strains with a range of expression levels, in line with genomic 'position effects' due to random integration of the expression cassette (Fig. 5-6b) (Baier *et al.,* 2018). The resulting strains (ΔEPYC1ΔRbcS::RbcS2 and ΔRbcS::Rbcs2) could then be grown under photosynthetic conditions (TP +3% CO_2 for 3 days, then switched to TP + ambient CO_2 for 12 hours) and blotted for the absence of EPYC1 in comparison to WT and ΔEPYC1 strains grown under the same conditions. EPYC1 blots clearly demonstrated the absence of EPYC1 in the expected strains and allowed conclusion that EPYC1 was removed in the parent ΔEPYC1ΔRbcS strains prior to RbcS2 reintroduction (Fig. 5-6c). Strain ΔEPYC1ΔRbcS-3 was taken forward as the sole ΔEPYC1ΔRbcS strain for further analyses.

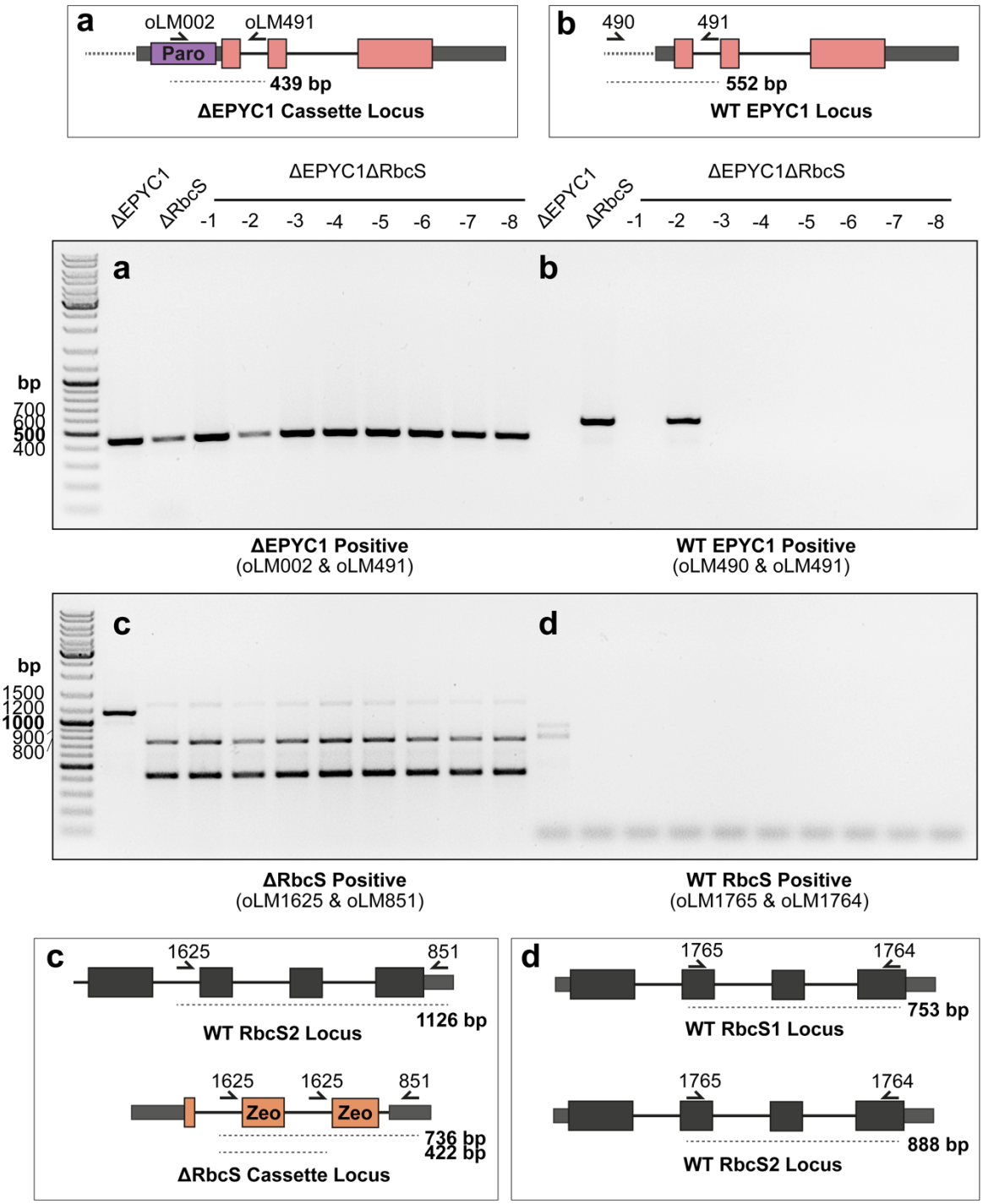


Figure 5-5 | PCR screening for genetic crossing in the ΔEPYC1ΔRbcS strain.
(a) PCR screening of the ΔEPYC1 locus as indicated by the schematic (top), and the corresponding gel electrophoresis below. (b) Screening for the WT EPYC1 locus. Insertion of the paromomycin cassette (~3 kb) ablates amplification of the product in the ΔEPYC1 locus. (c) Screening for ΔRbcS. Due to the presence of multiple Rbcs2 intron 1 copies in the Zeocin resistance cassette in the ΔRbcS strain, the same primers used for the WT locus can be used for the ΔRbcS locus. (d) Screening for the WT RbcS loci. The highly similar RbcS1 and RbcS2 gene copies can be screened for using the same primer set.

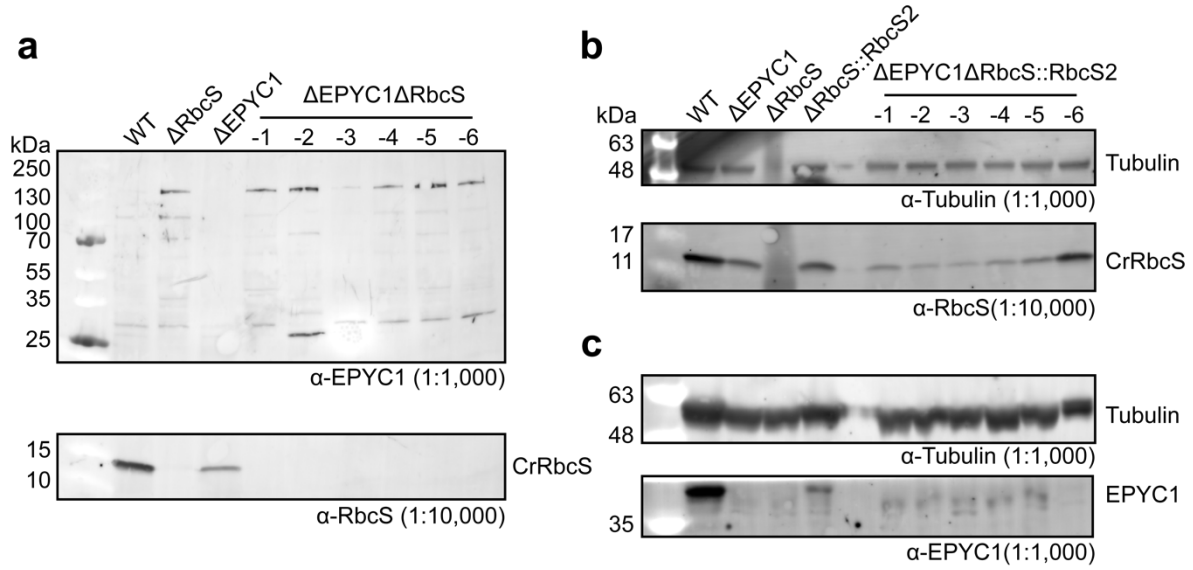


Figure 5-6 | Western blot screening for EPYC1 and RbcS knockout in crossed strains.
(a) Western blot analysis of the lysate of indicated stains grown in heterotrophic conditions in the dark. (b) Western blot analysis of RbcS2 complemented strains and origin strains grown in heterotrophic conditions in the dark. (c) Western blot analysis of the same strains as in Fig. 4-6b grown in photosynthetic conditions.

5.3.2. Recovery of native CCM function in the ΔEPYC1ΔRbcS strain

It has previously been shown that reintroduction of EPYC1 recovers the high-CO_2 requiring phenotype of the ΔEPYC1 strain (Mackinder *et al.,* 2016). To ensure appropriate tractability of the generated ΔEPYC1ΔRbcS strain, recovery of the native pyrenoid-based CCM was also attempted in this strain. To this end, EPYC1 was reintroduced into the previously generated ΔEPYC1ΔRbcS::RbcS2 strain, using the same plasmid previously used for ΔEPYC1 complementation (Mackinder *et al.,* 2016). Spot test analysis demonstrated the recovery of photosynthetic growth at high and low CO_2 in the ΔEPYC1ΔRbcS::RbcS2 and ΔEPYC1ΔRbcS::RbcS2::EPYC1 strains respectively (Fig. 5-7). Recovery of low CO_2 growth in the ΔEPYC1 strain was also recapitulated. Interestingly, even after complementing the CC-4691 ΔRbcS strain with RbcS2, the strain still showed a light sensitive phenotype, with very little growth even under 3% CO_2 conditions. This phenotype was somewhat recovered in the equivalent ΔEPYC1ΔRbcS::RbcS2 line indicating that the CC-4691 strain likely possesses additional mutations that render it light sensitive, which were crossed out in the creation of the ΔEPYC1ΔRbcS line. Further, comparison of the growth of WT and ΔEPYC1::EPYC1 lines with the ΔEPYC1ΔRbcS::RbcS1::EPYC1 line indicates that the strain is not of completely comparable health. Given the highly sensitive phenotype of CC-4691, it is likely some defects were inherited from

this strain during crossing. Regardless, taken together the evidence indicated that CCM function and presumably pyrenoid assembly can be appropriately recovered in the ΔEPYC1ΔRbcS strain.

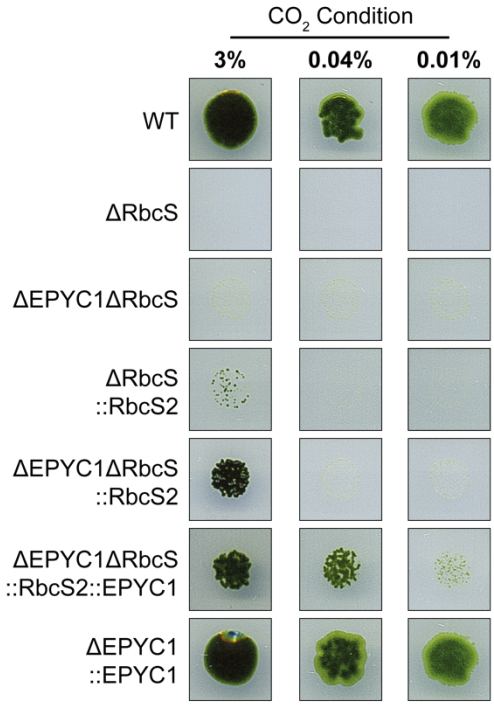


Figure 5-7 | Spot test analysis of CCM recovery in the ΔEPYC1ΔRbcS strain.
Images of spot test plates after 96 hours of growth of $1x10^5$ cells at the indicated CO_2 condition at pH 7.4.

5.3.3. Expression of Chlorella RbcS (CsRbcS) in the ΔEPYC1ΔRbcS strain

Having demonstrated that photosynthetic growth can be recovered in the ΔEPYC1ΔRbcS strain, the next goal was to introduce a foreign RbcS that lacks cross-compatibility with EPYC1. As it was previously demonstrated the Chlorella RbcS lacks cross-reactivity with EPYC1 *in vitro* (Fig. 5-4), this seemed an obvious choice. Initial attempts to express a codon-optimised Chlorella RbcS with the Chlamydomonas RbcS2 transit peptide and introns inserted at conserved positions, using the promoter and terminator pair previously used for Chlamydomonas RbcS expression (pAR/tRbcS2) (Caspari, 2020) was unsuccessful as compared to the equivalent Chlamydomonas sequence (Table 5-1). Genkov *et al.* (2010) reported similar results when trying to recover photosynthetic growth with a Spinach RbcS gene in a different ΔRbcS strain. This was accredited to a Met to Lys substitution present in position 2 of the mature sequence of the small subunit of Spinach, which may affect cleavage of the transit peptide as reported previously (Invernizzi *et al.,* 2002; Su *et al.,* 1999). When the Gln in position 2 of the mature

Chlorella RbcS was mutated back to a Met, the efficiency of the transformations was recovered to levels comparable to the WT Chlamydomonas sequence using the same promoter terminator pair (Table 5-1). Notably, the strong promoter and terminator pair of PSAD was apparently unable to drive expression of the RbcS even following the Q2M mutation.

Table 5-1: Summary of transformation efficiency of Chlorella RbcS constructs.

Construct	Description	Efficiency (%)
pLM1109	Codon-optimised, CrRbcS2 transit peptide and introns, pAR/tRBCS2	0.39
pLM1122	pLM1109 with **Q2M** mutation in mature RbcS2 sequence	116
pLM1118	pLM1122 with **pPSAD/tPSAD** instead of pAR/tRBCS2	0.25

To select the most appropriate strain to take forward, 96 colonies were screened for their phototrophic growth under CO_2-replete conditions for growth rate using a high throughput plate-based approach in which colony size was recorded over time (Fig. 5-8a) (see methods). Growth rate was reasonably consistent between colonies and the fastest growing colony after removal of outliers was selected for further characterisation and denoted ΔEPYC1ΔRbcS::CsRbcS (Fig. 5-8b, red).

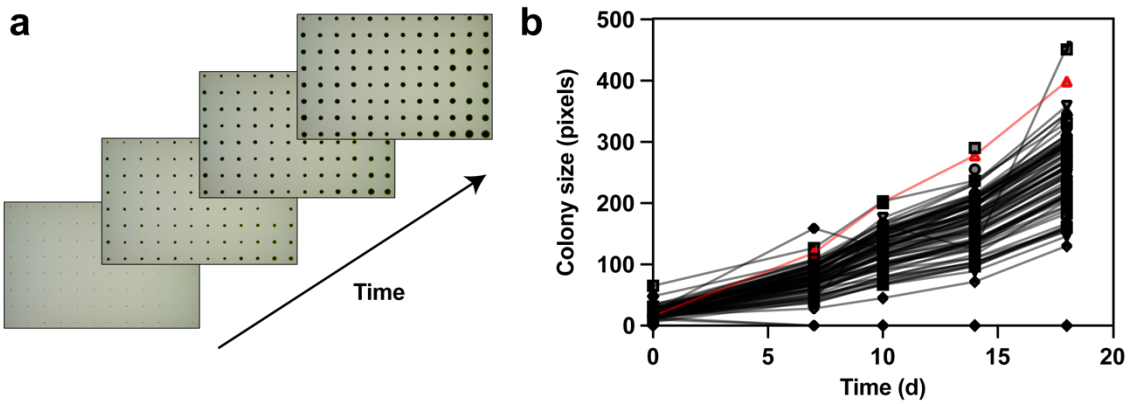


Figure 5-8 | Screening the growth rate of ΔEPYC1ΔRbcS strain expressing Chlorella RbcS.
(a) Representation of screening approach. Colony size was quantified at each timepoint and used to plot points in Fig. 1-8b. (b) Colony size plotted over time. The selected strain is indicated in red.

5.3.4. Optimisation of CsLinker expression in Chlamydomonas

With both the ΔEPYC1 (expressing WT Chlamydomonas RbcS) and ΔEPYC1ΔRbcS::CsRbcS (expressing Chlorella RbcS) strains established, the next challenge was to express the CsLinker protein in both strains. To probe for species cross-compatibility of CsLinker in Chlamydomonas ΔEPYC1 strains, it was assumed that expression of CsLinker would have to be of a comparable level to that of EPYC1 in the WT strain. Given the nuclear encoding of EPYC1 in WT Chlamydomonas, a nuclear transformation approach was pursued initially. The mature CsLinker coding sequence was codon-optimised for Chlamydomonas nuclear expression, had CrRbcS2 introns inserted and was fused to the EPYC1 chloroplast transit peptide. To allow for rapid on-plate screening of colonies, the coding sequence was also fused to a C-terminal mVenus tag. All 18 combinations of non-inducible promoter/terminator pairs from the Chlamydomonas MoClo toolkit were used (Crozet *et al.,* 2018), as well as the vector used for expression of EPYC1 previously (Mackinder *et al.,* 2016) (see methods for cloning details). The resulting level 1 vectors were assembled separately at level 2 alongside a hygromycin resistance cassette before transformation into ΔEPYC1. Although thousands of hygromycin resistant colonies were obtained in each of the transformations, none showed detectable fluorescence by plate scanning (data not shown). These results indicated expression of the gene is likely being silenced, as is often the case with the expression of transgenes in Chlamydomonas (Baier *et al.,* 2020; Neupert *et al.,* 2020; Schroda, 2019). Alternatively, processing and import of the protein into the chloroplast could also have been affected.

Table 5-2: Summary of nuclear expression vectors assembled for CsLinker expression in Chlamydomonas.

Number	Promoter	Terminator	Source
1	PSAD (pCM0-016)	PSAD (pCM0-114)	(Crozet *et al.,* 2018)
2	PSAD (pCM0-016)	RBCS2 (pCM0-115)	(Crozet *et al.,* 2018)
3	PSAD (pCM0-016)	RPS29 (pCM0-116)	(Crozet *et al.,* 2018)
4	PSAD (pCM0-016)	CA1 (pCM0-117)	(Crozet *et al.,* 2018)
5	PSAD (pCM0-016)	βTUB2 (pCM0-118)	(Crozet *et al.,* 2018)
6	PSAD (pCM0-016)	RPL23 (pCM0-119)	(Crozet *et al.,* 2018)
7	AR (pCM0-017)	PSAD (pCM0-114)	(Crozet *et al.,* 2018)
8	AR (pCM0-017)	RBCS2 (pCM0-115)	(Crozet *et al.,* 2018)
9	AR (pCM0-017)	RPS29 (pCM0-116)	(Crozet *et al.,* 2018)
10	AR (pCM0-017)	CA1 (pCM0-117)	(Crozet *et al.,* 2018)
11	AR (pCM0-017)	βTUB2 (pCM0-118)	(Crozet *et al.,* 2018)
12	AR (pCM0-017)	RPL23 (pCM0-119)	(Crozet *et al.,* 2018)
13	βTUB2 (pCM0-021)	PSAD (pCM0-114)	(Crozet *et al.,* 2018)
14	βTUB2 (pCM0-021)	RBCS2 (pCM0-115)	(Crozet *et al.,* 2018)
15	βTUB2 (pCM0-021)	RPS29 (pCM0-116)	(Crozet *et al.,* 2018)
16	βTUB2 (pCM0-021)	CA1 (pCM0-117)	(Crozet *et al.,* 2018)
17	βTUB2 (pCM0-021)	βTUB2 (pCM0-118)	(Crozet *et al.,* 2018)
18	βTUB2 (pCM0-021)	RPL23 (pCM0-119)	(Crozet *et al.,* 2018)
19	PSAD	CytC6	(Mackinder *et al.,* 2016)

One approach to circumvent both silencing and processing problems for expression of proteins that are targeted to the chloroplast is to move the expression location of the genetic material from the nuclear to the chloroplast genome (Jackson et al. 2021). By the advice of René Inckemann (Erb Group – MPI Marburg), based on results from the ongoing development of a large synthetic biology toolkit for the Chlamydomonas chloroplast, the intron-less CsLinker mature coding sequence was codon-optimised for chloroplast expression and assembled with a chloroplast codon-optimised N-terminal mVenus tag into a vector where expression was driven by the p16s promoter, psaA 5'UTR and psbA 3'UTR. An analogous vector in which the mVenus was replaced with a 3xFLAG tag was also assembled. The vectors were transformed by particle bombardment into the ΔEPYC1 and ΔEPYC1ΔRbcS::CsRbcS strains. Following 8 days of recovery on spectinomycin selection plates, most of the resulting colonies transformed with the mVenus-CsLinker vector showed peripheral fluorescence consistent with increasing expression of the inserted gene over time due to propagation of the genetic material by homologous recombination throughout the multiple copies of the chloroplast genome (Fig. 5-9b) (Karcher et al. 2009). Western blot analysis of resulting strains indicated expression of proteins with the expected molecular weight (Fig. 5-9a). Notably, expression of the constructs in the ΔEPYC1ΔRbcS::CsRbcS strain appeared more variable. Detection of the CsLinker protein by western blot proved challenging and is not discussed in detail here. Anecdotally it appeared the transfer of the protein from gel to blotting membrane was the limiting factor, which was somewhat mitigated by nitrocellulose membrane and a high voltage transfer system (iBlot 2). Addition of the mVenus tag also appeared to make the protein more readily detectable, possibly due to a shift in the isoelectric point of the fusion protein. Despite these challenges, by shifting expression to the chloroplast genome it was possible to achieve detectable levels of expression of CsLinker in Chlamydomonas, which could be used for further characterisation.

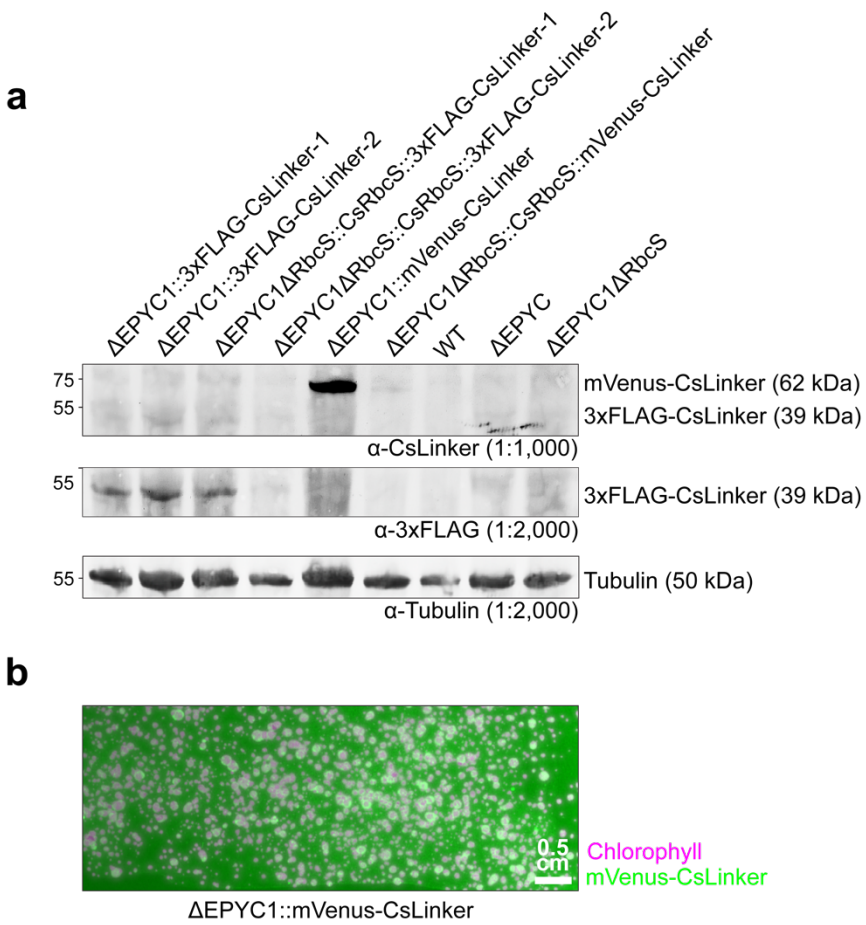


Figure 5-9 | Analysis of expression of CsLinker in Chlamydomonas strains.
(a) Western blot analysis of the lysate of Chlamydomonas strains transformed with the indicated tagged CsLinker constructs relative to the background strains. (b) Typhoon 8610 plate scan of a representative ΔEPYC1 transformation plate at day 8 showing mVenus fluorescence. Chlorophyll and mVenus channels are overlaid. Scale bar = 0.5 cm.

5.3.5. *Analysis of Rubisco condensation by CsLinker in Chlamydomonas*

To determine if CsLinker was able condense Chlamydomonas Rubisco as observed *in vitro* (Fig. 5-4), the ΔEPYC1 and ΔEPYC1ΔRbcS::CsRbcS strains expressing mVenus-CsLinker were analysed by confocal microscopy. In the ΔEPYC1::mVenus-CsLinker lines, a single apparent condensate of micron scale was observed in the canonical pyrenoid position (Fig. 5-10a). In contrast, in the ΔEPYC1ΔRbcS::CsRbcS strain multiple apparent condensates were observed, but were not positioned at the canonical pyrenoid position, though were often proximal to it (Fig. 5-10b). From this data it was tempting to conclude that the absence of the native Chlamydomonas RbcS, and hence the interaction interface for the native Chlamydomonas pyrenoid-assembly proteins, was the reason for the mis-positioning of the condensate in the chloroplast. However, when the reduced mVenus-CsLinker expression level (Fig. 5-9) in the ΔEPYC1ΔRbcS::CsRbcS and strain health (Fig. 5-7) of the parent ΔEPYC1ΔRbcS strains are considered, the confidence in this conclusion is weak. Based on the mis-positioning of the condensate and difficulty with comparable expression levels, further work with the ΔEPYC1ΔRbcS::CsRbcS strain was abandoned as it was assumed comparisons with ΔEPYC1 and WT strains would not be valid. Future work to develop a healthy ΔEPYC1ΔRbcS parent strains that would allow more direct comparison should be prioritised as this would allow a broad range of pyrenoid assembly hypotheses to be assessed *in vivo*, as discussed later.

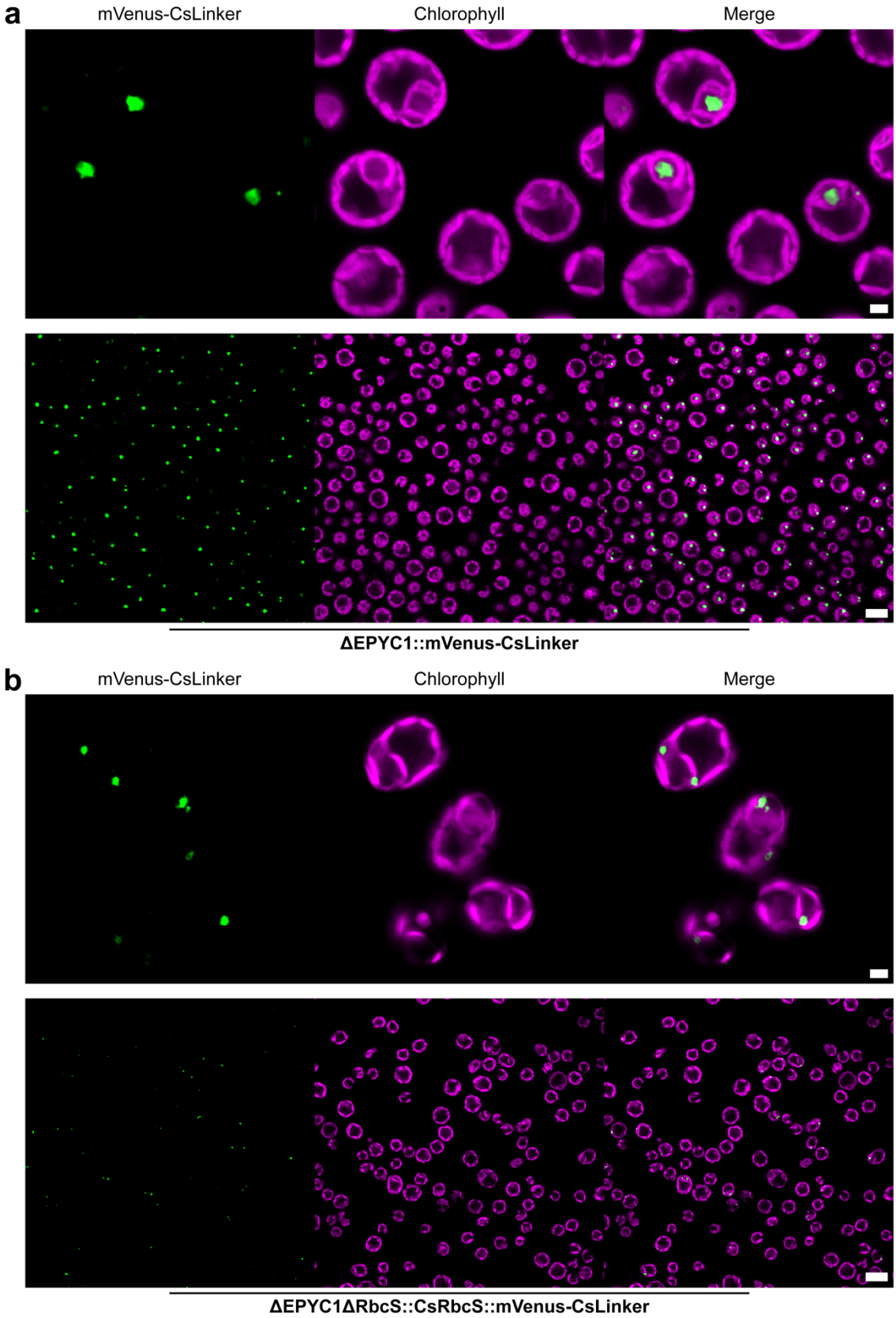


Figure 5-10 | Confocal microscopy of mVenus-CsLinker expression in Chlamydomonas strains.
(a) Confocal fluorescence microscopy images of mVenus-CsLinker expressed in ΔEPYC1. Scale bars = 2 µm (top), 10 µm (bottom). (b) Confocal images of mVenus-CsLinker expressed ΔEPYC1ΔRbcS::CsRbcS. Scale bars as above.

To determine if the apparent condensates observed in the ΔEPYC1 background also contained Rubisco, an mCherry-tagged copy of RbcS2 was introduced in the ΔEPYC1::mVenus-CsLinker strain. mCherry signal was observed to colocalise with the mVenus signal of the CsLinker in the pyrenoid and it was therefore assumed that Rubisco was also primarily present within the condensate (Fig. 5-11).

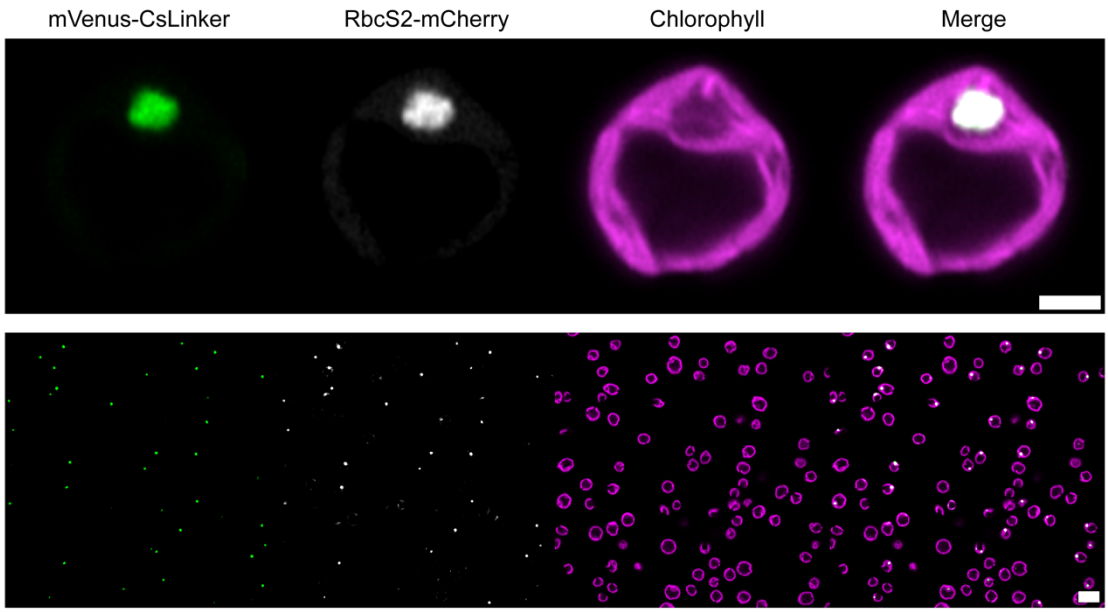


Figure 5-11 | Confocal microscopy of RbcS2-mCherry and mVenus-CsLinker colocalisation in Chlamydomonas.
Zoomed (top) and wide views of ΔEPYC1::mVenus-CsLinker::RbcS2-mCherry strain. Scale bars = 2 μm (top), 10 μm (bottom).

5.3.6. *Characterising functional cross-compatibility of the CsLinker-Rubisco condensate in Chlamydomonas*

Having demonstrated that CsLinker can condense Chlamydomonas Rubisco into a pyrenoid-like Rubisco condensate, it was next desirable to determine if the functionality of the condensate was also recovered. To avoid any detrimental effect of tags on the functionality of the CsLinker or the resulting condensate, the CsLinker was cloned and transformed into the ΔEPYC1 strain with no tag or epitope (Fig. 5-12). The empty transformation vector, conferring spectinomycin resistance through the aadA gene was also transformed into ΔEPYC1 to create a control strain (ΔEPYC1::aadA). Western blot analysis demonstrated largely consistent expression of the CsLinker in 7 ΔEPYC1::CsLinker strains (Fig. 5-12), relative to expression of the Chlamydomonas Rubisco large subunit (CrRbcL).

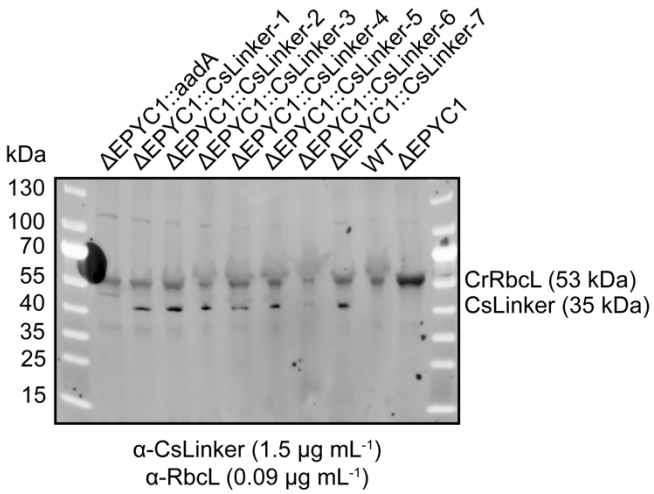


Figure 5-12 | Western blot analysis of untagged CsLinker expression in Chlamydomonas. Western blot analysis of Chlamydomonas Rubisco and CsLinker expression in the ΔEPYC1 background strain relative to control strains.

To determine if the pyrenoid was functionally recovered in these strains, spot test growth assay experiments were completed on the resulting strains. As expected, under high CO_2 and heterotrophic (TAP) conditions at both pH 8.0 and pH 8.2, all strains were able to grow indistinguishably from WT, as the pyrenoid-based CCM was not required (Fig. 5-13). Under air levels of CO_2 where a pyrenoid-based CCM is required for efficient growth, at pH 8.0 the ΔEPYC1::CsLinker strains showed almost WT levels of growth, suggesting functionality of the CsLinker-Rubisco condensate. Under these conditions the control ΔEPYC1::aadA strain showed no growth, though ΔEPYC1 retained a low level

of growth. At pH 8.2, growth of the ΔEPYC1 strain under ambient conditions was ablated, and again the ΔEPYC1::CsLinker strains showed growth and apparent functional pyrenoid recovery. It should be noted that recovery did not appear complete, as was observed when complementing the ΔEPYC1 with the native linker protein EPYC1 (Fig.5-7). Several possible explanations for the incomplete recovery exist. Despite the optimisation of CsLinker expression, it is possible the expression level, and *in vivo* abundance of the CsLinker protein is limiting when compared to the native levels of EPYC1. Future work to compare the *in vivo* abundance of CsLinker and EPYC1 in the strains using western blotting approaches compared to protein standards could be used to determine if this is the case. Alternatively, the pyrenoid could possess different mesoscopic properties – as observed with CsLinker *in vitro* (Chapter 3) – that also affect the function of the pyrenoid *in vivo*. Nevertheless, the data presented here suggest that CsLinker can functionally replace EPYC1 in the pyrenoid of Chlamydomonas in line with the observed *in vitro* cross-reactivity (Figs. 5-2, 5-3, 5-4) and *in vivo* co-localisation with Rubisco (Fig. 5-11), despite their distinct evolutionary origins. This exciting breakthrough demonstrates functional cross-compatibility of a disordered phase-separating protein that is likely a coincidence of rapid convergent evolution of the linker protein to interact with a protein that was itself evolving relatively slowly. This finding has great significance with respect to pyrenoid engineering approaches also as it shows for the first time that components from different pyrenoid systems can be functionally cross-compatible.

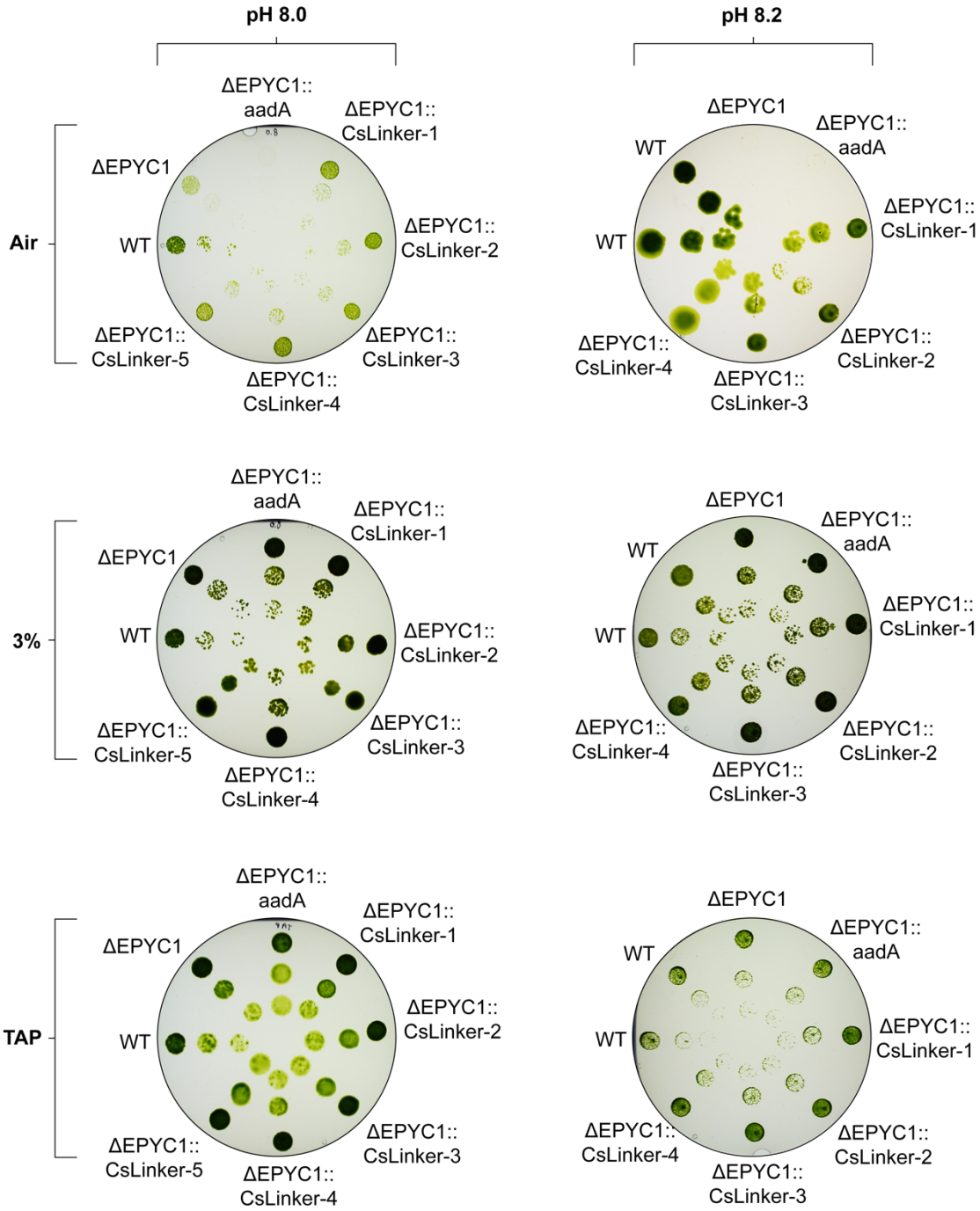


Figure 5-13 | Spot test growth assays of ΔEPYC1 strains complemented with CsLinker.
Images of spot test plates taken after 5 days of growth of 10,000 (outer ring), 1,000 (middle ring) and 100 (innermost ring) cells under the indicated conditions in the indicated strains.

5.4. Characterising in vitro cross-reactivity of CsLinker across the green lineage

Having demonstrated the functional cross-compatibility of CsLinker between Chlorella and Chlamydomonas, the next goal was to assess the extent of this cross-compatibility in the green lineage. Particular interest was paid to the green lineage given the clear phylogenetic relationship of Chlorella with higher crop plants (discussed in Chapter 2).

5.4.1. Comparison of the CsLinker binding site in RbcLs of the green lineage

Given the universal plastid encoding of the Rubisco large subunit (RbcL), a wealth of sequence data is available for RbcLs across the green lineage, which allowed analysis of the equivalent CsLinker binding sites in the respective sequences. Consensus sequences from MAFFT alignment of RbcLs in groups that represent the green lineage indicated a high level of sequence conservation of the CsLinker RbcL binding site, including the key interacting residues determined from the CsLinker-CsRubisco 3D structure (Fig. 5-14). Although there is inevitably species-level variation within groups, the consensus sequences demonstrated that the interface and the 4 CsLinker interacting residues are largely conserved in RbcLs representing all groups that evolved prior to the divergence of the spermatophytes. In line with the cross-compatibility observed with Chlamydomonas Rubisco which shared the same level of conservation, this analysis suggested that cross-compatibility of CsLinker likely extends some way across the green lineage.

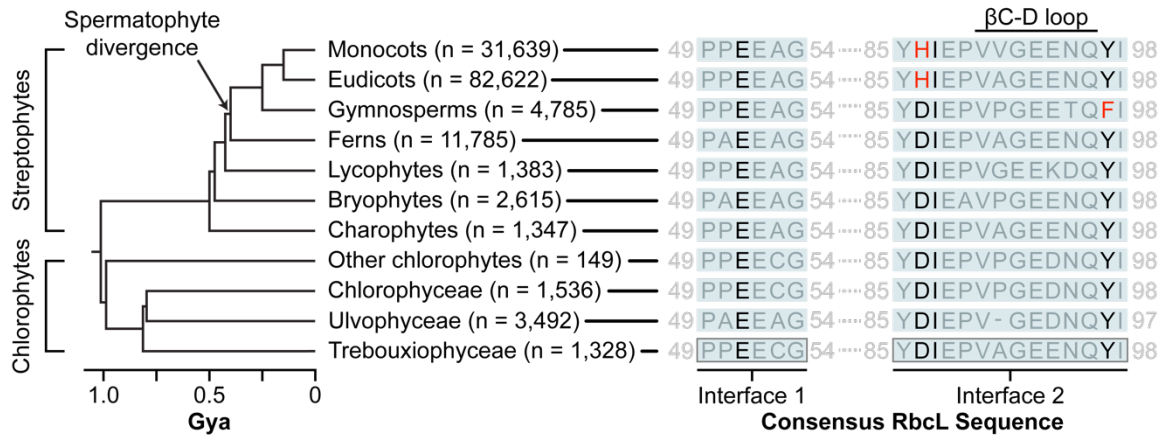


Figure 5-14 | Comparison of CsLinker RbcL binding sites in green lineage Rubiscos.
Alignment of the consensus sequences of the RbcL binding sites in Rubiscos of groups representing the green lineage. The time-calibrated phylogenetic tree (left) demonstrates the evolutionary relationship between the groups and the divergence of the spermatophytes is indicated. 'n' represents the number of sequences aligned in each group to determine the consensus sequence presented (right). The CsLinker interacting residues are shown in black, with non-conserved residues at the same relative positions in other Rubiscos shown in red. The consensus sequence of the group in which Chlorella resides is outlined (Trebouxiophyceae).

5.4.2. Characterising in vitro cross-compatibility of CsLinker with Ulva Rubisco

In addition to Chlorophyceae and Trebouxiophyceae which contain *Chlamydomonas* and *Chlorella* species respectively, the third primary chlorophytan class is Ulvophyceae, comprising a highly diverse class home to the main green seaweed lineages (Cortona *et al.,* 2020). As discussed in chapter 2, *Ulva* is the archetypal, model genus within the Ulvophyceae (Blomme *et al.,* 2020, 2023; Wichard *et al.,* 2015), in which a candidate linker protein was also previously identified (Chapter 2). Ulva possesses a similar phylogenetic relationship to Chlorella as Chlamydomonas (Fig. 2-22), and accordingly high conservation of the CsLinker binding site in the RbcL (Fig. 5-14) and it was therefore hypothesised that the cross-compatibility of CsLinker with Ulva Rubisco would also be similar.

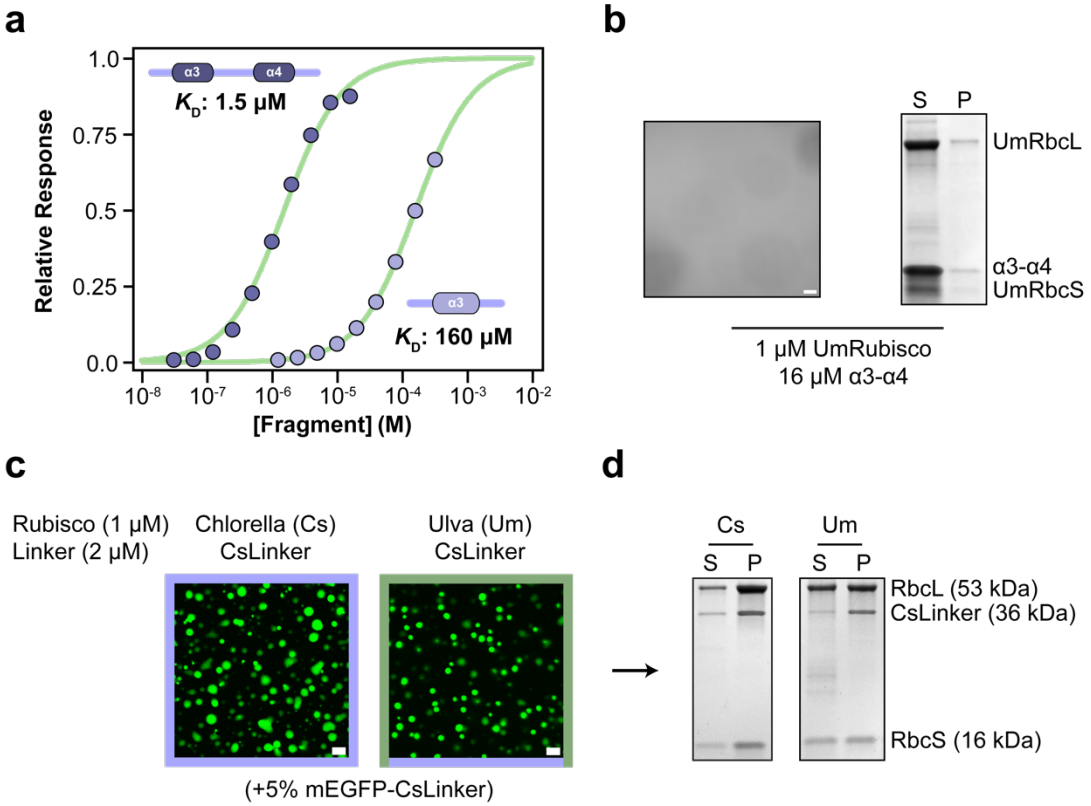


Figure 5-15 | *In vitro* cross-compatibility of CsLinker with Ulva Rubisco.
(a) SPR response normalised to Y_{max} value obtained from fit of raw data for Ulva Rubisco; n=3, error bars = S.D. (b) Brightfield microscopy of α3-α4-UmRubisco solution at the indicated concentration (left) and SDS-PAGE analysis of droplet sedimentation assays of the same solution composition (right). (c) Confocal fluorescence microscopy images of droplets formed at the indicated conditions with CsLinker ~10 minutes post-formation. mEGFP-tagged CsLinker was added to 5% molar ratio to allow visualisation. Scale bar = 5 μm. (d) SDS-PAGE analysis of droplet sedimentation assays of the solutions formed under the same conditions as in Fig. 5-15c.

As with Chlorella and Chlamydomonas Rubisco, SPR experiments demonstrated binding of both the α3 and α3-α4 fragments of CsLinker to Ulva Rubisco, and the α3-α4 fragment did not induce phase separation of the Ulva Rubisco (Fig. 5-15a,b). Interestingly, the affinity of both the α3 and α3-α4 fragments was markedly reduced for Ulva Rubisco (160 μM [CI_{95}: 155–166 μM] and 1.55 μM [CI_{95}: 1.39–1.72 μM] respectively), compared to the near identical values for Chlorella/Chlamydomonas Rubisco (103/108 μM and 1.2/1.3 μM respectively) (Fig. 5-3). When mixed with the full length CsLinker protein at concentrations that induced phase separation with the Chlorella and Chlamydomonas Rubiscos, Ulva Rubisco was phase separated into similar scale droplets (Fig. 5-15c), that could be sedimented and analysed (Fig. 5-15d). In agreement with the reduced apparent K_D of the CsLinker fragments, the efficiency of phase separation also appeared to be marginally reduced as compared to the cognate interaction (Fig. 5-15d), though this was not analysed quantitatively. Closer

analysis of the CsLinker interaction site of the Ulva Rubisco offered a possible explanation for the reduced apparent affinity and phase separation propensity of CsLinker for Ulva Rubisco. Unique to Ulvophyceae Rubiscos, there exists a single amino acid shortening of the βC-D loop (Fig. 5-14), that alters the loop conformation as predicted by AlphaFold2 modelling (Fig. 5-16). The same βC-D loop region of Chlorella RbcL was previously implicated in forming two hydrogen bonds with Gln170 of the α3-α4 CsLinker fragment (Fig. 4-13d), one of which was disrupted in the AlphaFold2 modelled Ulva Rubisco structure due to the amino acid change (Fig. 5-16). This discrepancy could account for the small affinity and phase separation differences observed, though further analysis to empirically determine the contribution of this hydrogen bond should be completed. Taken together the results with Ulva Rubisco were agreeable with previous findings with Chlamydomonas Rubisco, and consistent with the cross-compatibility of CsLinker being underpinned by its conserved interaction interface in the Rubisco large subunit. Given the rapid development of molecular tools for Ulva (Blomme *et al.,* 2020), and *in vivo* localisation of the Ulva linker in the pyrenoid (Fig. 2-11), confirming the *in vivo* functionality of CsLinker in Ulva is likely an achievable goal, and would add further weight to the functional cross-compatibility explored in Chlamydomonas (Fig. 5-13).

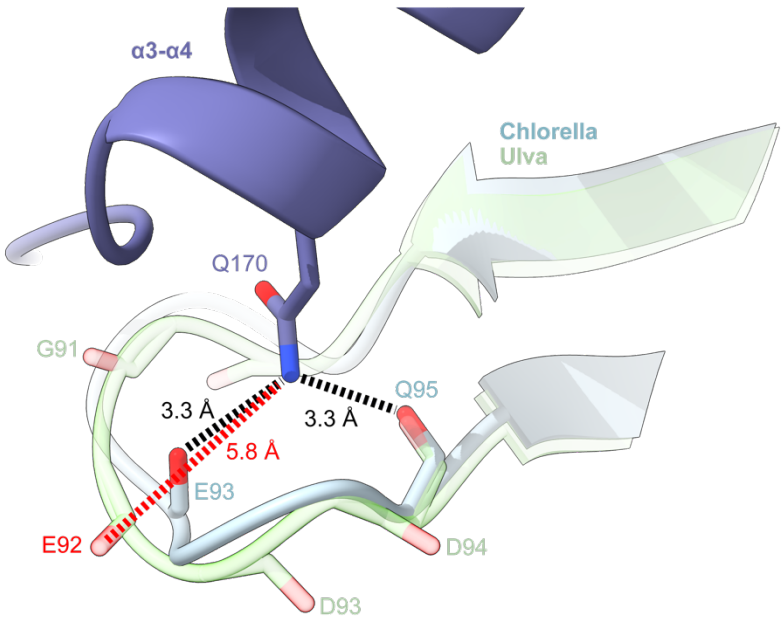


Figure 5-16 | Possible disruption of the CsLinker-RbcL hydrogen bond network in Ulva Rubisco.
Ribbon representation of α3-α4 interaction with Chlorella Rubisco (blue) from the cryo-EM determined 3D structure, with the putative hydrogen bonds of Gln170 indicated by black dashed lines. The same region of the AlphaFold2 modelled Ulva Rubisco is shown in green, with the disrupted hydrogen bond due to the loop change shown in red.

5.4.3. Probing in vitro cross-compatibility of CsLinker with representative plant Rubiscos

Having demonstrated phase separation of evolutionarily distinct Rubiscos in the chlorophytes presumably by use of the conserved RbcL interface, the next exciting goal was to determine if the cross-compatibility of CsLinker extended to plant Rubiscos. Sequence analysis of the CsLinker binding site in plant Rubiscos had previously demonstrated complete conservation in plant Rubiscos that evolved prior to divergence of the spermatophytes (in bryophytes, lycophytes, and ferns), and these Rubiscos were therefore of great interest with respect to compatibility with CsLinker. Ferns are the closest group of phylogenetic ancestors to spermatophyte plants, representing the most recently diverged group of streptophytes that possess all 4 of the CsLinker-interacting residues (Fig. 5-14). Testing the compatibility of Rubisco from ferns was therefore a primary goal. To this end, Rubisco was purified from the popular indoor fern *Adiantum raddianum*, whose RbcL sequence was completely conserved with the consensus sequence of ferns represented in Fig. 5-14. Besides the complete conservation of the consensus fern sequence, *Adiantum* was selected due to the commercial availability and amenability of the material to Rubisco purification. Previous attempts to purify fern Rubisco completed with environmental samples of common UK ferns (*Asplenium, Dryopteris, Polypodia* and *Polystichum* species) were wholly unsuccessful, possibly due to polyphenol presence in plant lysate (data not shown). As expected from the complete conservation of the CsLinker binding site (Fig. 5-14), fern Rubisco was readily phase separated by CsLinker *in vitro* (Fig. 5-17a), with a similar efficiency to the cognate interaction (Fig. 5-17b). This hugely significant result represented the first phase separation of native plant Rubisco with an algal linker protein.

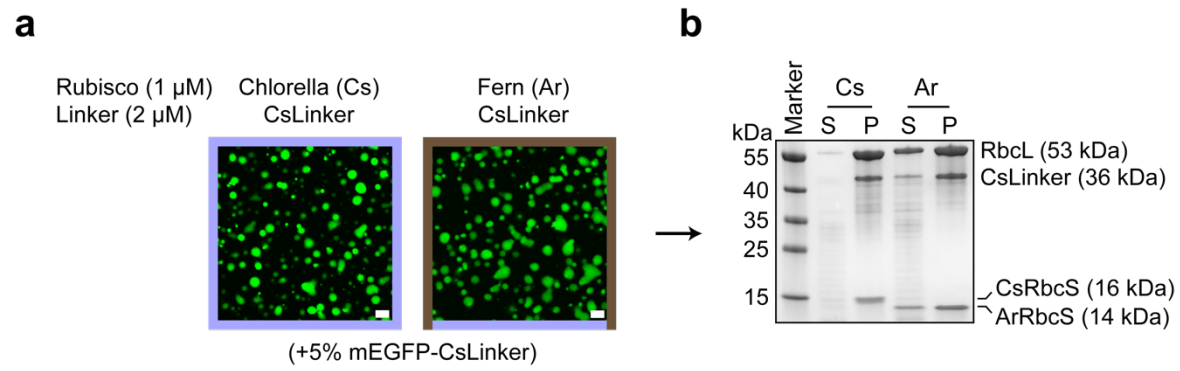


Figure 5-17 | *In vitro* phase separation of fern Rubisco by CsLinker.
(a) Confocal fluorescence microscopy images of droplets formed at the indicated conditions with CsLinker ~10 minutes post-formation. mEGFP-tagged CsLinker was added to 5% molar ratio to allow visualisation. Scale bar = 5 μm. (b) SDS-PAGE analysis of droplet sedimentation assays of the solutions formed under the same conditions as in Fig. 5-17a.

Encouraged by the phase separation of fern Rubisco, the next logical step was to determine whether CsLinker could phase separate flowering plant (spermatophyte) Rubiscos, with a particular focus on the sequences of angiosperms (Eudicots and Monocots), that are home to most crop plants. Most angiosperm plants possess a single substitution of Asp to His at position 86 in the RbcL (Fig. 5-14), and it was therefore uncertain whether the Rubisco would be able to phase separate with CsLinker. The Rubisco sequence of *Spinacea oleracea* (Spinach hereafter) represented complete conservation of the consensus eudicot CsLinker-interacting Rubisco sequence and was consequently purified for *in vitro* analysis. Presumably because of the ablation of the electrostatic interaction with Lys177 of CsLinker (Fig. 4-13), Spinach Rubisco was unable to phase separate at the concentrations used for the cognate reaction, or at elevated concentrations (Fig. 5-18a,b). In agreement, Arabidopsis Rubisco, which possesses the same Asp86 to His86 change in addition to a Tyr97 to Phe97 substitution, was also unable to phase separate under the same conditions, though some aggregation was observed at higher concentrations (Fig. 5-18c,d). These results suggested that the amino acid substitutions at the CsLinker interaction site in the selected plant Rubiscos were sufficient to prevent phase separation with CsLinker, but other Rubisco-specific effects cannot be ruled out from this analysis alone.

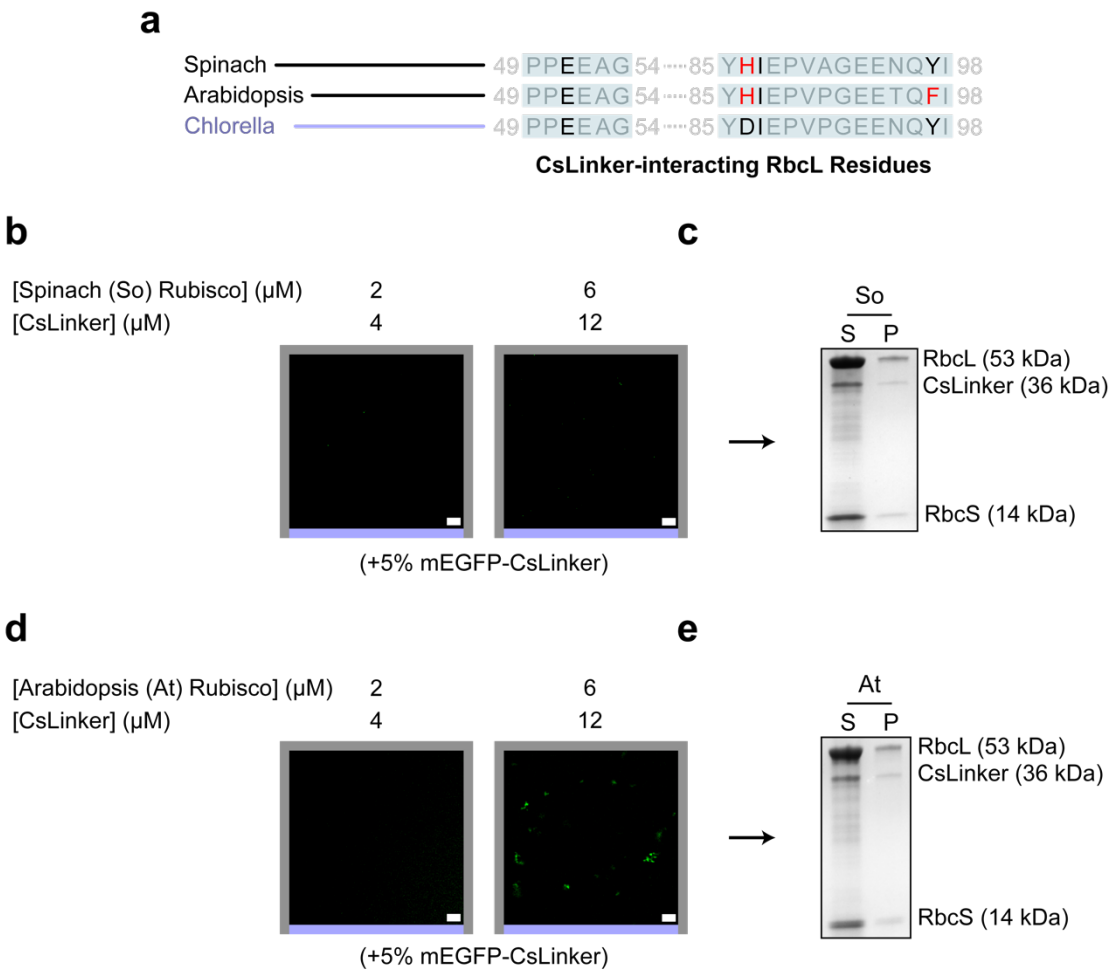


Figure 5-18 | Characterising cross-compatibility with Spinach Rubisco.
(a) Alignment of the CsLinker-interacting residues in the RbcL of Chlorella Rubisco and the equivalent regions in the RbcL of Spinach and Arabidopsis RbcLs. The interacting residues in Chlorella are shown in black, with non-compatible residues in equivalent positions in red. (b) Confocal fluorescence microscopy images of Spinach Rubisco solutions at the indicated conditions with CsLinker ~10 minutes post-mixing. mEGFP-tagged CsLinker was added to 5% molar ratio. Scale bar = 5 μm. (c) SDS-PAGE analysis of droplet sedimentation assays of the solution formed at 2 μM Spinach Rubisco, 4 μM CsLinker. (d) Confocal fluorescence microscopy images of Arabidopsis Rubisco solutions at the indicated conditions with CsLinker ~10 minutes post-mixing. mEGFP-tagged CsLinker was added to 5% molar ratio. Scale bar = 5 μm. (e) SDS-PAGE analysis of droplet sedimentation assays of the solution formed at 2 μM Arabidopsis Rubisco, 4 μM CsLinker.

5.4.4. Point mutation of Chlamydomonas Rubisco to probe the CsLinker interaction with the RbcL

To seek more definitive proof that the single amino acid substitution present in spinach Rubisco was sufficient to prevent phase separation, it was desirable to introduce the same substitution to a Rubisco that had previously been phase separated with CsLinker. Given the tractability and precedent of RbcL editing (Larson *et al.,* 1997; Ott *et al.,* 2000; Satagopan & Spreitzer, 2004), as well as the previous demonstration of functional phase separation with CsLinker (Figs. 5-4, 5-10, 5-13), Chlamydomonas

Rubisco was chosen for this purpose. A ΔEPYC1ΔRbcL strain was initially created by the homologous recombination of the aadA spectinomycin resistance cassette into the native Chlamydomonas RbcL locus. The knockout was confirmed at both the genetic and protein level (Fig. 5-19a,b). A Chlamydomonas RbcL in which the Asp86 was point mutated to His86 was then reintroduced to the ΔEPYC1ΔRbcL strain by homologous recombination replacement of the aadA cassette and selection for recovery of photosynthetic growth under CO_2-replete conditions. The RbcL locus of the phototrophic colonies was PCR amplified and sequenced to ensure the mutation was present (Fig. 5-19c). A control transformation in which unedited Chlamydomonas RbcL was used was also performed and the number of colonies obtained from both transformations was comparable (112 for WT, 109 for D86H), suggesting the D86H mutation has no detrimental effect on Rubisco function *in vivo* (data not shown). In contrast, transformation of Chlorella, *Arabidopsis*, Fern and Tobacco RbcL sequences in the same vector yielded no phototrophic colonies after transformation, suggesting that Rubisco assembly or function was affected in these strains, presumably due to species-specific assembly and activase requirements (Carmo-Silva & Salvucci, 2013).

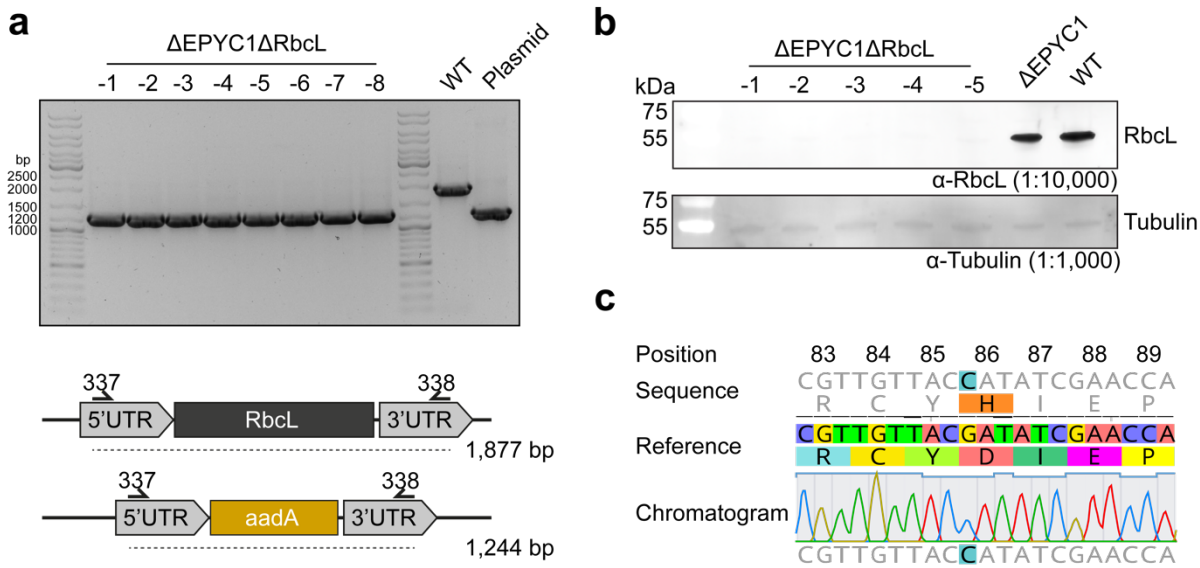


Figure 5-19 | Creation of D86H mutated Rubisco.
(a) PCR screening of ΔEPYC1ΔRbcL colonies following transformation with the aadA spectinomycin cassette relative to controls. A schematic representation of the expected PCR products is shown below. (b) Western blot confirmation of RbcL knockout in the ΔEPYC1ΔRbcL lines relative to WT and ΔEPYC1 controls. (c) Sequencing of the D86H mutated RbcL locus in the ΔEPYC1ΔRbcL::D86H_CrRubisco strain.

D86H mutated Rubisco was subsequently purified from the ΔEPYC1ΔRbcL::D86H_CrRubisco strain. In contrast to the unedited Chlamydomonas Rubisco, the D86H mutated Rubisco did not phase separate with CsLinker, in line with the results obtained with the Spinach Rubisco (Fig. 5-20a,b). Importantly the D86H mutated Rubisco maintained the ability to phase separate with the cognate Chlamydomonas linker EPYC1 (Fig. 5-20c,d), consistent with the maintenance of the EPYC1 binding site in the RbcS of the Rubisco. As expected, Spinach Rubisco was unable to phase separate with EPYC1, presumably due to the multiple amino acid substitutions in the equivalent EPYC1 binding site of the RbcS (Fig. 5-20c). Taken together these results demonstrated a single amino acid substitution in the RbcL, which likely ablates the electrostatic interaction between Asp86 of the RbcL and Lys177 of CsLinker (Fig. 4-13b), is sufficient to ablate phase separation of the Chlamydomonas Rubisco. This data offered biochemical support for the structurally determined CsLinker binding site on Rubisco, which was previously only biochemically confirmed by mutagenesis with respect to the CsLinker sequence (Fig. 4-15). The results of the D86H mutated Chlamydomonas Rubisco were consistent with the same substitution that is present in the Spinach Rubisco and offers a likely explanation as to the lack of phase separation of Spinach (Fig. 5-18b,c) and *Arabidopsis* (Fig. 5-18d,e) Rubiscos. More generally, the results indicated that the cross-compatibility of the CsLinker would likely not extend to Rubiscos that possess a substitution at the equivalent D86 position of the Rubisco (Fig. 5-14).

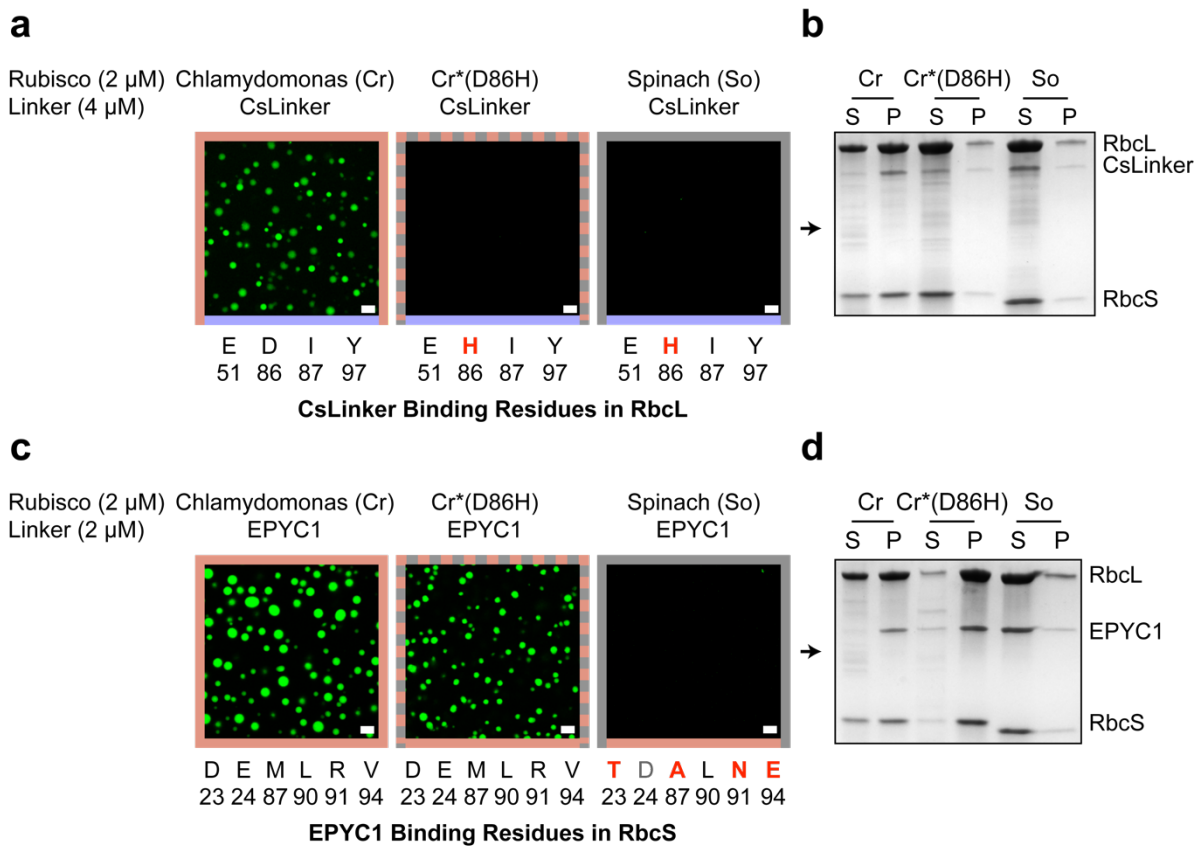


Figure 5-20 | *In vitro* characterisation of phase separation of D86H mutated Chlamydomonas Rubisco. (a) Confocal fluorescence microscopy images of solutions at the indicated conditions with CsLinker ~10 minutes post-formation. mEGFP-tagged CsLinker was added to 5% molar ratio to allow visualisation. The sequence of the CsLinker interacting residues in the equivalent position of each RbcL is indicated below each image. Scale bar = 5 μm. (b) SDS-PAGE analysis of droplet sedimentation assays of the solutions formed under the same conditions as in Fig. 5-20a. (c) Confocal fluorescence microscopy images of solutions at the indicated conditions with EPYC1 ~10 minutes post-formation. EGFP-tagged EPYC1 was added to 5% molar ratio to allow visualisation. The sequence of the EPYC1 interacting residues in the equivalent position of each RbcS is indicated below each image. Scale bar = 5 μm. (b) SDS-PAGE analysis of droplet sedimentation assays of the solutions formed under the same conditions as in Fig. 5-20c.

5.4.5. Exploring cross-compatibility of CsLinker for crop Rubiscos

Having demonstrated that the D86H substitution prevents phase separation with CsLinker, it was next desirable to determine the distribution of this substitution across C_3 crop Rubiscos. Comparison of the Rubisco large subunit sequences of the 24 most valuable C_3 crops indicated the CsLinker interaction interface was one of the more variable regions relative to the rest of the subunit (Fig. 5-21). Accordingly, there was some variation, but overwhelmingly position 86 of the RbcL was occupied by a His (Fig. 5-22). Clear phylogenetic grouping of the sequences was observed, with notable examples in the both the monocots (Wheat, Barley, and Oat) and eudicots (Beet and Spinach). The RbcL sequence of cotton was an interesting outlier in this subset, possessing all 4 of the key CsLinker interacting residues and

demonstrating almost total sequence conservation of the interaction interfaces with Chlorella, suggesting this Rubisco would natively phase separate with CsLinker. Other outliers included Soy and Cassava which possessed a Gly in position 86, which would presumably ablate the interaction with CsLinker and prevent phase separation. Broadly the conclusions of experiments completed with Spinach and Arabidopsis Rubiscos (Fig. 5-18) allowed prediction that most of the C_3 crop Rubiscos would not phase separate with CsLinker. Contrastingly, crop Rubiscos in the Solanales order demonstrated an almost totally conserved position 86 Arg in addition to several other changes in the βC and βC-D loop (Fig. 5-22). Although it was predicted the Arg86 substitution would likely ablate the interaction with CsLinker, no comparisons could be drawn between this sequence and the sequence of already tested Rubiscos, due to the large changes in sequence properties in these regions (D86R, P89R, E94K, N95D). It was therefore decided cross-compatibility with the Solanales Rubiscos should be tested empirically.

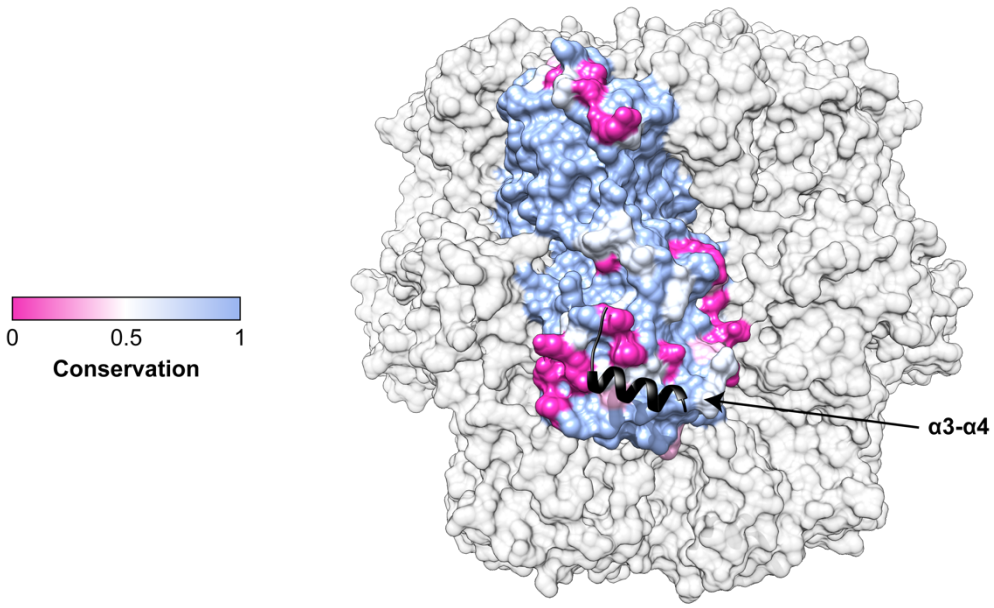


Figure 5-21 | Conservation of Rubisco large subunit residues in crop Rubiscos.
Surface representation of Rubisco with one large subunit coloured according to sequence conservation based on the alignment of the 24 C_3 crop Rubiscos presented in Fig. 5-22. The α3-α4 helix is shown in the binding position relative to the Chlorella Rubisco structure.

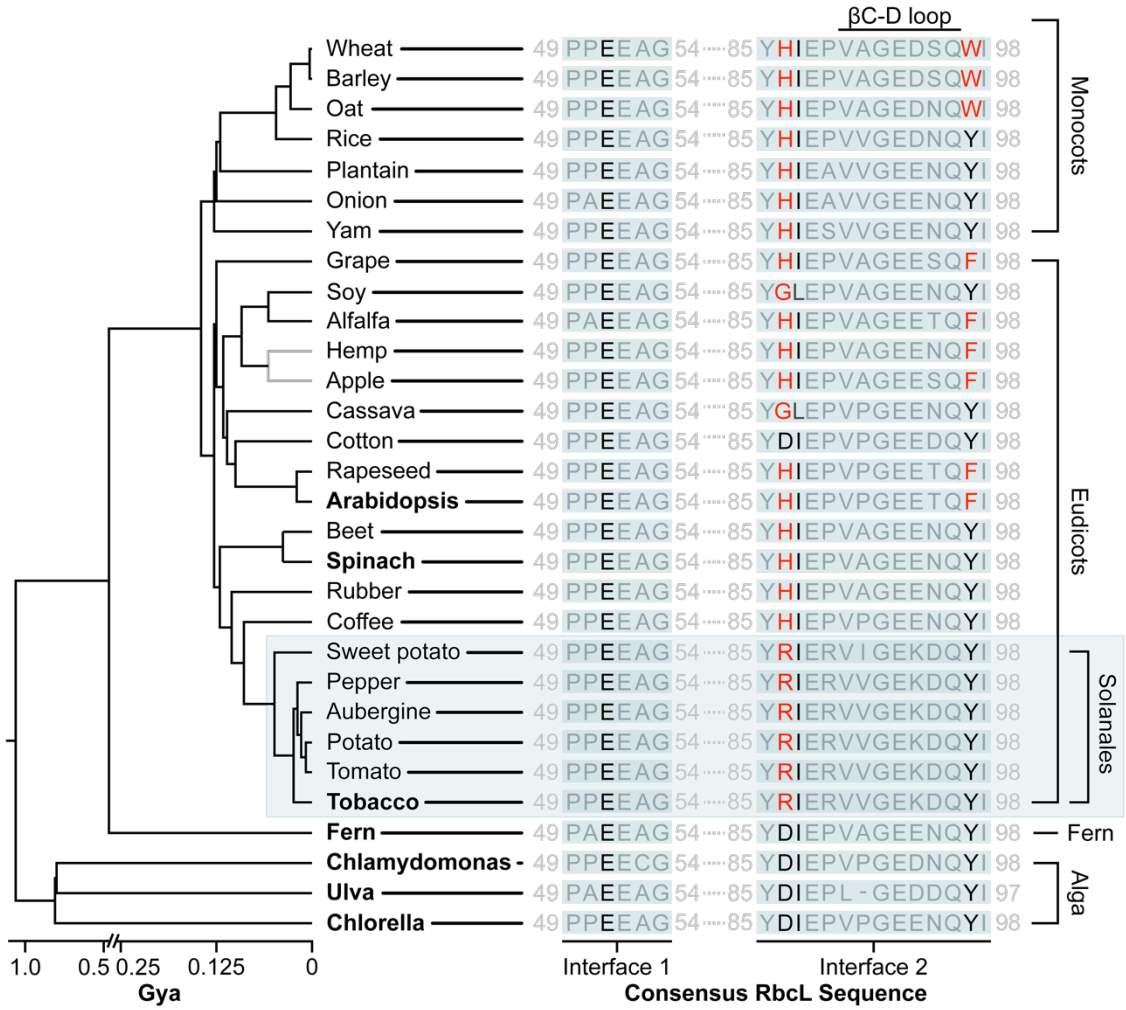


Figure 5-22 | Phylogeny of C_3 crop plants and comparison of their CsLinker RbcL binding sites. Consensus phylogenetic chronogram (left) of the 24 most valuable C_3 crop plants (from FAOSTAT data), as well as representative model species (bold) interrogated in this study. See methods for a detailed description of how the tree was built. An alignment of the CsLinker interacting residues of the RbcLs is shown adjacent. Black colouring indicates conserved interacting residues relative to the Chlorella RbcL sequence, and red indicates non-conserved residues.

5.4.6. Characterising in vitro cross-compatibility of CsLinker with Solanaceae Rubiscos

Nicotiana is the model Solanales genus (Goodin *et al.,* 2008), where *Nicotiana tabacum* (Tobacco) and *Nicotiana benthamiana* are the model species. Though the RbcL sequence of *N. benthamiana* was not available, it was assumed to be highly similar if not identical to that of *N. tabacum* in line with the high level of sequence conservation present in the Solanales (Fig. 5-22). *N. benthamiana* Rubisco was subsequently purified for *in vitro* experiments. Based on the largely different properties of the amino acid substitutions in the CsLinker binding interface, it was assumed that *N. benthamiana* Rubisco would not phase separate with CsLinker. Surprisingly, at the concentrations used for the cognate interaction,

small droplets that did not fuse over time were formed (Fig. 5-23a,c). Regardless of droplet state, droplet sedimentation assays indicated that both components entered the droplet phase at similar efficiency to the cognate interaction (Fig. 5-23b). With increasing global concentrations at the same ratio of components, droplets more readily fused and exhibited qualitative rearrangement over second timescales in bleaching experiments (Fig. 5-23d). No comparison could be made to the cognate droplets as FRAP experiments were not completed under these concentrations previously. As expected, no phase separation was observed with EPYC1, in line with the lack of conservation of the EPYC1 binding site in the RbcS (Fig. 5-23e,f), as compared to the Chlamydomonas sequence (Fig. 5-20c). The surprising result of CsLinker demixing *N. benthamiana* Rubisco indicated that other Solanales Rubiscos would also phase separate, due to their highly conserved RbcL sequence (Fig. 5-22). Regardless, this result represented the first phase separation of a native spermatophyte plant Rubisco. Motivated by this finding, the next aim was therefore to determine if CsLinker could phase separate crop Rubiscos of the *Solanum* genus. To this end, Potato (*Solanum tuberosum*) and Tomato (*Solanum lycopersicum*) Rubiscos were purified.

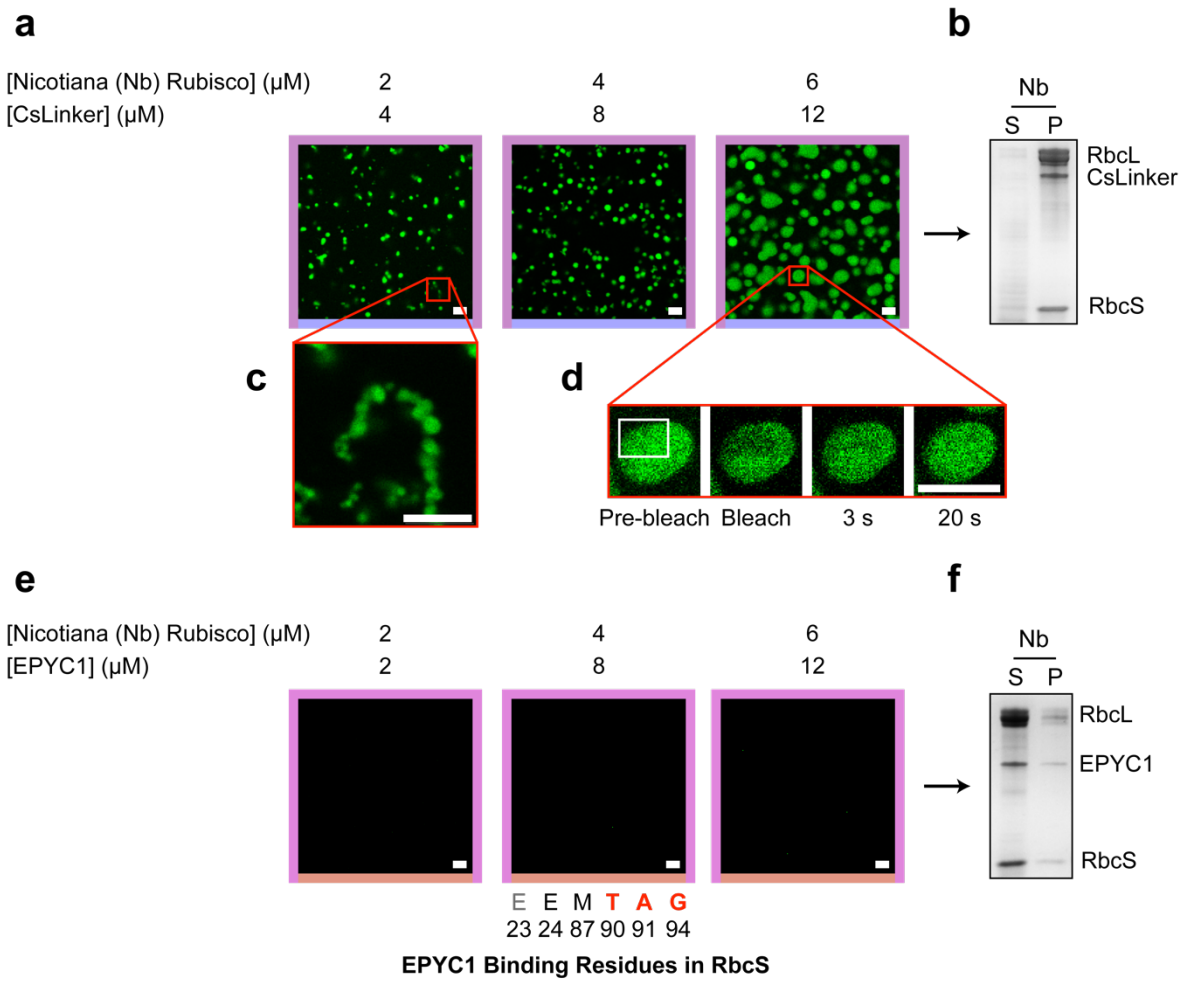


Figure 5-23 | Characterisation of cross-compatibility of CsLinker and EPYC1 with *Nicotiana benthamiana*. (a) Confocal fluorescence microscopy images of solutions at the indicated conditions with CsLinker ~10 minutes post-formation. mEGFP-tagged CsLinker was added to 5% molar ratio to allow visualisation. Scale bar = 5 μm. (b) SDS-PAGE analysis of droplet sedimentation assays of the solutions formed under the same conditions as in Fig. 5-23a, left panel. (c) Confocal fluorescence microscopy image of non-fusing droplets. Scale bar = 5 μm. (d) Confocal fluorescence microscopy images taken throughout a qualitative FRAP experiment completed on droplets formed at 6 μM Rubisco and 12 μM CsLinker. Scale bar = 5 μm. (e) Confocal fluorescence microscopy images of solutions at the indicated conditions with EPYC1 ~10 minutes post-formation. EGFP-tagged EPYC1 was added to 5% molar ratio to allow visualisation. Scale bar = 5 μm. (f) SDS-PAGE analysis of droplet sedimentation assays of the solutions formed under the same conditions as in Fig. 5-23e, left panel.

Although purification of Tomato and Potato Rubisco was non-trivial, enough material was obtained to complete preliminary *in vitro* trials at the lower global concentration points (2 μM and 4 μM Rubisco datapoints). Both Rubiscos demonstrated a similar trend to that observed with *N. benthamiana* Rubisco, forming small non-fusing droplets at the lower concentration that were more readily fused at higher concentrations (Fig. 5-24a,c). Droplet sedimentation demonstrated complete demixing of Rubisco and CsLinker with both Rubiscos at the lower concentration point (Fig. 5-24b,d). Following the results with *N. benthamiana* Rubisco demixing of the crop Solanales Rubiscos was expected, given the total

conservation of the presumed CsLinker binding site on Rubisco (Fig. 5-22) and demonstrated the first phase separation of a crop plant Rubisco. Presuming binding to the same interface on Rubisco, the surprising cross-compatibility of CsLinker with Solanales Rubiscos suggested that the interaction between CsLinker and Rubisco was somehow compensated in place of the ablated electrostatic interaction due to the Asp86 to Arg86 substitution.

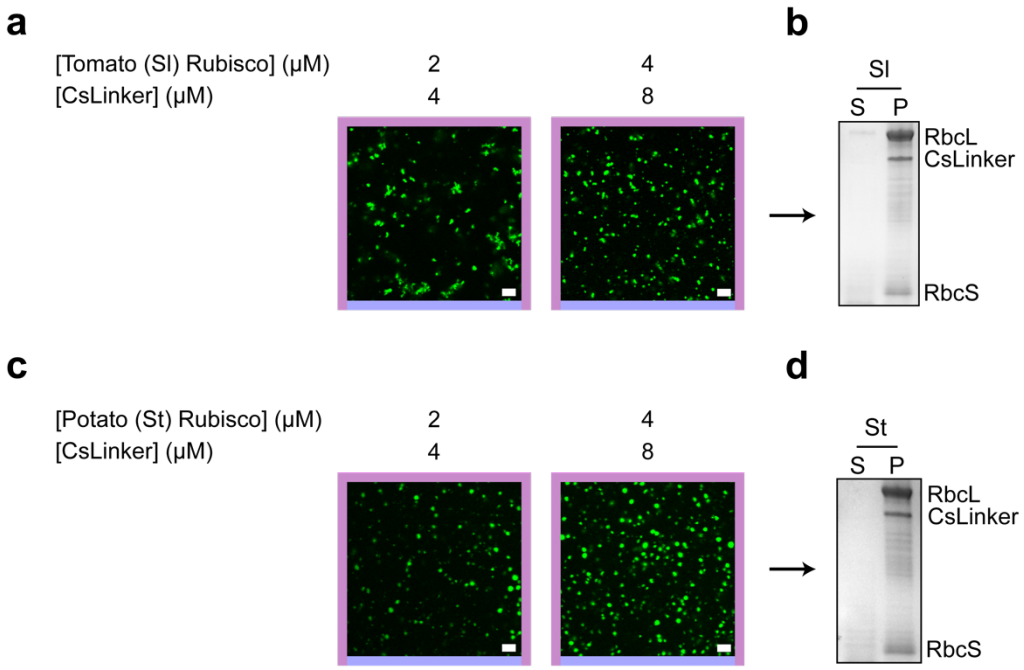


Figure 5-24 | Characterisation of *in vitro* cross-compatibility of CsLinker with Solanaceae crop Rubiscos. a) Confocal fluorescence microscopy images of solutions at the indicated conditions with CsLinker ~10 minutes post-formation. mEGFP-tagged CsLinker was added to 5% molar ratio to allow visualisation. Scale bar = 5 μm. (b) SDS-PAGE analysis of droplet sedimentation assays of the solutions formed under the same conditions as in Fig. 1-24a, left panel. (c) Confocal fluorescence microscopy images of solutions at the indicated conditions with CsLinker ~10 minutes post-formation. mEGFP-tagged CsLinker was added to 5% molar ratio to allow visualisation. Scale bar = 5 μm. (d) SDS-PAGE analysis of droplet sedimentation assays of the solutions formed under the same conditions as in Fig. 1-24c, left panel.

5.4.7. A possible rationale for CsLinker interaction with Solanales Rubiscos

A crystal structure of *Nicotiana tabacum* Rubisco in the apo state was available to make direct comparisons between the CsLinker binding interface on Chlorella Rubisco and the equivalent position on *Nictotiana* Rubisco (Duff *et al.,* 2000). As expected, the position 86 Arg in the *Nicotiana* structure was not positioned to form favourable interactions with the CsLinker, though the other electrostatic interaction between Glu51 of the RbcL and Arg176 of CsLinker was still accessible (Fig. 5-25b). The hydrophobicity of the interface between the structures was also largely similar in line with the conservation of the key Ile87 and Tyr97 residues (Fig. 5-25a,c), though small differences were accounted due to the Pro89 to Arg89 substitution in the *Nicotiana* RbcL (Fig. 5-22). As previously noted, the largest changes between the sequences and structures exist in the βC-D loop (Fig. 5-25d). The starkest change in residue property was present at position 94, where the Chlorella Glu94 is substituted for an Lys94 in the *Nicotiana* structure. Interestingly, the substituted Lys94 residue was positioned to feasibly form an electrostatic interaction with Glu169 of the CsLinker if an alternative side chain rotamer could be occupied (Fig. 5-25e). This conclusion comes with the caveat that the side chain position of Glu169 was poorly resolved in the cyro-EM structure of the CsLinker-CsRubisco complex (Fig. 4-12). Clearly this hypothesis requires biochemical and structural support but offers a reasonable rationale behind the observed cross-compatibility of CsLinker with Solanales Rubiscos, given the loss (D86R) and gain (E94K) of separate electrostatic interactions to maintain 2 electrostatic interactions at the interface. The general finding of unexpected cross-compatibility also offers huge promise with respect to synthetic engineering of linker proteins for phase separation of crop Rubisco, suggesting that small changes may be made to the linker-Rubisco interaction interface to bimodally tune phase separation propensity. Future studies could apply amino acid scanning or guided design approaches to the CsLinker helix to facilitate interaction with, and phase separation of, the non-Solanaceae crop Rubiscos which are largely incompatible due to a single amino acid substitution.

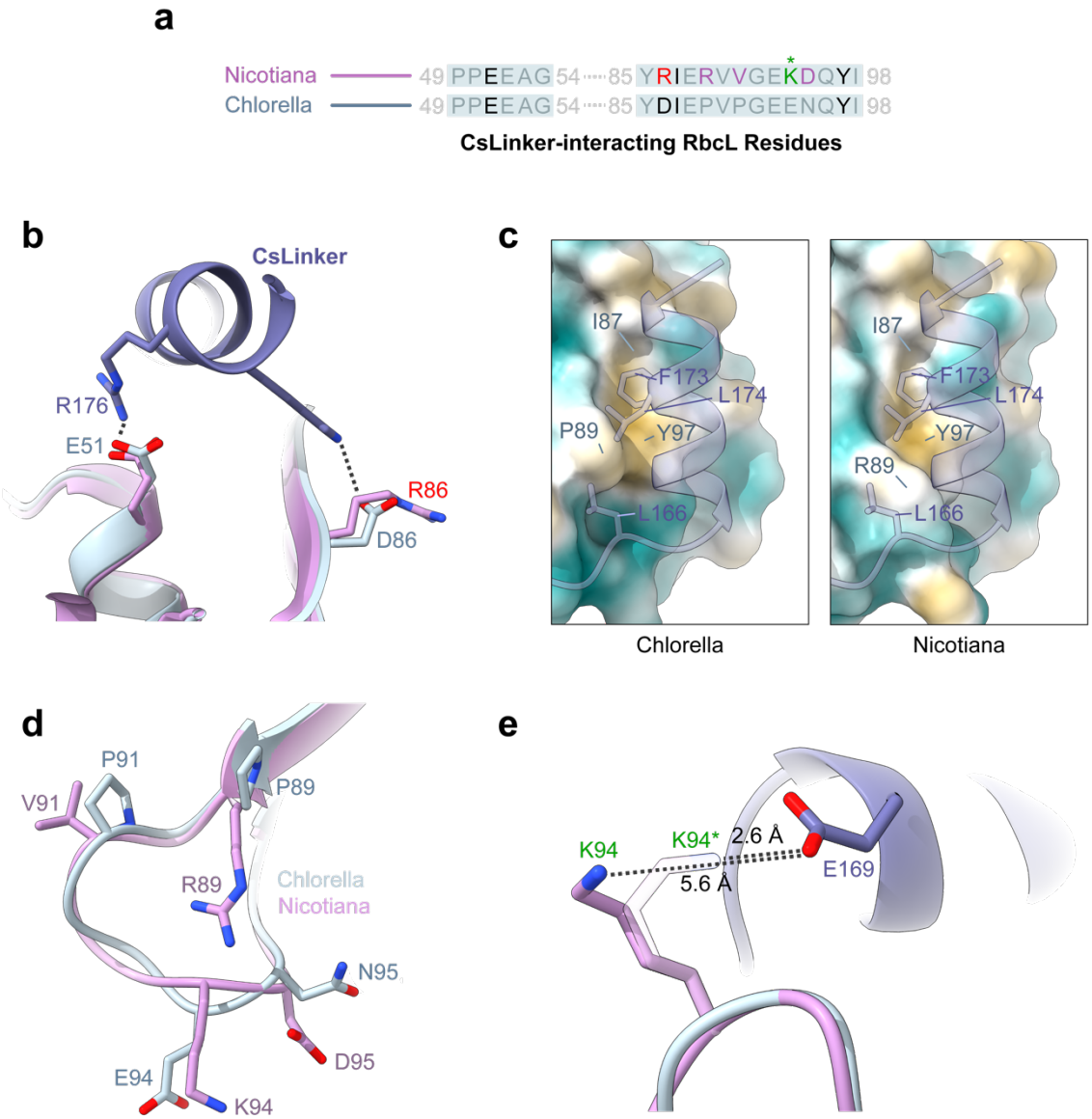


Figure 5-25 | Comparison of the Chlorella and *Nicotiana* CsLinker interactions sites in the RbcL.
(a) Alignment of the CsLinker interaction interfaces in the RbcL of *Nicotiana* and Chlorella. Interacting residues in the Chlorella sequence and equivalent positions in the *Nicotiana* sequence are shown in black, non-compatible interacting residues in red, non-interacting differences in purple and the possible additional interaction in green. (b) Electrostatic interactions at the CsLinker-RbcL interface. The structure of the *Nicotiana* RbcL (from PDB: 1EJ7; (Duff *et al.,* 2000)) and Chlorella RbcL (from the cryo-EM structure in Chapter 4, PDB: 8Q05) are displayed overlaid. The electrostatic interactions are indicated by dashed lines. The presumably non-interacting R86 is shown in red. (c) Hydrophobic interactions in the Chlorella (left) and *Nicotiana* (right) RbcL. The RbcL is shown as a surface representation coloured by hydrophobicity calculated according to Kyte-Doolittle scoring. (d) Differences in the βC-D loop region. Residues that are different between the two sequences are displayed and numbered accordingly. (e) The possible additional electrostatic interaction between CsLinker and *Nicotiana* Rubisco. The position of K94 in the 1EJ7 PDB structure is shown solid with green numbering. The alternative possible rotamer is shown transparent and labelled K94*. The distances of the closest electrostatic interactions are indicated.

5.4.8. Testing in planta cross-compatibility of CsLinker with Solanaceae Rubisco

THE WORK IN THIS SECTION WAS COMPLETED IN COLLABORATION WITH YUWEI MAO (UNIVERSITY OF EDINBURGH) AND IS PRESENTED HERE IN THE INTERESTS OF COMPLETENESS FOR DISCUSSION.

Having demonstrated *in vitro* cross-compatibility of Solanaceae Rubiscos with CsLinker, it was next desirable to test whether this translates *in vivo*. Separate transient expression in *Nicotiana benthamiana* of TurboGFP (tGFP)- and mNeonGreen-tagged mature CsLinker sequences fused to the *Arabidopsis* RbcS1A transit peptide resulted in clear condensate formation within the chloroplast, relative to control infiltrations with the fluorophore sequence fused directly to the transit peptide (Fig. 5-26). The number of condensates peaked with the level of expression on day 2, before reducing in abundance in the following 3 days (data not shown). Multiple condensates were observed in each chloroplast, suggesting that the condensates were not undergoing fusion or Ostwald ripening to form a single condensate, as observed previously for Chlamydomonas proto-pyrenoids in *Arabidopsis* (Atkinson *et al.,* 2020). This behaviour is perhaps analogous to that observed with lower concentrations of protein *in vitro* (Fig. 5-23). Although these results are very encouraging with respect to the phase separation of native plant Rubisco *in vivo*, only tentative conclusions can be made based on several caveats of the preliminary data. Firstly, so far, we have not demonstrated that Rubisco is present in the condensates, nor have we shown that the condensates are *bona fide* liquid-liquid phase separated droplets. Secondly, the results were obtained from the transient expression of CsLinker protein as opposed to the stable expression lines used previously in *Arabidopsis*. Regardless, the results present a potentially huge step for plant engineering that could allow the circumvention of complex Rubisco engineering to allow assembly of a functional pyrenoid *in planta*. These results, in conjunction with the apparently simpler pyrenoid assembly framework in Chlorella discussed in Chapter 2 could allow for the tangible assembly of prototype pyrenoids in Solanales crops imminently.

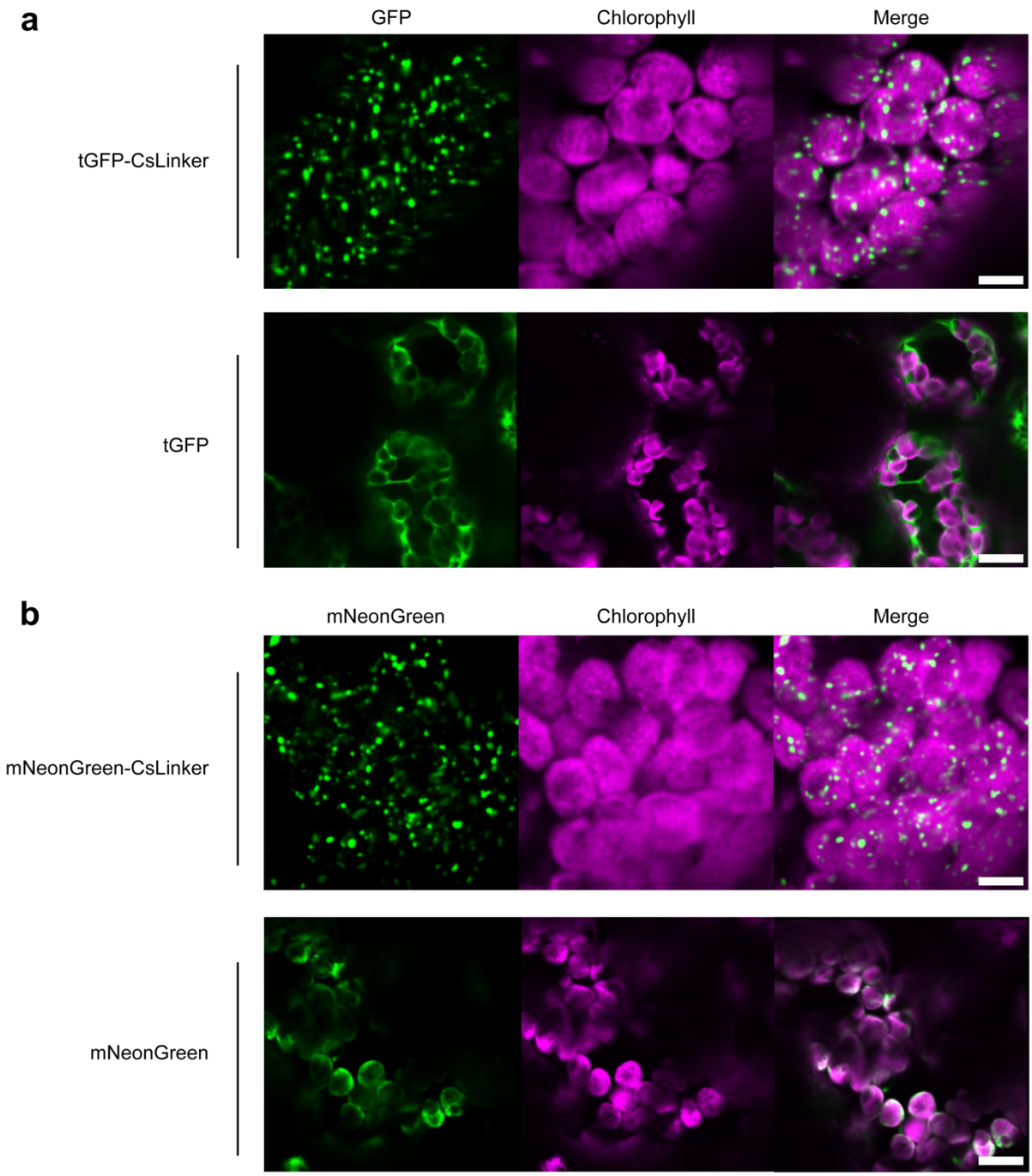


Figure 5-26 | Transient expression of CsLinker *in planta*.
(a) Confocal fluorescence microscopy images of tGFP-tagged CsLinker (top) and free tGFP (bottom), 2 days after infiltration of *Nicotiana benthamiana* leaves. Scale bar = 5 µm. (b) Confocal fluorescence microscopy images of mNeonGreen-tagged CsLinker (top) and free mNeonGreen (bottom), 2 days after infiltration of *Nicotiana benthamiana* leaves. Scale bar = 5 µm.

5.5. Discussion

This chapter contributed several major findings relevant to general phase separation biology, plant engineering and fundamental pyrenoid biology. These advances were afforded by the identification and detailed characterisation of CsLinker in previous chapters which enabled a set of hypotheses to be tested.

The finding that CsLinker interacts with the RbcL of Chlorella Rubisco in the previous chapter (Fig. 4-13) was of great excitement as it suggested an increased likelihood of cross-compatibility of CsLinker with plant Rubiscos given the similarity of RbcLs between algal chlorophytes and plant streptophytes (Fig. 5-14). This was of unique excitement relative to other previously characterised Rubisco linkers that are likely not cross-compatible with plant Rubiscos for a range of reasons. Although EPYC1 is also a Rubisco linker protein from an algal chlorophyte, its molecular interaction with the more species-variable RbcS was hypothesised to prevent cross-compatibility with other green lineage Rubiscos. This was empirically proven in this work (Fig. 5-4, Fig. 5-20, Fig. 5-23) and was consistent with previous findings of the lack of cross-compatibility of EPYC1 with rice Rubisco (Wunder *et al.,* 2018). PYCO1 has also been shown to lack cross-compatibility with Chlamydomonas Rubisco (Oh *et al.,* 2023), likely due to the lack of conservation of the interacting residues between the red-origin form ID Rubisco of Phaeodactylum and the green-origin form IB Rubisco of Chlamydomonas. Although not empirically tested, the same scenario is likely true of the α-carboxysomal linker protein CsoS2 due to lack of conservation of the RbcL residues between the form IA Rubisco of proteobacteria and the form IB Rubisco of the chlorophytes. It therefore emerged that CsLinker was a promising candidate to demonstrate cross-compatibility across the green lineage, given the utilisation of a highly conserved RbcL interaction interface that is a result of the close phylogenetic relationship relative to other linker proteins. This hypothesis was tested both *in vitro* and *in vivo* using previously introduced techniques as well as novel tools generated in this study.

As was expected from the similarity of the CsLinker interface of the RbcLs of Chlamydomonas and Chlorella, CsLinker was able to demix Chlamydomonas Rubisco into droplets in an almost

indistinguishably similar manner as the native interaction with Chlorella Rubisco (Fig. 5-4), consistent with SPR data with CsLinker fragments which demonstrated the same indistinction (Fig. 5-3). Uniquely, this finding was extended *in vivo* (Fig. 5-10), where for the first time a pyrenoid linker protein was transplanted across species. The major finding that CsLinker could functionally replace EPYC1 in the pyrenoid of Chlamydomonas has several broad implications. Firstly, specific to the findings within this study, the condensation of Chlamydomonas Rubisco into a condensate that recovers the function of the pyrenoid-based CCM and growth under ambient CO_2 conditions provides good support for the role of CsLinker in the pyrenoid of Chlorella, somewhat mitigating the lack of direct evidence in Chlorella precluded by the lack of genetic tractability. By transplanting CsLinker into a strain in which the native Chlamydomonas RbcS had been replaced with that of Chlorella, it was also possible to tentatively conclude that the correct localisation of the Rubisco condensate in Chlamydomonas is indeed underpinned by interaction of pyrenoid assembly proteins with the RbcS, as proposed previously (Meyer, Itakura, *et al.,* 2020). Although this ΔEPYC1ΔRbcS strain was marred by inheritance of historical off-site mutations that prevented direct comparison with ΔEPYC1 strains (Fig. 5-7), the approach pioneered here provides a useful tool for future pyrenoid engineering efforts. In particular, testing the localisation and functionality of the PMST and PMTT proteins from Chlorella in Chlamydomonas strains which possess a CsLinker-derived Rubisco condensate will likely be critical in demonstrating the nature of these proteins and confirming the hypothesised shared principals of RBM-guided pyrenoid assembly outlined in Chapter 2.

More generally, the replacement of EPYC1 with CsLinker *in vivo* exists as the only currently available example of functional complementation through phase separation by a protein with no sequence similarity to the replaced protein. In the field of general phase separation biology this is a compelling example that the physicochemical properties more so than the primary sequence of intrinsically disordered proteins underpin the functionality of the proteins. Extending this, the apparently concurrent convergent evolution of CsLinker and EPYC1 ~250 Mya (Fig. 2-21), yet disparate nature of their sequences, enforces that a certain set of constraints for the physicochemical nature of the interaction with Rubisco needed to be fulfilled to allow functionality. This fact was outlined by the similarity in

composition, yet altogether different locale of interacting residues at the interfaces of both EPYC1 and CsLinker with their cognate Rubiscos (Fig. 4-13) (He *et al.,* 2020). This provides a unique example in which two independently evolved proteins underpinning phase separation with the same functionality can be directly compared. Future work to compare the 'fitness' of pyrenoids underpinned by phase separation of linker protein variants from each species (*e.g.,* all Chlorella/Chlamydomonas linker variants identified in Chapter 2; Fig. 2-12) alongside the mesoscopic and molecular properties of the resultant condensates will likely advance understanding of both pyrenoid linker evolution as well as general constraints for the evolution of functional phase separation.

By mutation of the putatively shared CsLinker interface in the RbcL of Chlamydomonas Rubisco it was empirically demonstrated that the functional cross-compatibility was indeed underpinned by interaction with the same interface in Chlamydomonas Rubisco (Fig. 5-20). By comparison across the green lineage, it was subsequently demonstrated that the same interface was conserved in most lineages that evolved prior to divergence of the spermatophytes (Fig. 5-22) and *in vitro* characterisation of Ulva (Fig. 5-15) and fern (Fig. 5-17) Rubiscos accordingly demonstrated that this translated to phase separation also. Consistent with the mutation made in the Chlamydomonas Rubisco, deviation of this interface by even one of the key CsLinker-interacting residues completely abolished cross-compatibility of the CsLinker, as observed with Spinach and *Arabidopsis* Rubiscos (Fig. 5-18).

With the previous findings of cross-compatibility of CsLinker in mind, the most surprising finding of this chapter was the phase separation of Solanaceae Rubiscos (Fig. 5-23, Fig. 5-24). Given that sequence alignments demonstrated the ablation of at least one electrostatic interaction in addition to an altered βC-D loop (Fig.5-22), it was expected that these Rubiscos would behave in a similar manner to those of Spinach and *Arabidopsis*. Instead, all tested Solanaceae Rubiscos phase separated with CsLinker, consistent with shared use of the same interface. It should be emphasised that there were clear differences in the properties of droplets formed with Solanaceae Rubiscos when compared to those of the green algal and fern Rubiscos. It is likely these differences are underpinned by differences in interaction between the two proteins because of the highlighted sequence substitutions in the RbcL. To

try and rationalise the basis of interaction of Solanaceae Rubiscos with CsLinker, a hypothesised compensatory interaction that utilises another electrostatic interaction at the same interface was proposed based on the 3D structure of *Nicotiana* Rubisco (Fig. 5-25). This scenario appeared more likely than alternatives explanations, such as the use of an altogether different interface that gives rise to similar affinity and properties, though this cannot be ruled out from these data. For consideration, EPYC1 was previously shown to form droplets with cyanobacterial Rubisco of *Synechococcus* PCC630 likely through use of an alternative interface on Rubisco (Wunder *et al.,* 2018). Determining the structural basis of the interaction between CsLinker and Solanaceae Rubisco will allow conclusion of this hypothesis and provide further insight into the permitted tunability of the interaction between CsLinker and Rubisco. In a hugely significant step for plant engineering of pyrenoids, the *in vitro* phase separation of Rubisco also appeared to translate to *in planta* phase separation of native *Nicotiana benthamiana* Rubisco (Fig. 5-26). This was the first demonstration of phase separation of native plant Rubisco and demonstrated the feasibility of engineering pyrenoid-based CCMs *in planta* without the modification of Rubisco, thus shortcutting complex modifications previously required for EPYC1-based approaches (Atkinson *et al.,* 2020). Despite this, it has not yet been demonstrated that Rubisco is also present in the condensates formed *in planta*, though ongoing characterisations suggest this is the case. Additionally, it is unclear why there are multiple condensates in each chloroplast and whether this is artefactual or biological and whether it will have implications for functioning of *in planta* pyrenoids.

Finally, the apparent compensatory interaction mechanism observed with Solanaceae Rubiscos provides great promise for future engineering efforts in non-Solanaceae Rubiscos. As proposed to occur with Solanaceae Rubisco, it seems likely that a single electrostatic interaction may need to be compensated for in most non-Solanaceae Rubiscos to allow their phase separation by CsLinker (Fig. 5-22). Accordingly, a large focus of future work should be to modify CsLinker to provide this compensatory interaction in those Rubiscos. Based on the findings of this chapter and those previous, a general approach to pursue this goal is outlined here. Designing and testing of many full length linker protein variants is unfeasible due to difficulties in cloning, synthesis, and purification. Instead, an approach initially utilising the high affinity, ease of cloning and purification and amenability of α3-α4 fragment

variants could be pursued (Fig. 5-27). As such a library of α3-α4 fragment variants modified at the helices could be designed and probed for their affinity against representative target Rubiscos. Design of variants could be completed either intelligently based on 3D structures of target Rubiscos, using native RBM variants from bioinformatic analyses or by high throughput amino acid scanning approaches. α3-α4 variants with affinities like those of the cognate interaction (1 μM < K_D < 2 μM) could then be selected and full length CsLinker variants with the same modifications in the helices produced and tested for *in vitro* phase separation against the same Rubiscos. Finally, trialling transient expression *in planta* in species with the same conserved Rubisco interaction interface could be completed to identify plant-compatible linkers. This approach should allow fast-tracking of the design of compatible linkers for other plant Rubiscos, or indeed other Rubiscos in general. Equally the ongoing characterisation of linker proteins identified by FLIPPer may also reveal binding to an even more conserved region of Rubisco that is natively compatible with non-Nightshade Rubiscos.

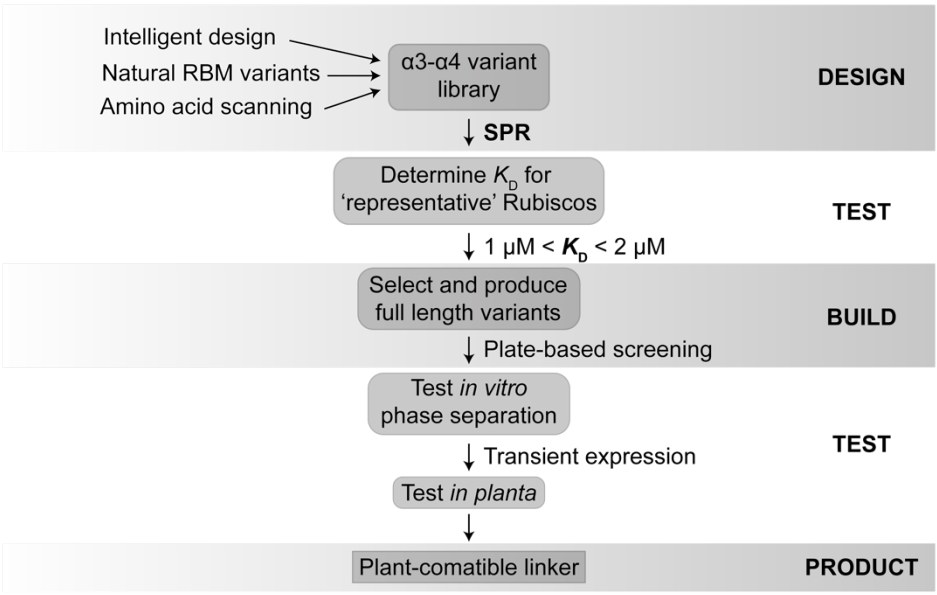


Figure 5-27 | A strategy for design of a plant-compatible linker for non-Solanaceae Rubiscos.
Schematic diagram of a proposed design strategy for design of a plant-compatible linker protein. The general phase at each stage is indicated to the right.

Taken together, the results of this chapter build on those of the previous and demonstrate that the RbcL-interaction interface of CsLinker affords the protein cross-compatibility with much of the green lineage

and provides great promise for future pyrenoid engineering approaches in plants. Furthermore, the findings reinforce the convergent evolution of function of CsLinker and EPYC1 and begin to provide clues as to the nature of the constraints during their evolution. These findings have broad implications for general phase separation as well as pyrenoid biologists.

5.6. Materials and methods

Unless otherwise stated, all strains and culture conditions, general molecular biology and protein biochemistry methods were performed consistent with the methods outlined in Chapters 3 and 4. General methods introduced in this chapter are outlined here.

5.6.1. General molecular biology methods

5.6.1.1. Chlamydomonas check PCR

For all PCR amplifications from *Chlamydomonas* nuclear and chloroplast DNA, whole cell DNA was extracted by resuspension of a colony in 50 µL of 10 mM EDTA (pH 8.0) and incubation for 10 minutes at 99 ºC in a thermal cycler. The resulting solution was centrifuged for ~10 minutes at 2,500 *g* and 1 µL of the supernatant was used in subsequent PCR reactions that had composition and conditions as in table 5-3 and table 5-4 respectively. The annealing temperature (T_m) for each reaction was calculated using the New England BioLabs online T_m calculator set to One*Taq* Hot Start Quick-Load 2x Master Mix with GC buffer, after which 2.5 ºC was subtracted from the calculated T_m. In all cases an extension time of 60 seconds per kilobase of target amplicon was used. 10 µL of the reaction mix following amplification was resolved by DNA gel electrophoresis as described previously.

Table 5-3: *Chlamydomonas* check PCR composition.

Component	Volume (µL)
dH_2O	8
2x One*Taq*® Hot Start Quick-Load® Master Mix w/ GC buffer (M0849)	12.5
10x One*Taq*® High GC enhancer (B9026)	2.5
Forward primer (10 µM stock)	0.5
Reverse primer (10 µM stock)	0.5
DNA	1
Total	**25**

Table 5-4: *Chlamydomonas* check PCR conditions.

Temperature (ºC)	Time (s)	Number of cycles
94	300	1
94	30	
Calculated	45	38
68	Calculated	
68	300	1

5.6.2. General Chlamydomonas methods

5.6.2.1. Chlamydomonas nuclear transformation

For all transformations of *Chlamydomonas* for nuclear expression of transgenes, a protocol based on previous work was used (Yamano & Fukuzawa, 2020). Relevant *Chlamydomonas* strains were grown to mid-exponential phase (2-8x10^6 cells mL^{-1}) in TAP media supplemented with the relevant antibiotics. 112 µL of cells at a concentration of ~2x10^8 cells mL^{-1} were incubated on ice in 0.2 cm gap electroporation cuvettes (EP-202 – Cell Projects) for ~30 minutes prior to electroporation. For each transformation 8 µL of linearised plasmid, containing ~1 µg of material was mixed with the cells immediately prior to electroporation, bringing the final volume to 120 µL. Electroporation was performed according to the conditions in Table 5-5 using a NEPA21 square pulse electroporator (NepaGene). Following electroporation, cells were immediately recovered in 7 mL of TAP + 40 mM sucrose in 15 mL falcon tubes overnight in very low light (~2-5 µmol photons m^{-2} s^{-1}) with agitation for ~24 hours before plating on TAP or TP plates. Recovery strategies varied between constructs and are indicataed at the relevant points.

Table 5-5: NEPA21 electroporation parameters.

Component	Voltage	Pulse length	Pulse interval	# of pulses	Decay rate	Polarity
Poring pulse	300 V	8 ms	50 ms	2	40	+
Transfer pulse	20 V	50 ms	50 ms	1	-	±

5.6.2.2. Chlamydomonas chloroplast transformation

Relevant *Chlamydomonas* strains were grown to mid-exponential phase (2-8x10^6 cells mL^{-1}) in TAP media supplemented with the relevant antibiotics. 1x10^7 cells mL^{-1} in a volume of ~100 µL were plated on TAP plates and distributed to a diameter of ~3 cm in the centre of the plate. The plated cells were allowed to recover for ~2-4 hours under low light conditions (~50 µmol photons m^{-2} s^{-1}). DNAdel™ 550 nm gold nanoparticles (Discontinued – Seashell Technologies) were resuspended to 50 mg mL^{-1} in the supplied binding buffer. For each transformation, 0.5 mg of gold was resuspended with 1-2 µg of plasmid DNA to a final gold concentration of 30 mg mL^{-1}. An equal volume of precipitation buffer (50 mM spermidine, 1M $CaCl_2$) was added to the suspension and allowed to incubate for 3 minutes following vortexing for ~10 seconds. The solution was subsequently centrifuged for ~10 seconds at 10,000 *g* and the supernatant removed. 500 µL of 100% ice-cold ethanol was added, the particles were resuspended, then pelleted in the same manner. The supernatant was removed and the gold solution was resuspended in 100% ethanol to 50 mg mL^{-1}. 10 µL of the solution was pipetted onto the centre of sterile macrocarriers (1652335 – Bio-Rad) and allowed to dry. A Bio-Rad PDS-1000/He™ system, with 1100 psi rupture disks (1652329 – Bio-Rad), was used for particle bombardment in which the transformation plate was placed ~12 cm from the rupture disk. Following bombardment, plates were recovered for ~24 hours in low light conditions. The recovered cells were resuspended in 250 µL of TAP media and plated on 3 separate 90 mm TAP or TP plates. Recovery strategies varied between constructs and are indicated at the relevant points.

5.6.2.3. Chlamydomonas spot tests

Relevant *Chlamydomonas* strains were grown to mid-exponential phase (2-8x10^6 cells mL^{-1}) in TAP media supplemented with the relevant antibiotics. Cells were pelleted and resuspended in TP media to a concentration of 1x10^6 cells mL^{-1} from which 10-fold serial dilutions were made. 10 µL of the relevant cellular concentrations were plated on either TP or TAP plates and allowed to dry. TAP plates were incubated at 10 µmol photons m^{-2} s^{-1} for ~2 weeks prior to imaging. TP plates were incubated under different CO_2 conditions in custom-built growth chambers with continuous supply of CO_2 in air. The TP plates were initially incubated for 24 hours at 50 µmol photons m^{-2} s^{-1}, followed by 150 µmol photons

m^{-2} s^{-1} for 24 hours and ~3-5 days at 300 µmol photons m^{-2} s^{-1}. Plate images were captured with Phenobooth (Singer Instruments).

5.6.2.4. Chlamydomonas growth rate screening

Relevant strains were arrayed from transformation plates to 96-pin format on TAP plates using the Stinger attachment for a RoToR+ colony propagation robot (Singer Instruments), using coordinates outputted from Phenobooth screening of the transformation plates. Colonies >5 pixel diameter, with clear separation from other colonies were selected for transfer. After re-array to TAP plates, the strains were grown for 2 generations without antibiotic selection and replicated using the pad attachment and 96-format long pins for the RoToR+. Ensuring even growth across the plates, the colonies were then replicated to TP plates using the same methodology. The plates were incubated in Ziploc bags (351126 – Ziploc) supplemented with 3% CO_2 under 100 µmol photons m^{-2} s^{-1} light and imaged periodically using the Phenobooth instrument. For quantification of colony size, the raw images from the Phenobooth were imported to Fiji and the threshold function was used to segment the colonies from the background. The area of the colony on the plate was then quantified and plotted using GraphPad Prism 9.

5.6.2.5. Screening of Chlamydomonas colonies plates for fluorescence

Chlamydomonas transformation plates were directly screened for fluorescence using a Typhoon 8610 Variable Mode Imager (Discontinued – GE Healthcare). Colonies were screened from 88 mL TAP agar plates poured in RoToR PlusPlates (PLU-003 – Singer Instruments). The screening was completed at 3 mm above the scanner plate surface. The excitation wavelength, emission filter and PMT settings used were according to Table 5-6.

Table 5-6: Typhoon 8610 screening parameters.

Channel	Excitation wavelength (nm)	Emission Filter	PMT (V)
Chlorophyll	633	670 nm BP 30	300
Venus	532	555 nm BP 20	800
mCherry	532	610 nm BP 30	800

5.6.2.6. Confocal fluorescence microscopy of Chlamydomonas strains

All confocal fluorescence imaging was completed using a Zeiss LSM 880 Airyscan microscope in Airyscan mode. Zeiss LSM 880 Airyscan microscope was used in confocal mode without use of the Airyscan module. In all cases a 63x 1.4 numerical aperture (NA) Plan-Apo oil-immersion lens (Carl Zeiss) was used with Carl Zeiss™ Immersol™ immersion oil (518F – Carl Zeiss). 10 µL of cells were left to settle in iBidi µ-Slide 18-well chamber slides (81811 – iBidi) for ~5-10 minutes, after which 30 µL of 1% (w/v) low melting point TP-agarose was added to the chamber and allowed to set.

5.6.3. Generation of the ΔEPYC1ΔRbcS strain

Paromomycin-resistant ΔEPYC1 strain CC-5360 (mt-) and Zeocin-resistant ΔRbcS strain CC-4691 (mt+) were restreaked on TAP plates with no nitrogen (TAP N0) for 3 generations (~3 weeks). Prior to restreaking on TAP N0, the cells were grown in a dense patch on TAP plates, which was subsequently transferred between each restreak. After growth on N0 plates, the strains were separately transferred to TAP N0 liquid media and resuspended to a concentration of ~$5x10^5$ cells mL^{-1}. The flasks were incubated at ~10 µmol photons m^{-2} s^{-1} for 30 minutes with shaking. After 30 minutes, the two strains were equally mixed in the same flask and incubated at ~50 µmol photons m^{-2} s^{-1} for 2 hours with no shaking. After this, 200 µL of the suspension was plated on TAP 3% agar (w/v) plates and allowed to dry without spreading and the plates were incubated at ~30 µmol photons m^{-2} s^{-1} overnight, then incubated in the dark for ~2 weeks. Using a sterilised scalpel, the vegetative cells were scraped away from the plate and the plate was exposed to the fumes of 100% chloroform at a height of ~6 cm for 30 seconds. The region containing the remainder of the cells was excised and incubated overnight in TAP liquid media with no shaking at ~10 µmol photons m^{-2} s^{-1}. The following day, TAP + antibiotics was added to a final concentration of 2.5 µg mL^{-1} Paromomcyin and 0.5 µg mL^{-1} Zeocin. After 3 weeks of growth at ~30 µmol photons m^{-2} s^{-1} with shaking, 250 µL of the cell suspension was plated on TAP + 20 µg mL^{-1} Paromomcyin and 10 µg mL^{-1} Zeocin 1% agar (w/v) plates and grown until single colonies could be restreaked for further analysis. PCR screening was completed according to Fig. 5-5, and the primers used are listed in Table 5-7.

Table 5-7: Primers used for screening ΔEPYC1ΔRbcS strains.

Name	Purpose	Oligonucleotide Sequence
oLM002	Paromomycin cassette forward	GACGTTACAGCACACCCTTG
oLM490	WT EPYC1 locus forward	GGCCGTGCACTCAATTAA
oLM491	WT EPYC1 locus reverse	CGAAGGACAGTAACAGTCG
oLM851	Rbcs2 locus reverse	ACTGATCAGCACGAAACGGG
oLM1625	RbcS2i1 forward	TCTGTCGCTGTCTCAAGCAG
oLM1764	WT RbcS1/2 locus reverse	CCTGCTTCTGGTTGTCGAAG
oLM1765	WT RbcS1/2 loccus forward	GCAGGATGTTCGAGACCTTC

5.6.4. Cloning of CsLinker for expression from the nuclear genome in Chlamydomonas

The TargetP 2.0 predicted mature CsLinker sequence (Fig. 3-1c) was back-translated and codon-optimised using the Geneious back translation tool and Kazusa *Chlamydomonas* codon usage table. 3 copies of the *Chlamydomonas reinhardtii* RbcS2 intron 1 were inserted in the sequence at ~500 bp intervals at between adjacent Gaunine nucleotides, according to previous recommendations (Baier *et al.,* 2018). The resulting sequence was synthesised by TWIST bioscience as 4 separate parts L-1 parts. Part 2 could not be synthesised and was instead PCR-amplified from *Chlorella sorokiniana* UTEX1230 genomic DNA using oLM2398/2399. The 4 parts were assembled at L0 by Golden Gate Assembly with BpiI to create pLM781. The resulting construct was sequenced with oLM271/272. The sequenced L0 CsLinker from pLM781 sequence was separately assembled with the indicated promoters and terminators (Table 5-2) from the *Chlamydomonas* MoClo toolkit (Crozet *et al.,* 2018), alongside the mVenus part (pCM0-113) into pICH47742. The resulting vectors were sequenced with oLM1271/1272 and assembled separately alongside the L1 Hygromycin resistance cassette (pLM631 – Table 3-16) into pAGM4723. The resulting L2 vectors were transformed in *Chlamydomonas* ΔEPYC1 strain as described. To create the pPSAD/tCytC6 construct (#19 – Table 5-2), the L0 CsLinker sequence was PCR-amplified from pLM781 using primers oLM3059/3060 and Gisbon Assembled into the backbone of pLM031 (Table 3-16) that was PCR-amplified with oLM3061/3062. The resulting colonies were screened by colony PCR with oLM054/258 and sequenced with the same primers.

Table 5-8: Primers used for cloning of CsLinker for expression from the nucleus.

Name	Purpose	Oligonucleotide Sequence
oLM2398	CsLinker part 2 gDNA forward	ATCATGCTAGAAGACTACTGAGCGCCGAGC AGAAGGCA
oLM2399	CsLinker part 2 gDNA reverse	CTGAATCAGGAAGACATGCCGCGGGTCTCC ACGGGGGCGGGGC
oLM271	L0 sequencing forward	AGTTGGAACCTCTTACGTGC
oLM272	L0 sequencing forward	TTTGAGTGAGCTGATACCGC
oLM1271	L1 colony PCR/sequencing forward	CACGCCCTTTTAAATATCCG
oLM1272	L1 colony PCR/sequencing reverse	TGGCAGGATATATTGTGGTG
oLM3059	CsLinker Gibson Assembly forward	CGTGAAGACCGTGAAGGTCGCCGCCCGCAG CGGTGGCCGTG
oLM3060	CsLinker Gibson Assembly forward	GCTGCCACGACGGCTGTTGCGACG
oLM3061	Backbone Gibson Assembly Forward	CGTCGCAACAGCCGTCGTGGCAGCGGCCAC CACCATCACCACC
oLM3062	Backbone Gibson Assembly Reverse	GGCGGCGACCTTCACGGTCTTCACGCGGCC AGAGACG
oLM054	PSAD promoter colony PCR/sequencing forward	TTAGGTGTTGCGCTCTTGAC
oLM258	PSAD 3'UTR colony PCR/sequencing reverse	AACCAAGTCGACGTGCCTAC

5.6.5. Cloning of CsLinker for expression from the chloroplast genome in Chlamydomonas

The TargetP 2.0 predicted mature CsLinker sequence (Fig. 3-1c) was back-translated and codon-optimised using the Geneious back translation tool and Kazusa *Chlamydomonas* chloroplast codon usage table. The sequence was synthesised in 3 parts by TWIST bioscience. The resulting plasmids (pLM1375, pLM1376 and pLM1377) were assembled by Golden Gate Assembly using BsaI into the L2 acceptor vector provided by René Inckemann (pLM1374). The resulting constructs were sequenced with oLM3354/3355 and gave rise to pLM1385 which had no epitope or fluorescent tag on the CsLinker sequence. To create the mVenus- and 3xFLAG-tagged variants, part 1 of the CsLinker sequence was PCR-amplified from pLM1375 using primers oLM3395/2403 to modify the 5' syntax. The amplicon was gel-extracted for Golden Gate Assembly. For the mVenus construct, the mVenus sequence was PCR amplified from an existing mVenus vector (pLM211), which was optimised for chloroplast expression using codon usage optimizer (https://github.com/khai-/CUO), using primers oLM271/oLM3396 to

modify the 3' syntax. The resulting amplicon was also gel-extracted. For the 3xFLAG epitope, the sequence was codon-optimised as described for the CsLinker sequence and ordered as forward and reverse primers (oLM3397/3398) which were hybridised by mixing at 100 µM concentration, heating to 99 ºC in a thermal cycler and allowing to cool to room temperature. The gel extracted syntax-modified part 1 of CsLinker, pLM1376, pLM1377 and the mVenus amplicon were assembled by Golden Gate Assembly using BsaI into pLM1374 to create pLM1450 which contained the mVenus-tagged CsLinker. Separately, the gel extracted syntax-modified part 1 of CsLinker, pLM1376, pLM1377 and the hybridised 3xFLAG part were assembled into pLM1374 by Golden Gate Assembly with BsaI to create pLM1449 which contained the 3xFLAG-tagged CsLinker. The resulting colonies were sequenced using oLM3354/3355.

Table 5-9: Primers used for cloning of CsLinker for expression from the nucleus.

Name	Purpose	Oligonucleotide Sequence
oLM3354	L2 sequencing primer forward	GCGTTGCTAATGGTGTAAATAATG
oLM3355	L2 sequencing primer reverse	GACGTCCTGCCAACTGC
oLM3395	CsLinker part 1 syntax-modification forward	TACAGACTACGAGGTCTCTTTCAGCTATGGC AACTGCTCAATTATCAGC
oLM2403	CsLinker part 1 reverse	AGGATTGTGCCCTCTGAGCG
oLM271	mVenus forward	AGTTGGAACCTCTTACGTGC
oLM3396	mVenus syntax-modification reverse	GAGTCGTAGTGAGGTCTCGTGAACCTTTGTA TAATTCATCCATACCTAAAGTAATACCAG
oLM3397	3xFLAG hybridisation forward	AGACATGATGGTCTCAAATGGACTACAAAG ACGATGACGATAAAGGTGATTACAAAGACG ATGACGATAAAGGTGACTATAAAGACGATGA CGATAAAGGTTCACGAGACCGTCAGCATC
oLM3398	3xFLAG hybridisation reverse	GATGCTGACGGTCTCGTGAACCTTTATCGTC ATCGTCTTTATAGTCACCTTTATCGTCATCGT CTTTGTAATCACCTTTATCGTCATCGTCTTTG TAGTCCATTTGAGACCATCATGTCT

5.6.6. Cloning of RbcL chloroplast knockout and expression vectors

The ΔRbcL plasmid (pLM084) was previously constructed by Professor Luke Mackinder by Gibson Assembly of the AadA gene, conferring spectinomycin resistance, amplified from *Chlamydomonas* strain CC-5168 (Economou *et al.,* 2014), into plasmid P-67 cpDNA EcoRI 14 (Chlamydomonas Resource Centre) to replace the RbcL CDS in frame. To create the *Chlamydomonas* RbcL chloroplast expression vector, the WT RbcL sequence was PCR-amplified from the chloroplast DNA of strain CC-4533 using primers oL3555/3556 and the amplicon was subsequently gel-extracted. The amplicon was Gibson Assembled with the backbone of pLM084 to replace the AadA CDS in frame. The pLM084 backbone was amplified using oLM3401/3402. The resulting vector (pLM1525) contained the WT *Chlamydomonas* RbcL and homology arms for the RbcL locus in the chloroplast genome. The vector was sequenced with oLM337/338. To create the D86H mutated RbcL expression vector primers containing the relevant mutation were used to amplify the pLM1525 vector, which was subsequently gel extracted. 1 µL of the gel-extracted product was used in a 10 µL KLD enzyme mix (M0554 – New England BioLabs) reaction and incubated . The resulting construct (pLM1533) was sequenced with oLM337/338 and contained the D86H mutated Rubisco.

Table 5-10: Primers used for cloning of RbcL chloroplast expression vectors.

Name	Purpose	Oligonucleotide Sequence
oLM3555	RbcL forward	TCGTCAAAAGAAGTTACATTTATTTATATAAA TGGTTCCACAAACAGAAACTAAAG
oLM3556	RbcL reverse	TCACATAAACATCATGAAAAATAAAAATTAA AGTTTGTCAATAGTATCAAATTCG
oLM3401	pLM084 backbone forward	TTATATAAATAAATGTAACTTCTTTTGACGA
oLM3402	pLM084 backbone reverse	TTTTTATTTTTCATGATGTTTATGTGA
oLM337	RbcL 5'UTR sequencing forward	AAAGGCCCTTTCTATGCTCG
oLM338	RbcL 3'UTR sequencing reverse	ACCGATTGAGTTACATCCGC
oLM3549	RbcL D86H forward	GCTCGATATGATAACAACGACC
oLM3550	RbcL D86H reverse	CAGTTCCAGGTGAAGAAAACC

5.6.7. Rubisco purifications

The D86H-mutated *Chlamydomonas* Rubisco was purified exactly as described previously in Chapter 3 for the purification of WT *Chlamydomonas* Rubisco. Rubisco purifications of *Ulva*, Fern, Spinach, *Arabidopsis*, *Nicotiana*, Tomato, Potato were completed in the same manner with the modifications described here.

All seaweed and plant material was separated from stem, root and debris and washed in distilled water 3 time prior to processing. Instead of sonication, the material was lysed by blending and homogenisation in the same buffer. For the blending, washed and processed material was mixed with a roughly equal volume of lysis buffer, which had been pre-cooled to ~4 ºC. The material was blended in 5, 30 second pulses with ~5 minute breaks to allow the solution to cool. The blended material was then homogenised in batches using a 55 mL Potter-Elvehjem homogenizer (15301331 – Fisher Scientific). The lysate was clarified and processed according to the protocol used previously for Chlorella and *Chlamydomonas* Rubiscos.

Ulva sp. material was sourced from intertidal rocks west of Staithes Harbour, North Yorkshire (Fig. 5-27). Material was transported in sea water and processed on the same day to avoid degradation of the sample. *Adiantum raddianum* (Fragrans variant) fern plants were sourced from Dean's Garden Centre, York. The plants were maintained in ambient conditions until processing of the leaf material. *Spinacia oleracea* (Spinach) leaf material was sourced as 'baby spinach, 300g' bags from Co-op Food, Hull Road, York and was processed the same day as purchased. *Arabidopsis thaliana* (Col-0) rosette material was kindly provided by Jason Daff (University of York Horticulture staff). 5- week old *Nicotiana benthamiana* plants and Potato leaf material were kindly provided by Dr. Liat Adler (University of York). Tomato plant material was kindly provided by Sara Barrett.

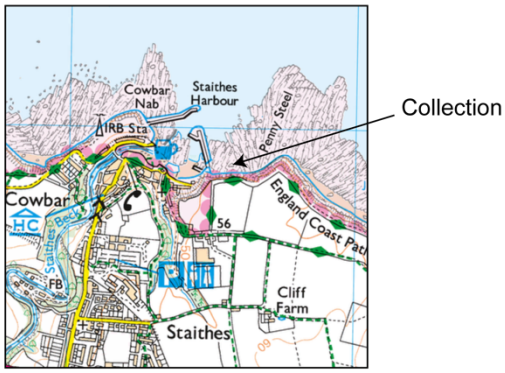


Figure 5-28 | *Ulva* material collection location.
Ordnance survey 1:25,000 map view of the coastline at Staithes with the collection location of the material indicated.

5.6.8. Construction of crop Rubisco phylogenetic chronogram

The consensus phylogenetic chronogram in Fig. 5-22 was built using previously determined divergence estimates from literature, where the average value was used in all cases. The values for divergence of Rice from oat and oat from wheat/barley were taken from (Peng *et al.,* 2022). The estimate for wheat and barley divergence was taken from (Middleton *et al.,* 2014). The estimate for sugarcane divergence was taken from (Kellogg, 2001). Estimates for the divergence of plantain, onion and yam were taken from (Hertweck *et al.,* 2015). Estimates for divergence points of most Rosid species (grape, soy, alfalfa, cassava, cotton and *Arabidopsis*/Rapeseed) and for divergence of rubber and coffee were taken from (Zeng *et al.,* 2017). An estimate for hemp and apple divergence is presented based on phylogeny, with no time-calibration. Rapeseed/*Arabidopsis* divergence timepoint was taken from (Yang *et al.,* 1999). Spinach and Beet divergence was taken from (Xu *et al.,* 2017). Divergence timepoints of the Solanales species (sweet potato, pepper, aubergine, potato, tomato and tobacco) were taken from (Särkinen *et al.,* 2013). The remainder of the divergence points were taken from previous analysis (see Chapter 2 methods).

5.6.9. In planta expression of CsLinker (with Yuwei Mao)

Cloning, transient expression and imaging of tagged CsLinker in *Nicotiana benthamiana* was completed by Yuwei Mao (University of Edinburgh). The mature CsLinker sequence was codon optimised and synthesised using GeneArt and the *Nicotiana benthamiana* codon usage table

(ThermoFisher Scientific). The sequence was golden gate assembled into pICH47732 with a 35S promoter, AtRbcS1A transit peptide, C-terminal Turbo-GFP or mNeonGreen and HSP+NOS double terminator according to (Engler *et al.,* 2014). Overnight cultures of electrocompetent GV3101 Agrobacterium harbouring the relevant plasmids were grown overnight in LB. Cultures were resuspended in 10 mM $MgCl_2$ to OD_{600} of 0.8 and syringe infiltrated into the youngest fully expanded leaves of four-week old *Nicotiana benthamiana* plants. Images were captured using a Leica SP8 confocal microscope 48 hours after infiltration and incubation of plants at 25 °C, 16h light 8h dark, 170 µM photons m^{-2} s^{-1}.

Table 5-11: Plasmids used in Chapter 5.

Name	Description	Use	Origin
pODC13	CrRbcS2 (pAR/tRbcS2)	Creation of ΔEPYC1ΔRbcS::RbcS2 and ΔRbcS::RbcS2	(Caspari, 2020)
pLM036	EPYC1	Complementation of EPYC1 in relevant strains	(Mackinder *et al.,* 2016)
pLM804	Part 1 nuclear-optimised CsLinker	Assembly of CsLinker nuclear expression vectors	TWIST Bioscience
pLM709	Part 3 nuclear-optimised CsLinker	Assembly of CsLinker nuclear expression vectors	TWIST Bioscience
pLM805	Part 4 nuclear-optimised CsLinker	Assembly of CsLinker nuclear expression vectors	TWIST Bioscience
pLM781	L0 nuclear-optimised CsLinker	Assembly of CsLinker nuclear expression vectors	This study
pICH47742	L1 Golden Gate acceptor vector	Assembly of CsLinker nuclear expression vectors	(Engler *et al.,* 2014)
pCM0-113	L0 nuclear expression mVenus part	Assembly of CsLinker nuclear expression vectors	(Crozet *et al.,* 2018)
pLM631	*Chlamydomonas* Hygromycin cassette	Assembly of CsLinker nuclear expression vectors	(Crozet *et al.,* 2018)
pAGM4723	L2 Acceptor vector	Assembly of CsLinker nuclear expression vectors	(Engler *et al.,* 2014)
pLM031	EPYC1 (pPSAD/tCytC6) vector	Assembly of CsLinker nuclear expression vectors	(Mackinder *et al.,* 2016)
pLM1374	L2 *Chlamydomonas* chloroplast expression vector	Assembly of CsLinker chloroplast expression vectors	René Inckemann
pLM1375	Part 1 chloroplast optimised CsLinker	Assembly of CsLinker chloroplast expression vectors	TWIST Bioscience
pLM1376	Part 1 chloroplast optimised CsLinker	Assembly of CsLinker chloroplast expression vectors	TWIST Bioscience
pLM1377	Part 1 chloroplast optimised CsLinker	Assembly of CsLinker chloroplast expression vectors	TWIST Bioscience
pLM1385	Untagged CsLinker chloroplast expression	Expression of untagged CsLinker in ΔEPYC1	This study
pLM1449	3xFLAG-tagged CsLinker chloroplast expression	Expression of 3xFLAG-CsLinker in ΔEPYC1	This study
pLM1450	mVenus-tagged CsLinker chloroplast expression	Expression of mVenus-CsLinker in ΔEPYC1	This study
pLM084	*Chlamydomonas* ΔRbcL expression	RbcL KO and cloning of WT RbcL and D86H RbcL	Prof. Luke Mackinder
pLM1525	*Chlamydomonas* RbcL WT expression vector	Expression of WT Chlamydomonas RbcL	This study
pLM1533	*Chlamydomonas* D86H RbcL expression vector	Expression of D86H Chlamydomonas RbcL	This study

6. DISCUSSION

At the outset of this study, an overarching hypothesis was defined:

> "The fundamental hypothesis of this study was that linker proteins commonly underpin the condensation of Rubisco in pyrenoids across the tree of life"

The motivation for interrogating this hypothesis was also outlined:

> "The primary motivation was to discover novel approaches to pyrenoid assembly in nature that could benefit ongoing efforts to engineer pyrenoid-based CCMs in plants"

The overarching hypothesis was interrogated through bioinformatic, biochemical, structural, and *in vivo* means throughout this thesis in Chapters 2, 3, 4 and 5 respectively. The results of these interrogations presented three broad major advances that are discussed separately below:

1) Pyrenoids in UTC algae convergently and concurrently evolved during a period late in the Paleozoic Era (~260-275 Mya).

2) The pyrenoid linker protein from *Chlorella sorokiniana* (CsLinker) can cross-complement functional Rubisco condensation in *Chlamydomonas* across an ~800 My evolutionary gap.

3) CsLinker demonstrates cross-compatibility with native crop Rubiscos of the nightshade family of plants – providing a route for easier plant engineering.

The context of these findings and their implications for future directions are subsequently discussed.

6.1. Pyrenoids in UTC algae evolved convergently and concurrently

The majority of the work completed in Chapter 2 was completed during a COVID-19 enforced lockdown period. Though frustrating, in hindsight this period provided an invaluable reflective window that was transformative to the work completed in this study. Through completing a comprehensive literature review (Barrett *et al.,* 2021), an appreciation of pyrenoid breadth and diversity was garnered that directly influenced the course of the study. The striking observation that pyrenoid assembly is likely commonly underpinned by liquid-liquid phase separation reinforced the hypothesis of Mackinder *et al.* (2016) that EPYC1-like proteins are likely present in diverse pyrenoids. Building directly on the results of the seminal Mackinder *et al.* (2016) study, during the same lockdown period, the development of a Fast Linker Identification Pipeline for Pyrenoids (FLIPPer) was begun.

The results of FLIPPer analysis were mixed with respect to the fundamental hypothesis of the study, as outlined in the discussion of Chapter 2. One clear story based on analysis of genomes in the UTC clade of algae did, however, emerge. The identification of sequences that are functionally and structurally analogous to EPYC1 in *Ulva* species of the Ulvophyceae class and *Chlorella* species of the Trebouxiophyceae class allowed two broad conclusions:

1) Pyrenoid linker proteins definitively evolved convergently in *Ulva*, *Chlorella* and *Chlamydomonas* long after the basal divergence from their shared most recent common ancestor ~800 Mya.

2) Evolution of the linker proteins and probably pyrenoids themselves likely occurred concurrently during a shared 'window of opportunity' ~260-275 Mya when the atmospheric O_2:CO_2 ratio was at its historic maximum.

These findings provided the first direct molecular evidence for the convergent evolution of pyrenoids in green algae and the first time-calibrated molecular evidence for the proposed emergence of pyrenoids in response to extreme challenges for Rubisco imposed by atmospheric conditions.

The proposed emergence timepoint for pyrenoids in UTC algae was largely consistent with previous predictions for CCM emergence (Badger & Price, 2003; Griffiths *et al.,* 2017; Raven *et al.,* 2017). Though the proposed timepoint displayed a satisfying and feasible alignment with atmospheric and terrestrial conditions, many questions emerged from this conclusion.

Why did other pyrenoids not evolve at this time, and why were they lost later?

There is clear indication that hornwort and diatom pyrenoids likely evolved much later than the proposed timepoint for the UTC algae. It remains unclear why pyrenoids were not evolved or passed down from their ancestors that existed during the same period. In hornworts it is also clear that the later evolved pyrenoids were lost in the same period they also evolved (Villarreal & Renner, 2012). A possible explanation for both these phenomena could be the habitats in which the species evolved and subsequently occupied. Most UTC algae exist in freshwater/brackish aqueous environments, whereas diatoms are largely marine-dwelling and hornworts are either semi-aqueous or non-aqueous. It is possible the subsumed environments of UTC algae provided the appropriate conditions for pyrenoid evolution, which were not equivalent in hornworts and diatom ancestors at the time, explaining the proposed evolution much later (in the last 100 My). Equally, hornwort lineages in which pyrenoids have been lost may have subsequently occupied non-aqueous environments that removed the requirement for pyrenoids.

Did functional CCM machinery evolve much earlier or was it co-opted?

The evolutionary timepoint of pyrenoid linker proteins far later than the basal divergence of *Ulva*, *Chlorella* and *Chlamydomonas* species presents an interesting dichotomy. It is well established that homologues of key *Chlamydomonas* CCM machinery (*e.g.,* carbonic anhydrases (CAH3/LCIB) and bicarbonate transporters (BSTs, LCIA)) are conserved across species in the UTC clade (Benlloch *et al.,* 2015; Jin *et al.,* 2016; Mukherjee *et al.,* 2019; Yamano *et al.,* 2015), suggesting they were present in the most recent common ancestor of all UTC algae ~800 Mya. This raises the question of whether a primordial CCM was present in the common ancestor that was presumably devoid of a pyrenoid or whether alternatively, these proteins were co-opted into a pyrenoid-based CCM prior to, or following

the evolution of Rubisco condensation. The latter case would imply a similar convergent evolutionary mechanism for CCM machinery as present for the linker proteins. Though discussed in Chapter 2 for the *Chlorella* pyrenoid, generally the order of pyrenoid-based CCM evolution is unclear and clearly requires further consideration, especially with consideration to ongoing efforts for plant engineering.

Did Rubisco's form and slow evolution play a role in the evolution of the linker proteins?

As outlined in the introduction, the catalytic properties of Rubisco are thought to be amongst the slowest evolving of metabolic enzymes (Bouvier *et al.,* 2023). The catalytic properties are also generally grouped by Rubisco form (Bouvier *et al.,* 2021; Flamholz *et al.,* 2019). These two constraints could have significant implications for the evolution of pyrenoids. Given Rubisco's slow evolution, it is likely that Rubisco is still evolving within pyrenoid-based CCMs to increase their carboxylation rate. Presumably, trajectory towards a faster, less specific Rubisco would provide advantages for pyrenoid-based CCMs. It would be interesting to test this empirically, perhaps by integrating a cyanobacterial Rubisco in the Chlamydomonas pyrenoid-based CCM. The form-specific Rubisco properties could have implications for the timepoint of linker evolution. For example, the relatively high specificity and low catalytic rate of diatom Rubiscos may have afforded their ancestors the ability to survive through the high O_2, low CO_2 environment in which green algal pyrenoids evolved presumably to overcome the relatively poor specificity of their Rubiscos. These considerations have important implications for considering the future efficiency of Rubiscos in engineered pyrenoid-based CCMs, given the ongoing rise in atmospheric CO_2.

In line with primary motivation of this study, identification of the Rubisco-binding motifs (RBMs) in the identified linker sequences allowed the further identification of protein sequences that putatively drive pyrenoid assembly with the ultrastructural features in those species (PMSTs and PMTTs). As hoped, in *Chlorella* this uncovered an apparently simpler blueprint for the assembly of the pyrenoid with direct implications for plant engineering, given the latterly uncovered cross-compatibility of *Chlorella* RBMs with plant Rubiscos. Work to validate the simplicity of this mechanism is already ongoing both *in planta* and *in capsa*, and there is a great deal of optimism surrounding this.

More generally, the approaches undertaken in Chapter 2 demonstrate the impetus of *in silico* analyses yet equally demonstrate the importance of the foundations laid by years of fundamental algal and pyrenoid research that allowed the formation of the hypotheses to be investigated. With this in mind, several key questions remain outstanding that may only be answered with primary biological investigations. Most intriguingly, there was a distinct lack of linker protein analogues in key pyrenoid-containing species with high quality reference genomes (*e.g., Thalassiosira pseudonana*, *Parachlorella kessleri* and all Prasinophytes). This reinforces the possibility that wholly alternative pyrenoid assembly mechanisms exist in nature that are still to be determined. For example, some evidence exists for the presence of DNA in the pyrenoid of *Dunaliella salina* (Oleksienko *et al.,* 2020) – a species in which no strong linker candidate was identified – raising the possibility for a DNA-driven Rubisco condensation mechanism. Equally, the lack of SAGA/PMST analogues in the genomes of *Ulva* species raised the intriguing question of how the starch sheath is attached to the pyrenoid matrix in those species. In the UTC algae, the evolutionary origin of the primary sequences of the linker proteins, and presumably subsequently evolved RBM-containing protein sequences, remains wholly unclear. Furthermore, given the completely independent origin of the linker proteins in each lineage it is intriguing that such a high degree of structural analogy is present between the proteins. This might indicate that the evolutionary drivers in each species were highly similar, relative to drivers in other species with less analogous sequences (*e.g., Phaeodactylum*). Equally, the available parts that could be co-opted into pyrenoid-based CCMs may have been more similar in UTC algae than in diatoms and driven this similarity. The number of outstanding questions surrounding pyrenoid evolution and assembly only serve to showcase the relative infancy of this field of study, and the exciting biology that lies ahead.

6.2. CsLinker cross-complements functional Rubisco condensation across 800 My

Chapter 5 presented the *magnum opus* of this study, which was pursued based on the foundations of the extensive biochemical and structural characterisations of CsLinker undertaken in Chapters 3 and 4. At the outset of the characterisation of CsLinker it was expected that the system would be unremarkably like that of Chlamydomonas. In particular, it was anticipated that the interaction between CsLinker and Chlorella Rubisco would be contributed by a similar Rubisco binding site as observed for EPYC1. The discovery that CsLinker binding to Rubisco occurred in the highly cross-kingdom conserved RbcL was somewhat of a "*Eureka*!" moment for the study and ultimately led to the latter two major findings of the study as outlined above.

From initial *in vitro* observations of CsLinker's cross-reactivity with Rubisco from Chlamydomonas, it was immediately clear that the RbcL-binding of CsLinker would afford a greater degree of cross-compatibility of the protein as compared to EPYC1. In an undeniable showcase of this behaviour the CsLinker was expressed in a Chlamydomonas strain lacking the native linker EPYC1 and demonstrated functional recovery of the pyrenoid-based CCM by driving Rubisco condensation. This behaviour demonstrated 3 key findings empirically:

1) The disordered proteins CsLinker and EPYC1 have definitively convergently evolved the same functionality in the phase separation of Rubisco, though the *modus operandi* may not be so important.

2) Condensation of Rubisco is critical to pyrenoid function and enables assembly with the ultrastructural features.

3) Chlamydomonas can act as a chassis organism to test functionality and assembly of heterologous pyrenoid components.

The recovery of pyrenoid function by complementation of Rubisco condensation with a completely disordered protein represents as far as can be interpreted, the first demonstration of complementation of function by phase separation with a protein bearing no sequence homology to the original. Considering the purely *in silico* identification of CsLinker, this represents a unique finding that demonstrates the power of the integrated approach implemented in this study. Other studies have explored a similar type of complementation, collectively termed 'orthogonal' (Boeynaems *et al.,* 2023), though have always relied on retention of some sequence-conserved domains in the chimeric 'orthogonal' proteins (King & Petry, 2020; Rawat *et al.,* 2021; Shi *et al.,* 2021). The fact that Rubisco condensation by EPYC1 and CsLinker are underpinned by different sequence interactions and give rise to different apparent material properties is the most damning indictment that phase separation, rather than EPYC1-specific interactions are key to the function of the pyrenoid. As discussed in Chapter 5, this sheds some insight into the evolutionary constraints of these proteins. It is posited that a linker protein must have initially evolved from some specific interaction with a surface-exposed Rubisco ('sticker') that had a certain appropriate affinity but had limited apparent positional constraint on the surface of Rubisco. The 'spacer' region between the stickers may then have evolved compensatorily to give rise to the appropriate properties of the interaction (*i.e.,* phase separation of Rubisco).

In addition to important insights into cross-compatibility and linker evolution, the functionality of the hybrid condensate formed in Chlamydomonas provides a useful proof-of-concept that Chlamydomonas can be used as a chassis strain for so-called '*in capsa*' testing of pyrenoid assembly components from heterologous pyrenoids. This is currently being employed to interrogate the blueprints for assembly of the *Chlorella* pyrenoid outlined in Chapter 2 and may more generally offer a rapid testing platform for parts prior to, or in combination with, testing *in planta*. In this sense, the replacement of characterised pyrenoid assembly proteins from Chlamydomonas may be functionally replaced by those of other pyrenoids to demonstrate their functionality. This framework will likely provide a useful tool for interrogating the pyrenoids of non-tractable species (*e.g., Chlorella*).

6.3. CsLinker provides a promising route for plant engineering

Undoubtedly the most exciting output of this study from a personal perspective was the finding that CsLinker can phase separate nightshade plant Rubiscos both *in vitro* and crucially, *in planta*. Whilst this was outlined as a possibility in early hypotheses, it was surprising that indeed a natural, plant-compatible linker protein was identified that could complete this role. This represents a hugely significant step forward for pyrenoid engineering in plants. Utilisation of the RbcL-binding interface of CsLinker in plants will likely overcome the only major drawback of engineering a Chlamydomonas pyrenoid-based CCM into plants, which is that the RbcSs of the host plant need to be replaced in each case to allow interaction with EPYC1. In nightshade plants, there is optimism that CsLinker-based pyrenoid assembly can be completed rapidly with little modification to either the host plant or donor Chlorella sequences. Outside of nightshade Rubiscos, *in vitro* results suggested that the CsLinker was non-compatible and as such, a roadmap for the development of compatible linker proteins for those Rubiscos was outlined in the discussion of Chapter 5.

Currently, all pyrenoid-based CCM engineering efforts are taking place using Chlamydomonas as a donor and *Arabidopsis* as a host (Atkinson *et al.,* 2020), and will likely come to proof-of-concept fruition in the near future. The next stage of engineering for this ambitious field will be to transfer a working prototype into a C_3 crop plant. Here, the continued parallel characterisation of the Chlamydomonas and *Chlorella* pyrenoids may prove invaluable. As outlined above, the major contribution of the Chlorella system is of a plant-compatible linker protein and RBM sequence. Therefore, the molecular insight gained from the engineering of the *Chlamydomonas* pyrenoid *in planta* may be combined with the advantages provided by the Chlorella sequences. For example, there exists the distinct possibility that RBMs may be able to be substituted between proteins of both origins to create functional chimeric proteins. As an example, if the functionality of a Chlamydomonas assembly protein (*e.g.,* SAGA1) is demonstrated *in planta*, the Chlamydomonas RBMs could easily be substituted with those of CsLinker to render the chimera 'plant-compatible'. Through this approach, combined with future efforts in the characterisation of other pyrenoids, a toolbox of proteins from different species could be assembled that can be tested in combination to assemble chimeric pyrenoids in C_3 crops that

would otherwise require extensive engineering to make them amenable to pyrenoid transfer. Previous work outlined in this study has developed the tools to test these proteins rapidly and accelerate the assembly of truly ‘plant-compatible’ pyrenoids. The ultimate goal of this combined approach should be to build a pyrenoid-based CCM that can be transplanted as a gene cassette into plants to enable pyrenoid assembly without modification of the host plant genome. The highly collaborative nature of the pyrenoid engineering community, in particular within the CAPP consortium gives great reason to be optimistic that these goals can be achieved.

6.4. Concluding remarks

By pursuing pyrenoid assembly outside of model species, an entirely novel linker protein was discovered. Characterisations of this protein provided transformative contributions to pyrenoid biology, biological phase separation, and plant engineering prospects that far exceeded expectations at the outset. I am hugely grateful to have been afforded the opportunity to pursue this "blue-sky" approach during my studies and appreciate the contribution this freedom had to my findings. I am hugely excited by the future of this field, and hope that my contributions have accelerated progress towards its transformative potential.

7. REFERENCES

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 2), 213–221.

Adler, L., Díaz-Ramos, A., Mao, Y., Pukacz, K. R., Fei, C., & McCormick, A. J. (2022). New horizons for building pyrenoid-based CO_2-concentrating mechanisms in plants to improve yields. *Plant Physiology*. https://doi.org/10.1093/plphys/kiac373

Adler, L., Lau, C. S., Shaikh, K. M., van Maldegem, K. A., Payne-Dwyer, A. L., Lefoulon, C., Girr, P., Atkinson, N., Barrett, J., Emrich-Mills, T. Z., Dukic, E., Blatt, M. R., Leake, M. C., Peltier, G., Spetea, C., McCormick, A. J., Mackinder, L. C. M., & Walker, C. E. (2023). The role of BST4 in the pyrenoid of *Chlamydomonas reinhardtii*. In *bioRxiv* (p. 2023.06.15.545204). https://doi.org/10.1101/2023.06.15.545204

Ahiahonu, E. K., Anku, W. W., Roopnarain, A., Green, E., Govender, P. P., & Serepa-Dlamini, M. H. (2022). Bioresource potential of *Tetradesmus obliquus* UJEA_AD : critical evaluation of biosequestration rate, biochemical and fatty acid composition in BG11 media. *Journal of Chemical Technology and Biotechnology* , *97*(3), 689–697.

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.

Ainsworth, E. A., & Long, S. P. (2005). What have we learned from 15 years of free-air CO2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO2. *New Phytol.*, *165*(2), 351–372.

Alberti, S. (2017). Phase separation in biology. *Current Biology: CB*, *27*(20), R1097–R1102.

Almagro Armenteros, J. J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A., & Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance*, *2*(5). https://doi.org/10.26508/lsa.201900429

Andersen, R. A. (2011). *Ochromonas moestrupii* sp. nov. (Chrysophyceae), a new golden flagellate from Australia. *Phycologia*, *50*(6), 600–607.

Andersson, I. (1996). Large structures at high resolution: the 1.6 A crystal structure of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase complexed with 2-carboxyarabinitol bisphosphate. *Journal of Molecular Biology*, *259*(1), 160–174.

Ang, W. S. L., How, J. A., How, J. B., & Mueller-Cajar, O. (2022). The stickers and spacers of Rubiscondensation: assembling the centrepiece of biophysical CO_2-concentrating mechanisms. *Journal of Experimental Botany*, erac321.

Antal, T. K., Kovalenko, I. B., Rubin, A. B., & Tyystjärvi, E. (2013). Photosynthesis-related quantities for education and modeling. *Photosynthesis Research*, *117*(1–3), 1–30.

Arakaki, Y., Kawai-Toyooka, H., Hamamura, Y., Higashiyama, T., Noga, A., Hirono, M., Olson, B. J. S. C., & Nozaki, H. (2013). The simplest integrated multicellular organism unveiled. *PloS One*, *8*(12), e81641.

Arriola, M. B., Velmurugan, N., Zhang, Y., Plunkett, M. H., Hondzo, H., & Barney, B. M. (2018). Genome sequences of *Chlorella sorokiniana* UTEX 1602 and *Micractinium conductrix* SAG 241.80: implications to maltose excretion by a green alga. *The Plant Journal: For Cell and Molecular Biology*, *93*(3), 566–586.

Aspar, A. H. (2013). *Studying the ultrastructure of the marine diatom Seminavis robusta* (A. Bones, Ed.) [Master of Science in Molecular Medicine, Norwegian University of Science and Technology]. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/263917

Atkinson, N., Feike, D., Mackinder, L. C. M., Meyer, M. T., Griffiths, H., Jonikas, M. C., Smith, A. M., & McCormick, A. J. (2016). Introducing an algal carbon-concentrating mechanism into higher plants: location and incorporation of key components. *Plant Biotechnology Journal*, *14*(5), 1302–1315.

Atkinson, N., Leitão, N., Orr, D. J., Meyer, M. T., Carmo-Silva, E., Griffiths, H., Smith, A. M., & McCormick, A. J. (2017). Rubisco small subunits from the unicellular green alga

Chlamydomonas complement Rubisco-deficient mutants of *Arabidopsis*. *The New Phytologist*, *214*(2), 655–667.
Atkinson, N., Mao, Y., Chan, K. X., & McCormick, A. J. (2020). Condensation of Rubisco into a proto-pyrenoid in higher plant chloroplasts. *Nature Communications*, *11*(1), 6303.
Atkinson, N., Velanis, C. N., Wunder, T., Clarke, D. J., Mueller-Cajar, O., & McCormick, A. J. (2019). The pyrenoidal linker protein EPYC1 phase separates with hybrid *Arabidopsis-Chlamydomonas* Rubisco through interactions with the algal Rubisco small subunit. *Journal of Experimental Botany*. https://doi.org/10.1093/jxb/erz275
Azaman, S. N. A., Nagao, N., Yusoff, F. M., Tan, S. W., & Yeap, S. K. (2017). A comparison of the morphological and biochemical characteristics of *Chlorella sorokiniana* and *Chlorella zofingiensis* cultured under photoautotrophic and mixotrophic conditions. *PeerJ*, *5*, e3473.
Badger, M. R., Hanson, D., & Price, G. D. (2002). Evolution and diversity of CO_2 concentrating mechanisms in cyanobacteria. *Functional Plant Biology: FPB*, *29*(3), 161–173.
Badger, M. R., Kaplan, A., & Berry, J. A. (1980). Internal Inorganic Carbon Pool of *Chlamydomonas reinhardtii*: evidence for a carbon dioxide-concentrating mechanism. *Plant Physiology*, *66*(3), 407–413.
Badger, M. R., Pfanz, H., Büdel, B., Heber, U., & Lange, O. (1993). Evidence for the functioning of photosynthetic CO_2-concentrating mechanisms in lichens containing green algal and cyanobacterial photobionts. *Planta*, *191*(1). https://doi.org/10.1007/bf00240896
Badger, M. R., & Price, G. D. (2003). CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *Journal of Experimental Botany*, *54*(383), 609–622.
Baier, T., Jacobebbinghaus, N., Einhaus, A., Lauersen, K. J., & Kruse, O. (2020). Introns mediate post-transcriptional enhancement of nuclear gene expression in the green microalga Chlamydomonas reinhardtii. *PLoS Genetics*, *16*(7), e1008944.
Baier, T., Wichmann, J., Kruse, O., & Lauersen, K. J. (2018). Intron-containing algal transgenes mediate efficient recombinant gene expression in the green microalga Chlamydomonas reinhardtii. *Nucleic Acids Research*, *46*(13), 6909–6919.
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 in Enzymology*, *398*, 540–554.
Banani, S. F., Lee, H. O., Hyman, A. A., & Rosen, M. K. (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews. Molecular Cell Biology*, *18*(5), 285–298.
Banani, S. F., Rice, A. M., Peeples, W. B., Lin, Y., Jain, S., Parker, R., & Rosen, M. K. (2016). Compositional Control of Phase-Separated Cellular Bodies. *Cell*, *166*(3), 651–663.
Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., & Milo, R. (2011). The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry*, *50*(21), 4402–4410.
Bar-On, Y. M., & Milo, R. (2019). The global mass and average rate of rubisco. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(10), 4738–4743.
Barrett, J., Girr, P., & Mackinder, L. C. M. (2021). Pyrenoids: CO2-fixing phase separated liquid organelles. *Biochimica et Biophysica Acta, Molecular Cell Research*, *1868*(5), 118949.
Bedoshvili, Y. D., Popkova, T. P., & Likhoshway, Y. V. (2009). Chloroplast structure of diatoms of different classes. *Cell and Tissue Biology*, *3*(3), 297–310.
Belkin, S., Mehlhorn, R. J., & Packer, L. (1987). Proton gradients in intact cyanobacteria. *Plant Physiology*, *84*(1), 25–30.
Benlloch, R., Shevela, D., Hainzl, T., Grundström, C., Shutova, T., Messinger, J., Samuelsson, G., & Sauer-Eriksson, A. E. (2015). Crystal structure and functional characterization of photosystem II-associated carbonic anhydrase CAH3 in Chlamydomonas reinhardtii. *Plant Physiology*, *167*(3), 950–962.
Benson, A. A., Bassham, J. A., Calvin, M., Goodale, T. C., Haas, V. A., & Stepka, W. (1950). The Path of Carbon in Photosynthesis. V. Paper Chromatography and Radioautography of the Products1. *Journal of the American Chemical Society*, *72*(4), 1710–1718.
Berner, R. A. (2006). GEOCARBSULF: A combined model for Phanerozoic atmospheric O2 and CO2. *Geochimica et Cosmochimica Acta*, *70*(23), 5653–5664.

Berner, R. A. (2009). Phanerozoic atmospheric oxygen: New results using the GEOCARBSULF model. *American Journal of Science*, *309*(7), 603–606.
Berner, R. A., & Kothavala, Z. (2001). Geocarb III: A Revised Model of Atmospheric CO2 over Phanerozoic Time. *American Journal of Science*, *301*(2), 182–204.
Bertagnolli, B. L., & Nadakavukaren, M. J. (1970). An ultrastructural study of pyrenoids from Chlorella pyrenoidosa. *Journal of Cell Science*, *7*(3), 623–630.
Bhat, J. Y., Miličić, G., Thieulin-Pardo, G., Bracher, A., Maxwell, A., Ciniawsky, S., Mueller-Cajar, O., Engen, J. R., Hartl, F. U., Wendler, P., & Hayer-Hartl, M. (2017). Mechanism of Enzyme Repair by the AAA+ Chaperone Rubisco Activase. *Molecular Cell*, *67*(5), 744-756.e6.
Biegert, A., & Söding, J. (2008). HHrepID: de novo protein repeat identification by probabilistic consistency. *Bioinformatics* , *24*, 807–814.
Blank, R. J., & Trench, R. K. (1988). Immunogold localization of Ribulose-1.5-bisphosphate carboxylase-oxygenase in Symbiodinium kawahutii trench et blank - An Endosymbiotic Dinoflagellate. *Endocytobiosis Cell Res.*, *5*(1), 75–82.
Block, M. A., Tewari, A. K., Albrieux, C., Maréchal, E., & Joyard, J. (2002). The plant S-adenosyl-L-methionine:Mg-protoporphyrin IX methyltransferase is located in both envelope and thylakoid chloroplast membranes. *European Journal of Biochemistry / FEBS*, *269*(1), 240–248.
Blomme, J., Liu, X., Jacobs, T. B., & De Clerck, O. (2020). A molecular toolkit for the green seaweed Ulva mutabilis. In *bioRxiv* (p. 2020.12.15.422947). https://doi.org/10.1101/2020.12.15.422947
Blomme, J., Liu, X., Jacobs, T. B., & De Clerck, O. (2021). A molecular toolkit for the green seaweed Ulva mutabilis. *Plant Physiology*, *186*(3), 1442–1454.
Blomme, J., Wichard, T., Jacobs, T. B., & De Clerck, O. (2023). Ulva: An emerging green seaweed model for systems biology. *Journal of Phycology*, *59*(3), 433–440.
Boeynaems, S., Alberti, S., Fawzi, N. L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L., Tompa, P., & Fuxreiter, M. (2018). Protein Phase Separation: A New Phase in Cell Biology. *Trends in Cell Biology*, *28*(6), 420–435.
Boeynaems, S., Chong, S., Gsponer, J., Holt, L., Milovanovic, D., Mitrea, D. M., Mueller-Cajar, O., Portz, B., Reilly, J. F., Reinkemeier, C. D., Sabari, B. R., Sanulli, S., Shorter, J., Sontag, E., Strader, L., Stachowiak, J., Weber, S. C., White, M., Zhang, H., Zweckstetter, M., Elbaum-Garfinkle, S., & Kriwacki, R. (2023). Phase Separation in Biology and Disease; Current Perspectives and Open Questions. *Journal of Molecular Biology*, *435*(5), 167971.
Bonaventura, C., & Myers, J. (1969). Fluorescence and oxygen evolution from Chlorella pyrenoidosa. *Biochimica et Biophysica Acta*, *189*(3), 366–383.
Borkhsenious, O. N., Mason, C. B., & Moroney, J. V. (1998). The intracellular localization of ribulose-1,5-bisphosphate Carboxylase/Oxygenase in chlamydomonas reinhardtii. *Plant Physiology*, *116*(4), 1585–1591.
Bouvier, J., Emms, D., & Kelly, S. (2023). Rubisco is evolving for improved catalytic efficiency and CO2 assimilation in plants. In *bioRxiv* (p. 2022.07.06.498985). https://doi.org/10.1101/2022.07.06.498985
Bouvier, J. W., Emms, D. M., Rhodes, T., Bolton, J. S., Brasnett, A., Eddershaw, A., Nielsen, J. R., Unitt, A., Whitney, S. M., & Kelly, S. (2021). Rubisco Adaptation Is More Limited by Phylogenetic Constraint Than by Catalytic Trade-off. *Molecular Biology and Evolution*, *38*(7), 2880–2896.
Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R. P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Verret, F., Berges, J. A., Brownlee, C., Cadoret, J.-P., Chiovitti, A., Choi, C. J., Coesel, S., De Martino, A., Detter, J. C., Durkin, C., Falciatore, A., Fournet, J., Haruta, M., Huysman, M. J. J., Jenkins, B. D., Jiroutova, K., Jorgensen, R. E., Joubert, Y., Kaplan, A., Kröger, N., Kroth, P. G., La Roche, J., Lindquist, E., Lommer, M., Martin-Jézéquel, V., Lopez, P. J., Lucas, S., Mangogna, M., McGinnis, K., Medlin, L. K., Montsant, A., Oudot-Le Secq, M.-P., Napoli, C., Obornik, M., Parker, M. S., Petit, J.-L., Porcel, B. M., Poulsen, N., Robison, M., Rychlewski, L., Rynearson, T. A., Schmutz, J., Shapiro, H., Siaut, M., Stanley, M., Sussman, M. R., Taylor, A. R., Vardi, A., von

Dassow, P., Vyverman, W., Willis, A., Wyrwicz, L. S., Rokhsar, D. S., Weissenbach, J., Armbrust, E. V., Green, B. R., Van de Peer, Y., & Grigoriev, I. V. (2008). The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature*, *456*(7219), 239–244.

Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., & Hyman, A. A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science*, *324*(5935), 1729–1732.

Brawley, S. H., Blouin, N. A., Ficko-Blean, E., Wheeler, G. L., Lohr, M., Goodson, H. V., Jenkins, J. W., Blaby-Haas, C. E., Helliwell, K. E., Chan, C. X., Marriage, T. N., Bhattacharya, D., Klein, A. S., Badis, Y., Brodie, J., Cao, Y., Collén, J., Dittami, S. M., Gachon, C. M. M., Green, B. R., Karpowicz, S. J., Kim, J. W., Kudahl, U. J., Lin, S., Michel, G., Mittag, M., Olson, B. J. S. C., Pangilinan, J. L., Peng, Y., Qiu, H., Shu, S., Singer, J. T., Smith, A. G., Sprecher, B. N., Wagner, V., Wang, W., Wang, Z.-Y., Yan, J., Yarish, C., Zäuner-Riek, S., Zhuang, Y., Zou, Y., Lindquist, E. A., Grimwood, J., Barry, K. W., Rokhsar, D. S., Schmutz, J., Stiller, J. W., Grossman, A. R., & Prochnik, S. E. (2017). Insights into the red algae and eukaryotic evolution from the genome of Porphyra umbilicalis (Bangiophyceae, Rhodophyta). *Proceedings of the National Academy of Sciences of the United States of America*, *114*(31), E6361–E6370.

Buchan, D. W. A., Minneci, F., Nugent, T. C. O., Bryson, K., & Jones, D. T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research*, *41*(Web Server issue), W349-57.

Buck, S., Rhodes, T., Gionfriddo, M., Skinner, T., Yuan, D., Birch, R., Kapralov, M. V., & Whitney, S. M. (2023). Escherichia coli expressing chloroplast chaperones as a proxy to test heterologous Rubisco production in leaves. *Journal of Experimental Botany*, *74*(2), 664–676.

Budd, T. W., Tjostem, J. L., & Duysen, M. E. (1969). Ultrastructure of Chlorella pyrenoidosa as affected by environmental changes. *American Journal of Botany*, *56*(5), 540–545.

Burki, F., Roger, A. J., Brown, M. W., & Simpson, A. G. B. (2020). The New Tree of Eukaryotes. *Trends Ecol. Evol.*, *35*(1), 43–55.

Busch, A., & Hess, S. (2022). A diverse group of underappreciated zygnematophytes deserves in-depth exploration. *Applied Phycology*, *3*(1), 306–323.

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* , *5*(2), 1141–1171.

Callaway, E. (2015). The revolution will not be crystallized: a new method sweeps through structural biology. *Nature*, *525*(7568), 172–174.

Callaway, E. (2020, February 10). *Revolutionary cryo-EM is taking over structural biology*. Nature Publishing Group UK. https://doi.org/10.1038/d41586-020-00341-9

Calvert, H. E., Dawes, C. J., & Borowitzka, M. A. (1976). Phlylogenetic relationships of Caulerpa (Chlorophyta) based on comparative ultrastructure. *J. Phycol.*, *12*(2), 149–162.

Cardon, Z. G., Peredo, E. L., Dohnalkova, A. C., Gershone, H. L., & Bezanilla, M. (2018). A model suite of green algae within the Scenedesmaceae for investigating contrasting desiccation tolerance and morphology. *Journal of Cell Science*, *131*(7). https://doi.org/10.1242/jcs.212233

Carmo-Silva, A. E., & Salvucci, M. E. (2013). The regulatory properties of Rubisco activase differ among species and affect photosynthetic induction during light transitions. *Plant Physiology*, *161*(4), 1645–1655.

Caspari, O. D. (2020). Introduction of a leaky stop codon as molecular tool in Chlamydomonas reinhardtii. *PloS One*, *15*(8), e0237405.

Catling, D. C., & Zahnle, K. J. (2020). The Archean atmosphere. *Science Advances*, *6*(9), eaax1420.

Cecchin, M., Marcolungo, L., Rossato, M., Girolomoni, L., Cosentino, E., Cuine, S., Li-Beisson, Y., Delledonne, M., & Ballottari, M. (2019). Chlorella vulgaris genome assembly and annotation reveals the molecular basis for metabolic acclimation to high light conditions. *The Plant Journal: For Cell and Molecular Biology*, *100*(6), 1289–1305.

Chan, K.-Y. (1974). Ultrastructure of Pyrenoid Division in *Coelastrum* sp. *Cytologia*, *39*(3), 531–536.

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., Yang, Q.-L., Xu, J.-X., Deng, X., Zhang, Y.-J., Liu, T., Rots, M. G., Xu, G.-L., & Huang, K.-Y. (2023). Efficient methods for multiple types of precise gene-editing in Chlamydomonas. *The Plant Journal: For Cell and Molecular Biology*, *115*(3), 846–865.

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 α-carboxysomes into plant chloroplasts to support autotrophic photosynthesis. *Nature Communications*, *14*(1), 2118.

Choi, J.-M., Holehouse, A. S., & Pappu, R. V. (2020). Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annual Review of Biophysics*, *49*, 107–133.

Choi, J.-M., & Pappu, R. V. (2020). The Stickers and Spacers Framework for Describing Phase Behavior of Multivalent Intrinsically Disordered Proteins. *Biophysical Journal*, *118*(3), 492a.

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & de Hoon, M. J. L. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* , *25*(11), 1422–1423.

Cortona, A. D., Jackson, C. J., Bucchini, F., Van Bel, M., D'hondt, S., Škaloud, P., Delwiche, C. F., Knoll, A. H., Raven, J. A., Verbruggen, H., Vandepoele, K., Clerck, O. D., & Leliaert, F. (2020). Neoproterozoic origin and multiple transitions to macroscopic growth in green seaweeds. *Proceedings of the National Academy of Sciences*, *117*(5), 2551–2559.

Cox, E. J. (1981). Observations on the morphology and vegetative cell division of the diatom Donkinia recta. *Helgoländer Meeresuntersuchungen*, *34*(4), 497–506.

Craig, R. J., Gallaher, S. D., Shu, S., Salomé, P. A., Jenkins, J. W., Blaby-Haas, C. E., Purvine, S. O., O'Donnell, S., Barry, K., Grimwood, J., Strenkert, D., Kropat, J., Daum, C., Yoshinaga, Y., Goodstein, D. M., Vallon, O., Schmutz, J., & Merchant, S. S. (2023). The Chlamydomonas Genome Project, version 6: Reference assemblies for mating-type plus and minus strains reveal extensive structural mutation in the laboratory. *The Plant Cell*, *35*(2), 644–672.

Craig, R. J., Hasan, A. R., Ness, R. W., & Keightley, P. D. (2021). Comparative genomics of Chlamydomonas. *The Plant Cell*, *33*(4), 1016–1041.

Crooks, G. E., Hon, G., Chandonia, J.-M., & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Research*, *14*(6), 1188–1190.

Crozet, P., Navarro, F. J., Willmund, F., Mehrshahi, P., Bakowski, K., Lauersen, K. J., Pérez-Pérez, M.-E., Auroy, P., Gorchs Rovira, A., Sauret-Gueto, S., Niemeyer, J., Spaniol, B., Theis, J., Trösch, R., Westrich, L.-D., Vavitsas, K., Baier, T., Hübner, W., de Carpentier, F., Cassarini, M., Danon, A., Henri, J., Marchand, C. H., de Mia, M., Sarkissian, K., Baulcombe, D. C., Peltier, G., Crespo, J.-L., Kruse, O., Jensen, P.-E., Schroda, M., Smith, A. G., & Lemaire, S. D. (2018). Birth of a Photosynthetic Chassis: A MoClo Toolkit Enabling Synthetic Biology in the Microalga Chlamydomonas reinhardtii. *ACS Synthetic Biology*, *7*(9), 2074–2086.

Cuevas-Velazquez, C. L., & Dinneny, J. R. (2018). Organization out of disorder: liquid-liquid phase separation in plants. *Current Opinion in Plant Biology*, *45*(Pt A), 68–74.

Dangeard, P.-A. (1888). Recherches sur les algues inférieures. *Ann Sci Nat Bot Ser VII*, *7*, 105–175.

Daugbjerg, N., Norlin, A., & Lovejoy, C. (2018). Baffinella frigidus gen. et sp. nov. (Baffinellaceae fam. nov., Cryptophyceae) from Baffin Bay: Morphology, pigment profile, phylogeny, and growth rate response to three abiotic factors. *Journal of Phycology*, *54*(5), 665–680.

Davidi, L., Gallaher, S. D., Ben-David, E., Purvine, S. O., Filmore, T. L., Nicora, C. D., Craig, R. J., Schmollinger, S., Roje, S., Blaby-Haas, C. E., Auber, R. P., Wisecaver, J., & Merchant, S. S. (2023). Pumping Iron: A Multi-omics Analysis of Two Extremophilic Algae Reveals Iron Economy Management. In *bioRxiv* (p. 2023.02.09.527888). https://doi.org/10.1101/2023.02.09.527888

De Clerck, O., Bogaert, K. A., & Leliaert, F. (2012). Chapter Two - Diversity and Evolution of Algae: Primary Endosymbiosis. In G. Piganeau (Ed.), *Advances in Botanical Research* (Vol. 64, pp. 55–86). Academic Press.

De Clerck, O., Kao, S.-M., Bogaert, K. A., Blomme, J., Foflonker, F., Kwantes, M., Vancaester, E., Vanderstraeten, L., Aydogdu, E., Boesger, J., Califano, G., Charrier, B., Clewes, R., Del

Cortona, A., D'Hondt, S., Fernandez-Pozo, N., Gachon, C. M., Hanikenne, M., Lattermann, L., Leliaert, F., Liu, X., Maggs, C. A., Popper, Z. A., Raven, J. A., Van Bel, M., Wilhelmsson, P. K. I., Bhattacharya, D., Coates, J. C., Rensing, S. A., Van Der Straeten, D., Vardi, A., Sterck, L., Vandepoele, K., Van de Peer, Y., Wichard, T., & Bothwell, J. H. (2018). Insights into the Evolution of Multicellularity from the Sea Lettuce Genome. *Current Biology: CB*, *28*(18), 2921-2933.e5.

De Martino, A., Bartual, A., Willis, A., Meichenin, A., Villazán, B., Maheswari, U., & Bowler, C. (2011). Physiological and molecular evidence that environmental changes elicit morphological interconversion in the model diatom Phaeodactylum tricornutum. *Protist*, *162*(3), 462–481.

Dean, C., Pichersky, E., & Dunsmuir, P. (1989). Structure, Evolution, and Regulation of RbcS Genes in Higher Plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, *40*(1), 415–439.

Deane, J., Hill, D., Brett, S., & McFadden, G. (1998). Hanusia phi gen. et sp. nov. (Cryptophyceae): characterization of 'Cryptomonas sp. Φ.' *Eur. J. Phycol.*, *33*(2), 149–154.

Decelle, J., Stryhanyuk, H., Gallet, B., Veronesi, G., Schmidt, M., Balzano, S., Marro, S., Uwizeye, C., Jouneau, P.-H., Lupette, J., Jouhet, J., Maréchal, E., Schwab, Y., Schieber, N. L., Tucoulou, R., Richnow, H., Finazzi, G., & Musat, N. (2019). Algal Remodeling in a Ubiquitous Planktonic Photosymbiosis. *Current Biology: CB*, *29*(6), 968-978.e4.

del Pino Plumed, M., Villarejo, A., de los Róos, A., García-Reina, G., & Ramazanov, Z. (1996). The CO 2-concentrating mechanism in a starchless mutant of the green unicellular alga Chlorella pyrenoidosa. *Planta*, *200*(1), 28–31.

Demchenko, E., Mikhailyuk, T., Coleman, A. W., & Pröschold, T. (2012). Generic and species concepts in Microglena (previously the Chlamydomonas monadina group) revised using an integrative approach. *European Journal of Phycology*, *47*(3), 264–290.

Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S., & Ralser, M. (2020). DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods*, *17*(1), 41–44.

Deschamps, P., Haferkamp, I., Dauvillée, D., Haebel, S., Steup, M., Buléon, A., Putaux, J.-L., Colleoni, C., d'Hulst, C., Plancke, C., Gould, S., Maier, U., Neuhaus, H. E., & Ball, S. (2006). Nature of the periplastidial pathway of starch synthesis in the cryptophyte Guillardia theta. *Eukaryotic Cell*, *5*(6), 954–963.

Dignon, G. L., Best, R. B., & Mittal, J. (2020). Biomolecular Phase Separation: From Molecular Driving Forces to Macroscopic Properties. *Annual Review of Physical Chemistry*, *71*, 53–75.

Dodge, J. D. (1968). The fine structure of chloroplasts and pyrenoids in some marine dinoflagellates. *Journal of Cell Science*, *3*(1), 41–48.

dos Reis, M., Donoghue, P. C. J., & Yang, Z. (2016). Bayesian molecular clock dating of species divergences in the genomics era. *Nature Reviews. Genetics*, *17*(2), 71–80.

Dosztányi, Z., Csizmok, V., Tompa, P., & Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* , *21*(16), 3433–3434.

Drum, R. W. (1963). THE CYTOPLASMIC FINE STRUCTURE OF THE DIATOM, NITZSCHIA PALEA. *The Journal of Cell Biology*, *18*(2), 429–440.

Duanmu, D., Wang, Y., & Spalding, M. H. (2009). Thylakoid lumen carbonic anhydrase (CAH3) mutation suppresses air-Dier phenotype of LCIB mutant in Chlamydomonas reinhardtii. *Plant Physiology*, *149*(2), 929–937.

Duff, A. P., Andrews, T. J., & Curmi, P. M. (2000). The transition between the open and closed states of rubisco is triggered by the inter-phosphate distance of the bound bisphosphate. *Journal of Molecular Biology*, *298*(5), 903–916.

Economou, C., Wannathong, T., Szaub, J., & Purton, S. (2014). A Simple, Low-Cost Method for Chloroplast Transformation of the Green Alga Chlamydomonas reinhardtii. In P. Maliga (Ed.), *Chloroplast Biotechnology: Methods and Protocols* (pp. 401–411). Humana Press.

Edwards, E. J., & Ogburn, R. M. (2012). Angiosperm Responses to a Low-CO2 World: CAM and C4 Photosynthesis as Parallel Evolutionary Trajectories. *International Journal of Plant Sciences*, *173*(6), 724–733.

Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C. C.-H., Eckmann, C. R., Myong, S., & Brangwynne, C. P. (2015). The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(23), 7189–7194.

Ellis, R. J. (1979). The most abundant protein in the world. *Trends in Biochemical Sciences*, *4*(11), 241–244.

Emrich-Mills, T. Z., Yates, G., Barrett, J., Girr, P., Grouneva, I., Lau, C. S., Walker, C. E., Kwok, T. K., Davey, J. W., Johnson, M. P., & Mackinder, L. C. M. (2021). A recombineering pipeline to clone large and complex genes in Chlamydomonas. *The Plant Cell*, *33*(4), 1161–1181.

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 4), 486–501.

Engel, B. D., Schaffer, M., Cuellar, L. K., Villa, E., Plitzko, J. M., & Baumeister, W. (2015). Correction: Native architecture of the Chlamydomonas chloroplast revealed by in situ cryo-electron tomography. *ELife*, *4*. https://doi.org/10.7554/eLife.11383

Engler, C., Youles, M., Gruetzner, R., Ehnert, T.-M., Werner, S., Jones, J. D. G., Patron, N. J., & Marillonnet, S. (2014). A golden gate modular cloning toolbox for plants. *ACS Synthetic Biology*, *3*(11), 839–843.

Evans, L. V. (1966). Distribution of pyrenoids among some brown algae. *Journal of Cell Science*, *1*(4), 449–454.

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 α-carboxysome. *Structure* , *31*(6), 677-688.e4.

Fan, J., Ning, K., Zeng, X., Luo, Y., Wang, D., Hu, J., Li, J., Xu, H., Huang, J., Wan, M., Wang, W., Zhang, D., Shen, G., Run, C., Liao, J., Fang, L., Huang, S., Jing, X., Su, X., Wang, A., Bai, L., Hu, Z., Xu, J., & Li, Y. (2015). Genomic Foundation of Starch-to-Lipid Switch in Oleaginous Chlorella spp. *Plant Physiology*, *169*(4), 2444–2461.

Fan, J., Xu, H., Luo, Y., Wan, M., Huang, J., Wang, W., & Li, Y. (2015). Impacts of CO2 concentration on growth, lipid accumulation, and carbon-concentrating-mechanism-related gene expression in oleaginous Chlorella. *Applied Microbiology and Biotechnology*, *99*(5), 2451–2462.

Fawley, M. W., Yun, Y., & Qin, M. (2001). Phylogenetic analyses of 18s rdna sequences reveal a new coccoid lineage of the prasinophyceae (Chlorophyta). *Journal of Phycology*, *36*(2), 387–393.

Fei, C., Wilson, A. T., Mangan, N. M., Wingreen, N. S., & Jonikas, M. C. (2022). Modelling the pyrenoid-based CO2-concentrating mechanism provides insights into its operating principles and a roadmap for its engineering into crops. *Nature Plants*, *8*(5), 583–595.

Feng, X., Holzinger, A., Permann, C., Anderson, D., & Yin, Y. (2021). Characterization of Two Zygnema Strains (Zygnema circumcarinatum SAG 698-1a and SAG 698-1b) and a Rapid Method to Estimate Nuclear Genome Size of Zygnematophycean Green Algae. *Frontiers in Plant Science*, *12*, 610381.

Feric, M., & Brangwynne, C. P. (2013). A nuclear F-actin scaffold stabilizes ribonucleoprotein droplets against gravity in large cells. *Nature Cell Biology*, *15*(10), 1253–1259.

Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*, *281*(5374), 237–240.

Flamholz, A. I., Prywes, N., Moran, U., Davidi, D., Bar-On, Y. M., Oltrogge, L. M., Alves, R., Savage, D., & Milo, R. (2019). Revisiting Trade-offs between Rubisco Kinetic Parameters. *Biochemistry*, *58*(31), 3365–3376.

Flamholz, A., & Shih, P. M. (2020). Cell biology of photosynthesis over geologic time. *Current Biology: CB*, *30*(10), R490–R494.

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.e20.

Freeman Rosenzweig, E. S., Xu, B., Kuhn Cuellar, L., Martinez-Sanchez, A., Schaffer, M., Strauss, M., Cartwright, H. N., Ronceray, P., Plitzko, J. M., Förster, F., Wingreen, N. S., Engel, B. D.,

Mackinder, L. C. M., & Jonikas, M. C. (2017). The Eukaryotic CO2-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization. *Cell*, *171*(1), 148-162.e19.
Fresnel, J., & Probert, I. (2005). The ultrastructure and life cycle of the coastal coccolithophorid Ochrosphaera neapolitana (Prymnesiophyceae). *Eur. J. Phycol.*, *40*(1), 105–122.
Friedl, T. (1989). Comparative ultrastructure of pyrenoids inTrebouxia (Microthamniales, Chlorophyta). *Plant Systematics and Evolution = Entwicklungsgeschichte Und Systematik Der Pflanzen*, *164*(1–4), 145–159.
Gaffal, K. P., Arnold, C.-G., Friedrichs, G. J., & Gemple, W. (1995). Morphodynamical Changes of the Chloroplast of Chlamydomonas reinhardtii During the 1st Round of Division. *Archiv Für Protistenkunde*, *145*(1), 10–23.
Galmés, J., Kapralov, M. V., Andralojc, P. J., Conesa, M. À., Keys, A. J., Parry, M. A. J., & Flexas, J. (2014). Expanding knowledge of the Rubisco kinetics variability in plant species: environmental and evolutionary trends. *Plant, Cell & Environment*, *37*(9), 1989–2001.
Gantt, E., & Conti, S. F. (1965). The ultrastructure of Porphyridium cruentum. *The Journal of Cell Biology*, *26*(2), 365–381.
Gao, C., Wang, Y., Shen, Y., Yan, D., He, X., Dai, J., & Wu, Q. (2014). Oil accumulation mechanisms of the oleaginous microalga Chlorella protothecoides revealed through its genome, transcriptomes, and proteomes. *BMC Genomics*, *15*(1), 582.
Gartner, G., & Ingolic, E. (1989). Some ultrastructural aspects of the pyrenoids in Chlorokybus atmophyticus GEITLER (Charophyceae, Chlorokybales). *Phyton; Annales Rei Botanicae*.
Gärtner, G., Stoyneva, M. P., Uzunov, B. A., Mancheva, A. D., & Ingolić, E. (2012). Ultrastructure of vegetative cells and autospores of an aerophytic strain of Vischeria stellata (Chodat ex Poulton) Pascher (Eustigmatophyceae) from Bulgaria. *Fottea* , *12*(2), 273–280.
Gärtner, G., Uzunov, B., Ingolic, E., Kofler, W., Gacheva, G., Pilarski, P., Zagorchev, L., Odjakova, M., & Stoyneva, M. (2015). Microscopic investigations (LM, TEM and SEM) and identification of Chlorella isolate R-06/2 from extreme habitat in Bulgaria with a strong biological activity and resistance to environmental stress factors. *Biotechnol. Biotechnol. Equip.*, *29*(3), 536–540.
Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the Expasy Server; In J. M. Walker (Ed.), *The Proteomics Protocols Handbook* (pp. 561–607). Humana Press.
Genkov, T., Meyer, M., Griffiths, H., & Spreitzer, R. J. (2010). … Hybrid Rubisco Enzymes with Plant Small Subunits and Algal Large Subunits: ENGINEERED rbcS cDNA FOR EXPRESSION IN CHLAMYDOMONAS*♦. *Journal of Biological*. https://www.jbc.org/article/S0021-9258(18)49658-5/abstract
Gergis, M. S. (1972). Influence of carbon dioxide supply on the chloroplast structure of Chlorella pyrenoidosa. *Archiv Fur Mikrobiologie*, *83*(4), 321–327.
Gibbs, S. P. (1962). The ultrastructure of the pyrenoids of green algae. *J. Ultrastruct. Res.*, *7*(3–4), 262–272.
Giustiniani, A., Drenckhan, W., & Poulard, C. (2017). Interfacial tension of reactive, liquid interfaces and its consequences. *Advances in Colloid and Interface Science*, *247*, 185–197.
Goodenough, U. W. (1970). CHLOROPLAST DIVISION AND PYRENOID FORMATION IN CHLAMYDOMONAS REINHARDI. *Journal of Phycology*, *6*(1), 1–6.
Goodenough, U. W., & Levine, R. (1970). Chloroplast structure and function in ac-20, a mutant strain of Chlamydomonas reinhardi. 3. Chloroplast ribosomes and membrane organization. *J. Cell Biol.*, *44*(3), 547–562.
Goodin, M. M., Zaitlin, D., Naidu, R. A., & Lommel, S. A. (2008). Nicotiana benthamiana: its history and future as a model for plant-pathogen interactions. *Molecular Plant-Microbe Interactions: MPMI*, *21*(8), 1015–1026.
Gorelova, O., Baulina, O., Ismagulova, T., Kokabi, K., Lobakova, E., Selyakh, I., Semenova, L., Chivkunova, O., Karpova, O., Scherbakov, P., Khozin-Goldberg, I., & Solovchenko, A. (2019). Stress-induced changes in the ultrastructure of the photosynthetic apparatus of green microalgae. *Protoplasma*, *256*(1), 261–277.
Goudet, M. M. M., Orr, D. J., Melkonian, M., Müller, K. H., Meyer, M. T., Carmo-Silva, E., & Griffiths, H. (2020). Rubisco and carbon concentrating mechanism (CCM) co-evolution

across Chlorophyte and Streptophyte green algae. *The New Phytologist*. https://doi.org/10.1111/nph.16577
GrandPre, T., Zhang, Y., Pyo, A. G. T., Weiner, B., Li, J.-L., Jonikas, M. C., & Wingreen, N. S. (2023). Effects of linker length on phase separation: lessons from the Rubisco-EPYC1 system of the algal pyrenoid. In *bioRxiv* (p. 2023.06.11.544494). https://doi.org/10.1101/2023.06.11.544494
Griffiths, D. J. (1970). The pyrenoid. *The Botanical Review; Interpreting Botanical Progress*, *36*(1), 29–58.
Griffiths, H., Meyer, M. T., & Rickaby, R. E. M. (2017). Overcoming adversity through diversity: aquatic carbon concentrating mechanisms. *Journal of Experimental Botany*, *68*(14), 3689–3695.
Guillou, L., Eikrem, W., Chrétiennot-Dinet, M.-J., Le Gall, F., Massana, R., Romari, K., Pedrós-Alió, C., & Vaulot, D. (2004). Diversity of picoplanktonic prasinophytes assessed by direct nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. *Protist*, *155*(2), 193–214.
Guiry, M. D. (2012). How Many Species of Algae Are There? *J. Phycol.*, *48*(5), 1057–1063.
Guo, W., Feng, L., Wang, Z., Guo, J., Park, D., Carroll, B. L., Zhang, X., Liu, J., & Cheng, J. (2022). In-situ high-resolution 3D imaging combined with proteomics and metabolomics reveals enlargement of subcellular architecture and enhancement of photosynthesis pathways in nuclear-irradiated Chlorella pyrenoidosa. *Chemical Engineering Journal* , *430*, 133037.
Hammel, A., Zimmer, D., Sommer, F., Mühlhaus, T., & Schroda, M. (2018). Absolute Quantification of Major Photosynthetic Protein Complexes in Chlamydomonas reinhardtii Using Quantification Concatamers (QconCATs). *Frontiers in Plant Science*, *9*, 1265.
Han, K. Y., Graf, L., Reyes, C. P., Melkonian, B., Andersen, R. A., Yoon, H. S., & Melkonian, M. (2018). A Re-investigation of Sarcinochrysis marina (Sarcinochrysidales, Pelagophyceae) from its Type Locality and the Descriptions of Arachnochrysis, Pelagospilus, Sargassococcus and Sungminbooa genera nov. *Protist*, *169*(1), 79–106.
Hanson, M. R., Lin, M. T., Carmo-Silva, A. E., & Parry, M. A. J. (2016). Towards engineering carboxysomes into C3 plants. *The Plant Journal: For Cell and Molecular Biology*, *87*(1), 38–50.
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*. https://doi.org/10.7554/eLife.30294
Harris, E. H. (2001). CHLAMYDOMONAS AS A MODEL ORGANISM. *Annual Review of Plant Physiology and Plant Molecular Biology*, *52*, 363–406.
Hatch, M. D. (1987). C4 photosynthesis: a unique elend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics*, *895*(2), 81–106.
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. *Nature Plants*, 1–11.
He, S., Crans, V. L., & Jonikas, M. C. (2023). The pyrenoid: the eukaryotic CO2-concentrating organelle. *The Plant Cell*. https://doi.org/10.1093/plcell/koad157
Heger, A., & Holm, L. (2000). Rapid automatic detection and alignment of repeats in protein sequences. *Proteins*, *41*(2), 224–237.
Heldt, W. H., Werdan, K., Milovancev, M., & Geller, G. (1973). Alkalization of the chloroplast stroma caused by light-dependent proton flux into the thylakoid space. *Biochimica et Biophysica Acta*, *314*(2), 224–241.
Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E., & Downing, K. H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *Journal of Molecular Biology*, *213*(4), 899–929.
Henderson, R., Sali, A., Baker, M. L., Carragher, B., Devkota, B., Downing, K. H., Egelman, E. H., Feng, Z., Frank, J., Grigorieff, N., Jiang, W., Ludtke, S. J., Medalia, O., Penczek, P. A., Rosenthal, P. B., Rossmann, M. G., Schmid, M. F., Schröder, G. F., Steven, A. C., Stokes, D.

L., Westbrook, J. D., Wriggers, W., Yang, H., Young, J., Berman, H. M., Chiu, W., Kleywegt, G. J., & Lawson, C. L. (2012). Outcome of the first electron microscopy validation task force meeting. *Structure* , *20*(2), 205–214.

Hennacy, J. H., & Jonikas, M. C. (2020). Prospects for Engineering Biophysical CO2 Concentrating Mechanisms into Land Plants to Enhance Yields. *Annual Review of Plant Biology*. https://doi.org/10.1146/annurev-arplant-081519-040100

Herron, M. D., Hackett, J. D., Aylward, F. O., & Michod, R. E. (2009). Triassic origin and early radiation of multicellular volvocine algae. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(9), 3254–3258.

Hertweck, K. L., Kinney, M. S., Stuart, S. A., Maurin, O., Mathews, S., Chase, M. W., Gandolfo, M. A., & Pires, J. C. (2015). Phylogenetics, divergence times and diversification from three genomic partitions in monocots. *Botanical Journal of the Linnean Society. Linnean Society of London*, *178*(3), 375–393.

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. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(33). https://doi.org/10.1073/pnas.2107425118

Hirakawa, Y., & Ishida, K.-I. (2015). Prospective function of FtsZ proteins in the secondary plastid of chlorarachniophyte algae. *BMC Plant Biology*, *15*, 276.

Hirooka, S., Hirose, Y., Kanesaki, Y., Higuchi, S., Fujiwara, T., Onuma, R., Era, A., Ohbayashi, R., Uzuka, A., Nozaki, H., Yoshikawa, H., & Miyagishima, S.-Y. (2017). Acidophilic green algal genome provides insights into adaptation to an acidic environment. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(39), E8304–E8313.

Ho, S. Y. W., & Duchêne, S. (2014). Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology*, *23*(24), 5947–5965.

Hoffman, L. R. (1968a). Observations on the Fine Structure of Oedogonium IV. The Mature Pyrenoid of Oe. cardiacum. *Trans. Am. Microsc. Soc.*, *87*(2), 178–185.

Hoffman, L. R. (1968b). Observations on the fine structure of Oedogonium. V. Evidence for the de novo Formation of Pyrenoids in Zoospores of OE. cardiacum. *J. Phycol.*, *4*(3), 212–218.

Holdsworth, R. H. (1971). The isolation and partial characterization of the pyrenoid protein of Eremosphaera viridis. *J. Cell Biol.*, *51*(2), 499–513.

Hoops, H. J., & Floyd, G. L. (1979). Ultrastructure of the centric diatom, Cyclotella meneghiniana: vegetative cell and auxospore development. *Phycologia*, *18*(4), 424–435.

Hori, T., & Inouye, I. (1981). The ultrastructure of mitosis in Cricosphaera roscoffensis var. haptonemofera (Prymnesiophyceae). *Protoplasma*, *106*(1–2), 121–135.

Hori, T., Norris, R. E., & Chihara, M. (1986). Studies on the ultrastructure and taxonomy of the genusTetraselmis (Prasinophyceae). *Botanical Magazine Tokyo*, *99*(1), 123–135.

Hoshina, R., Iwataki, M., & Imamura, N. (2010). Chlorella variabilis and Micractinium reisseri sp. nov. (Chlorellaceae, Trebouxiophyceae): Redescription of the endosymbiotic green algae of Paramecium bursaria (Peniculia, Oligohymenophorea) in the 120th year. *Phycological Research*, *58*(3), 188–201.

Hovde, B. T., Hanschen, E. R., Steadman Tyler, C. R., Lo, C.-C., Kunde, Y., Davenport, K., Daligault, H., Msanne, J., Canny, S., Eyun, S.-I., Riethoven, J.-J. M., Polle, J., & Starkenburg, S. R. (2018). Genomic characterization reveals significant divergence within Chlorella sorokiniana (Chlorellales, Trebouxiophyceae). *Algal Research*, *35*, 449–461.

Huang, F., Kong, W.-W., Sun, Y., Chen, T., Dykes, G. F., Jiang, Y.-L., & Liu, L.-N. (2020). Rubisco accumulation factor 1 (Raf1) plays essential roles in mediating Rubisco assembly and carboxysome biogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(29), 17418–17428.

Huang, X., Zhu, S., Zhang, H., Huang, Y., Wang, X., Wang, Y., & Chen, D. (2021). Biochar Nanoparticles Induced Distinct Biological Effects on Freshwater Algae via Oxidative Stress, Membrane Damage, and Nutrient Depletion. *ACS Sustainable Chemistry & Engineering*, *9*(32), 10761–10770.

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

Ifuku, K., Ishihara, S., Shimamoto, R., Ido, K., & Sato, F. (2008). Structure, function, and evolution of the PsbP protein family in higher plants. *Photosynthesis Research*, *98*(1–3), 427–437.

Ikeda, T., & Takeda, H. (1995). SPECIES-SPECIFIC DIFFERENCES OF PYRENOIDS IN CHLORELLA (CHLOROPHYTA)1. *Journal of Phycology*, *31*(5), 813–818.

Invernizzi, C., Imhof, J., Burkard, G., Schmid, K., & Boschetti, A. (2002). Effects of mutations at the two processing sites of the precursor for the small subunit of ribulose-bisphosphate carboxylase in Chlamydomonas reinhardtii. *Biochemical Journal*, *366*(Pt 3), 989–998.

Ishida, K.-I., Ishida, N., & Hara, Y. (2000). Lotharella amoeboformis sp. nov.: A new species of chlorarachniophytes from Japan. *Phycological Research*, *48*(4), 221–229.

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. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(37), 18445–18454.

Jackson, C., Knoll, A. H., Chan, C. X., & Verbruggen, H. (2018). Plastid phylogenomics with broad taxon sampling further elucidates the distinct evolutionary origins and timing of secondary green plastids. *Scientific Reports*, *8*(1), 1523.

Jackson, H. O., Taunt, H. N., Mordaka, P. M., Smith, A. G., & Purton, S. (2021). The Algal Chloroplast as a Testbed for Synthetic Biology Designs Aimed at Radically Rewiring Plant Metabolism. *Frontiers in Plant Science*, *12*, 708370.

Jenks, A., & Gibbs, S. P. (2000). Immunolocalization and distribution of form ii rubisco in the pyrenoid and chloroplast stroma of amphidinium carterae and form i rubisco in the symbiont-derived plastids of peridinium foliaceum (dinophyceae). *Journal of Phycology*, *36*(1), 127–138.

Jin, S., Sun, J., Wunder, T., Tang, D., Cousins, A. B., Sze, S. K., Mueller-Cajar, O., & Gao, Y.-G. (2016). Structural insights into the LCIB protein family reveals a new group of β-carbonic anhydrases. *Proc. Natl. Acad. Sci. U. S. A.*, *113*(51), 14716–14721.

Jouenne, F., Eikrem, W., Le Gall, F., Marie, D., Johnsen, G., & Vaulot, D. (2011). Prasinoderma singularis sp. nov. (Prasinophyceae, Chlorophyta), a solitary coccoid Prasinophyte from the South-East Pacific Ocean. *Protist*, *162*(1), 70–84.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589.

Kadereit, G., Borsch, T., Weising, K., & Freitag, H. (2003). Phylogeny of Amaranthaceae and Chenopodiaceae and the Evolution of C4 Photosynthesis. *International Journal of Plant Sciences*, *164*(6), 959–986.

Kajikawa, H., Okada, M., Ishikawa, F., Okada, Y., & Nakayama, K. (1988). Immunochemical Studies on Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase in the Chloroplasts of the Marine Alga Bryopsis maxima. *Plant Cell Physiol.*, *29*(4), 549–556.

Kanai, R., & Edwards, G. E. (1999). The biochemistry of C4 photosynthesis. *C4 Plant Biology*, *49*, 87.

Kanevski, I., Maliga, P., Rhoades, D. F., & Gutteridge, S. (1999). Plastome engineering of ribulose-1,5-bisphosphate carboxylase/oxygenase in tobacco to form a sunflower large subunit and tobacco small subunit hybrid. *Plant Physiology*, *119*(1), 133–142.

Kapust, R. B., Tözsér, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D., & Waugh, D. S. (2001). Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Engineering*, *14*(12), 993–1000.

Karcher, D., Köster, D., Schadach, A., Klevesath, A., & Bock, R. (2009). The Chlamydomonas chloroplast HLP protein is required for nucleoid organization and genome maintenance. *Molecular Plant*, *2*(6), 1223–1232.

Karlsson, J., Clarke, A. K., Chen, Z. Y., Hugghins, S. Y., Park, Y. I., Husic, H. D., Moroney, J. V., & Samuelsson, G. (1998). A novel alpha-type carbonic anhydrase associated with the thylakoid membrane in Chlamydomonas reinhardtii is required for growth at ambient CO2. *EMBO J.*, *17*(5), 1208–1216.

Kasting, J. F. (2013). What caused the rise of atmospheric O2? *Chemical Geology*, *362*, 13–25.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.

Katsaros, C., Weiss, A., Llangos, I., Theodorou, I., & Wichard, T. (2017). Cell structure and microtubule organisation during gametogenesis of Ulva mutabilis Føyn (Chlorophyta). *Botanica Marina*, *60*(2), 123–135.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*(6), 845–858.

Kellogg, E. A. (2001). Evolutionary history of the grasses. *Plant Physiology*, *125*(3), 1198–1205.

Kerby, N. W., & Evans, L. V. (1978). Isolation and partial characterization of pyrenoids from the brown alga Pilayella littoralis (L.) Kjellm. *Planta*, *142*(1), 91–95.

Kerfeld, C. A., & Melnicki, M. R. (2016). Assembly, function and evolution of cyanobacterial carboxysomes. *Current Opinion in Plant Biology*, *31*, 66–75.

Kevin, M. J., Hall, W. T., McLaughlin, J. J., & Zahl, P. A. (1969). SYMBIODINIUM MICROADRIATICUM FREUDENTHAL, A REVISED TAXONOMIC DESCRIPTION, ULTRASTRUCTURE. *Journal of Phycology*, *5*(4), 341–350.

Kikutani, S., Nakajima, K., Nagasato, C., Tsuji, Y., Miyatake, A., & Matsuda, Y. (2016). Thylakoid luminal θ-carbonic anhydrase critical for growth and photosynthesis in the marine diatom Phaeodactylum tricornutum. *Proc. Natl. Acad. Sci. U. S. A.*, *113*(35), 9828–9833.

Kim, G. H., Klochkova, T. A., Lee, D. J., & Im, S. H. (2016). Chloroplast virus causes green-spot disease in cultivated Pyropia of Korea. *Algal Research*, *17*, 293–299.

Kimura, K., & Tomaru, Y. (2015). Discovery of two novel viruses expands the diversity of single-stranded DNA and single-stranded RNA viruses infecting a cosmopolitan marine diatom. *Applied and Environmental Microbiology*, *81*(3), 1120–1131.

King, M. R., & Petry, S. (2020). Phase separation of TPX2 enhances and spatially coordinates microtubule nucleation. *Nature Communications*, *11*(1), 270.

Kiss, J. Z., Vasoconcelos, A. C., & Triemer, R. E. (1986). Paramylon Synthesis and Chloroplast Structure Associated with Nutrient Levels in Euglena (Euglenophyceae). *J. Phycol.*, *22*(3), 327–333.

Klöpper, S., John, U., Zingone, A., Mangoni, O., Kooistra, W. H. C. F., & Cembella, A. D. (2013). Phylogeny and morphology of a Chattonella (Raphidophyceae) species from the Mediterranean Sea: what is C. subsalsa? *Eur. J. Phycol.*, *48*(1), 79–92.

Kochert, G., & Olson, L. W. (1970). Ultrastructure of Volvox carteri. *Archives of Microbiology*, *74*(1), 19–30.

Kozlowski, L. P., & Bujnicki, J. M. (2012). MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics*, *13*, 111.

Krayesky-Self, S., Phung, D., Schmidt, W., Sauvage, T., Butler, L., & Fredericq, S. (2020). First report of endolithic members of Rhodosorus marinus (stylonematales, Rhodophyta) growing inside rhodoliths offshore Louisiana, northwestern gulf of Mexico. *Frontiers in Marine Science*, *7*. https://doi.org/10.3389/fmars.2020.00007

Krienitz, L., Huss, V. A. R., & Bock, C. (2015). Chlorella: 125 years of the green survivalist. *Trends in Plant Science*, *20*(2), 67–69.

Krienitz, L., Ustinova, I., Friedl, T., & Huss, V. A. R. (2001). TRADITIONAL GENERIC CONCEPTS VERSUS 18S rRNA GENE PHYLOGENY IN THE GREEN ALGAL FAMILY SELENASTRACEAE (CHLOROPHYCEAE, CHLOROPHYTA). *Journal of Phycology*, *37*(5), 852–865.

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580.

Kromdijk, J., Głowacka, K., Leonelli, L., Gabilly, S. T., Iwai, M., Niyogi, K. K., & Long, S. P. (2016). Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Science*, *354*(6314), 857.

Kropat, J., Hong-Hermesdorf, A., Casero, D., Ent, P., Castruita, M., Pellegrini, M., Merchant, S. S., & Malasarn, D. (2011). A revised mineral nutrient supplement increases biomass and growth rate in Chlamydomonas reinhardtii. *The Plant Journal: For Cell and Molecular Biology*, *66*(5), 770–780.

Ku, M. S., Kano-Murakami, Y., & Matsuoka, M. (1996). Evolution and expression of C4 photosynthesis genes. *Plant Physiology*, *111*(4), 949–957.

Kuchitsu, K., Tsuzuki, M., & Miyachi, S. (1988). Changes of Starch Localization within the Chloroplast Induced by Changes in CO2 Concentration during Growth of Chlamydomonas reinhardtii: Independent Regulation of Pyrenoid Starch and Stroma Starch. *Plant Cell Physiol.*, *29*(8), 1269–1278.

Kuchitsu, K., Tsuzuki, M., & Miyachi, S. (1991). Polypeptide composition and enzyme activities of the pyrenoid and its regulation by CO2 concentration in unicellular green algae. *Can. J. Bot.*, *69*(5), 1062–1069.

Küken, A., Sommer, F., Yaneva-Roder, L., Mackinder, L. C., Höhne, M., Geimer, S., Jonikas, M. C., Schroda, M., Stitt, M., Nikoloski, Z., & Mettler-Altmann, T. (2018). Effects of microcompartmentation on flux distribution and metabolic pools in Chlamydomonas reinhardtii chloroplasts. *ELife*, *7*. https://doi.org/10.7554/eLife.37960

Kumar, M., Jeon, J., Choi, J., & Kim, S.-R. (2018). Rapid and efficient genetic transformation of the green microalga Chlorella vulgaris. *Journal of Applied Phycology*, *30*(3), 1735–1745.

Lacoste-Royal, G., & Gibbs, S. P. (1987). Immunocytochemical Localization of Ribulose-1,5-Bisphosphate Carboxylase in the Pyrenoid and Thylakoid Region of the Chloroplast of Chlamydomonas reinhardtii. *Plant Physiology*, *83*(3), 602–606.

Lacoste-Royal, Ginette, & Gibbs, S. P. (1987). Immunocytochemical localization of ribulose-1, 5-bisphosphate carboxylase in the pyrenoid and thylakoid region of the chloroplast of Chlamydomonas reinhardtii. *Plant Physiology*, *83*(3), 602–606.

Lander, C. A. (1935). The Relation of the Plastid to Nuclear Division in Anthoceros laevis. *American Journal of Botany*, *22*(1), 42–51.

Langley, C. H., & Fitch, W. M. (1974). An examination of the constancy of the rate of molecular evolution. *Journal of Molecular Evolution*, *3*(3), 161–177.

Larson, E. M., O'Brien, C. M., Zhu, G., Spreitzer, R. J., & Portis, A. R., Jr. (1997). Specificity for activase is changed by a Pro-89 to Arg substitution in the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. *The Journal of Biological Chemistry*, *272*(27), 17033–17037.

Lau, C. S., Dowle, A., Thomas, G. H., Girr, P., & Mackinder, L. C. M. (2023). A phase-separated CO2-fixing pyrenoid proteome determined by TurboID in Chlamydomonas reinhardtii. *The Plant Cell*. https://doi.org/10.1093/plcell/koad131

Lee, S. Y., Jeong, H. J., Kang, N. S., Jang, T. Y., Jang, S. H., & Lajeunesse, T. C. (2015). Symbiodinium tridacnidorum sp. nov., a dinoflagellate common to Indo-Pacific giant clams, and a revised morphological description of Symbiodinium microadriaticum Freudenthal, emended Trench & Blank. *European Journal of Phycology*, *50*(2), 155–172.

Leliaert, F., Smith, D. R., Moreau, H., Herron, M. D., Verbruggen, H., Delwiche, C. F., & De Clerck, O. (2012). Phylogeny and Molecular Evolution of the Green Algae. *Critical Reviews in Plant Sciences*, *31*(1), 1–46.

Leliaert, F., Tronholm, A., Lemieux, C., Turmel, M., DePriest, M. S., Bhattacharya, D., Karol, K. G., Fredericq, S., Zechman, F. W., & Lopez-Bautista, J. M. (2016). Chloroplast phylogenomic analyses reveal the deepest-branching lineage of the Chlorophyta, Palmophyllophyceae class. nov. *Scientific Reports*, *6*, 25367.

Leliaert, F., Verbruggen, H., & Zechman, F. W. (2011). Into the deep: new discoveries at the base of the green plant phylogeny. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *33*(9), 683–692.

Li, F. W., Villarreal, J. C., & Szovenyi, P. (2017). Hornworts: An Overlooked Window into Carbon-Concentrating Mechanisms. *Trends Plant Sci.*, *22*(4), 275–277.

Li, L., Wang, S., Wang, H., Sahu, S. K., Marin, B., Li, H., Xu, Y., Liang, H., Li, Z., Cheng, S., Reder, T., Çebi, Z., Wittek, S., Petersen, M., Melkonian, B., Du, H., Yang, H., Wang, J., Wong, G. K.-S., Xu, X., Liu, X., Van de Peer, Y., Melkonian, M., & Liu, H. (2020). The genome of Prasinoderma coloniale unveils the existence of a third phylum within green plants. *Nature Ecology & Evolution*, *4*(9), 1220–1231.

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, Q., Jiang, Y.-L., Xia, L.-Y., Chen, Y., & Zhou, C.-Z. (2022). Structural insights into cyanobacterial RuBisCO assembly coordinated by two chaperones Raf1 and RbcX. *Cell Discovery*, *8*(1), 93.

Lin, M. T., Orr, D. J., Worrall, D., Parry, M. A. J., Carmo-Silva, E., & Hanson, M. R. (2021). A procedure to introduce point mutations into the Rubisco large subunit gene in wild-type plants. *The Plant Journal: For Cell and Molecular Biology*, *106*(3), 876–887.

Lin, S., & Carpenter, E. J. (1997). Rubisco of Dunaliella tertiolecta is redistributed between the pyrenoid and the stroma as a light /shade response. *Marine Biology*, *127*(3), 521–529.

Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., & Russell, R. B. (2003). Protein disorder prediction: implications for structural proteomics. *Structure* , *11*(11), 1453–1459.

Liu, D., Ramya, R. C. S., & Mueller-Cajar, O. (2017). Surveying the expanding prokaryotic Rubisco multiverse. *FEMS Microbiology Letters*, *364*(16). https://doi.org/10.1093/femsle/fnx156

Liu, J., & Chen, F. (2016). Biology and Industrial Applications of Chlorella: Advances and Prospects. *Advances in Biochemical Engineering/Biotechnology*, *153*, 1–35.

Lokhorst, G. M., Sluiman, H. J., & Star, W. (1988). The Ultrastructure of Mitosis and Cytokinesis in the Sarcinoid Chlorokybus atmophticus (Chlorophyta, Charophyceae) Revealed by Rapid Freeze Fixation and Freeze Substitution. *J. Phycol.*, *24*(2), 237–248.

Lokhorst, G. M., & Star, W. (1980). Pyrenoid Ultrastructure in Ulothrix (Chlorophyceae). *Acta Bot. Neerl.*, *29*(1), 1–15.

Long, B. M., Förster, 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. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(18). https://doi.org/10.1073/pnas.2014406118

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 CO2-fixing enzyme Rubisco in tobacco chloroplasts. *Nat. Commun.*, *9*(1), 3570.

Long, S. P., Burgess, S., & Causton, I. (2019). Redesigning crop photosynthesis. In R. Zeigler (Ed.), *Sustaining Global Food Security: The Nexus of Science and Policy* (pp. 128–147). CSIRO Publishing.

López-Calcagno, P. E., Brown, K. L., Simkin, A. J., Fisk, S. J., Vialet-Chabrand, S., Lawson, T., & Raines, C. A. (2020). Stimulating photosynthetic processes increases productivity and water-use efficiency in the field. *Nature Plants*, *6*(8), 1054–1063.

Ludwig, M., Sültemeyer, D., & Price, G. D. (2000). Isolation of ccmKLMN genes from the marine cyanobacterium, Synechococcus sp. PCC7002 (Cyanophyceae), and evidence that CcmM is essential for carboxysome assembly. *Journal of Phycology*, *36*(6), 1109–1119.

Lundholm, N., & Moestrup, Ø. (2002). The marine diatom Pseudo-nitzschia galaxiae sp. nov. (Bacillariophyceae): morphology and phylogenetic relationships. *Phycologia*, *41*(6), 594–605.

Lupo, U., Sgarbossa, D., & Bitbol, A.-F. (2022). Protein language models trained on multiple sequence alignments learn phylogenetic relationships. *Nature Communications*, *13*(1), 6298.

Mackinder, L. C. M. (2018). The Chlamydomonas CO2 -concentrating mechanism and its potential for engineering photosynthesis in plants. *The New Phytologist*, *217*(1), 54–61.

Mackinder, L. C. M., Chen, C., Leib, R. D., Patena, W., Blum, S. R., Rodman, M., Ramundo, S., Adams, C. M., & Jonikas, M. C. (2017). A Spatial Interactome Reveals the Protein Organization of the Algal CO2-Concentrating Mechanism. *Cell*, *171*(1), 133-147.e14.

Mackinder, L. C. M., Meyer, M. T., Mettler-Altmann, T., Chen, V. K., Mitchell, M. C., Caspari, O., Freeman Rosenzweig, E. S., Pallesen, L., Reeves, G., Itakura, A., Roth, R., Sommer, F.,

Geimer, S., Mühlhaus, T., Schroda, M., Goodenough, U., Stitt, M., Griffiths, H., & Jonikas, M. C. (2016). A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(21), 5958–5963.

Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research*, *50*(W1), W276–W279.

Mangan, N., & Brenner, M. (2014). Systems analysis of the CO2 concentrating mechanism in cyanobacteria. *ELife*, *3*, e02043.

Mangan, N. M., Flamholz, A., Hood, R. D., Milo, R., & Savage, D. F. (2016). pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(36), E5354-62.

Mann, D. G. (1985). In vivo Observations of Plastid and Cell Division in Raphid Diatoms and Their Relevance to Diatom Systematics. *Annals of Botany*, *55*(1), 95–108.

Manton, I. (1967). Further observations on the fine structure of Chrysochromulina chiton with special reference to the haptonema, “peculiar” golgi structure and scale production. *J. Cell Sci.*, *2*(2), 265–272.

Manton, I., & Ettl, H. (1965). Observations on the fine structure of Mesostigma viride Lauterborn. *The Journal of the Linnean Society of London*, *59*(378), 175–184.

Manton, I., & Leedale, G. F. (1961). Further Observations on the Fine Structure of Chrysochromulina Minor and C. Kappa with Special Reference to the Pyrenoids. *Journal of the Marine Biological Association of the United Kingdom. Marine Biological Association of the United Kingdom*, *41*(2), 519–526.

Manuel, L. J., & Moroney, J. V. (1988). Inorganic Carbon Accumulation by Chlamydomonas reinhardtii: New Proteins are made During Adaptation to Low CO(2). *Plant Physiology*, *88*(2), 491–496.

Mao, Y., Catherall, E., Díaz-Ramos, A., Greiff, G. R. L., Azinas, S., Gunn, L., & McCormick, A. J. (2022). The small subunit of Rubisco and its potential as an engineering target. *Journal of Experimental Botany*. https://doi.org/10.1093/jxb/erac309

Mariani, V., Biasini, M., Barbato, A., & Schwede, T. (2013). lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* , *29*(21), 2722–2728.

Matusiak-Mikulin, K., Tukaj, C., & Tukaj, Z. (2006). Relationships between growth, development and photosynthetic activity during the cell cycle of Desmodesmus armatus (Chlorophyta) in synchronous cultures. *European Journal of Phycology*, *41*(1), 29–38.

McAllister, F. (1914). The Pyrenoid of Anthoceros. *American Journal of Botany*, *1*(2), 79–95.

McAllister, F. (1927). The Pyrenoids of Anthoceros and Notothylas with Especial Reference to Their Presence in Spore Mother Cells. *American Journal of Botany*, *14*(5), 246–257.

McDonald, K. L., & Pickett-Heaps, J. D. (1976). Ultrastructure and differentiation in Cladophora glomerata. I. Cell division. *Am. J. Bot.*, *63*(5), 592–601.

McGrath, J. M., & Long, S. P. (2014). Can the Cyanobacterial Carbon-Concentrating Mechanism Increase Photosynthesis in Crop Species? A Theoretical Analysis. *Plant Physiology*, *164*(4), 2247.

McKay, R. M., & Gibbs, S. P. (1990). Phycoerythrin is absent from the pyrenoid of Porphyridium cruentum: photosynthetic implications. *Planta*, *180*(2), 249–256.

McKay, R. M. L., & Gibbs, S. P. (1989). Immunocytochemical localization of ribulose 1,5-bisphosphate carboxylase/oxygenase in light-limited and light-saturated cells of Chlorella pyrenoidosa. *Protoplasma*, *149*(1), 31–37.

McKay, R. M. L., & Gibbs, S. P. (1991). Composition and function of pyrenoids: cytochemical and immunocytochemical approaches. *Can. J. Bot.*, *69*(5), 1040–1052.

McKay, R. M. L., Gibbs, S. P., & Vaughn, K. C. (1991). RuBisCo activase is present in the pyrenoid of green algae. *Protoplasma*, *162*(1), 38–45.

McSwiggen, D. T., Mir, M., Darzacq, X., & Tjian, R. (2019). Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes & Development*, *33*(23–24), 1619–1634.

Mészáros, B., Erdos, G., & Dosztányi, Z. (2018). IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Research*, *46*(W1), W329–W337.

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. *Nature Communications*, *13*(1), 4863.

Meyer, M. T., Goudet, M. M. M., & Griffiths, H. (2020). The Algal Pyrenoid. In A. W. D. Larkum, A. R. Grossmann, & J. A. Raven (Eds.), *Photosynthesis in Algae: Biochemical and Physiological Mechanisms* (pp. 179–203). Springer International Publishing.

Meyer, M. T., Itakura, A. K., Patena, W., Wang, L., He, S., Emrich-Mills, T., Lau, C. S., Yates, G., Mackinder, L. C. M., & Jonikas, M. C. (2020). Assembly of the algal CO2-fixing organelle, the pyrenoid, is guided by a Rubisco-binding motif. *Science Advances*, *6*(46). https://doi.org/10.1126/sciadv.abd2408

Meyer, M. T., McCormick, A. J., & Griffiths, H. (2016). Will an algal CO2-concentrating mechanism work in higher plants? *Current Opinion in Plant Biology*, *31*, 181–188.

Meyer, M. T., Whittaker, C., & Griffiths, H. (2017). The algal pyrenoid: key unanswered questions. *Journal of Experimental Botany*, *68*(14), 3739–3749.

Middleton, C. P., Senerchia, N., Stein, N., Akhunov, E. D., Keller, B., Wicker, T., & Kilian, B. (2014). Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the Triticeae tribe. *PloS One*, *9*(3), e85761.

Mikhailyuk, T., Holzinger, A., Massalski, A., & Karsten, U. (2014). Morphology and ultrastructure of Interfilum and Klebsormidium (Klebsormidiales, Streptophyta) with special reference to cell division and thallus formation. *European Journal of Phycology*, *49*(4), 395–412.

Mikhailyuk, T., Lukešová, A., Glaser, K., Holzinger, A., Obwegeser, S., Nyporko, S., Friedl, T., & Karsten, U. (2018). New Taxa of Streptophyte Algae (Streptophyta) from Terrestrial Habitats Revealed Using an Integrative Approach. *Protist*, *169*(3), 406–431.

Ming, R., VanBuren, R., Wai, C. M., Tang, H., Schatz, M. C., Bowers, J. E., Lyons, E., Wang, M.-L., Chen, J., Biggers, E., Zhang, J., Huang, L., Zhang, L., Miao, W., Zhang, J., Ye, Z., Miao, C., Lin, Z., Wang, H., Zhou, H., Yim, W. C., Priest, H. D., Zheng, C., Woodhouse, M., Edger, P. P., Guyot, R., Guo, H.-B., Guo, H., Zheng, G., Singh, R., Sharma, A., Min, X., Zheng, Y., Lee, H., Gurtowski, J., Sedlazeck, F. J., Harkess, A., McKain, M. R., Liao, Z., Fang, J., Liu, J., Zhang, X., Zhang, Q., Hu, W., Qin, Y., Wang, K., Chen, L.-Y., Shirley, N., Lin, Y.-R., Liu, L.-Y., Hernandez, A. G., Wright, C. L., Bulone, V., Tuskan, G. A., Heath, K., Zee, F., Moore, P. H., Sunkar, R., Leebens-Mack, J. H., Mockler, T., Bennetzen, J. L., Freeling, M., Sankoff, D., Paterson, A. H., Zhu, X., Yang, X., Smith, J. A. C., Cushman, J. C., Paull, R. E., & Yu, Q. (2015). The pineapple genome and the evolution of CAM photosynthesis. *Nature Genetics*, *47*(12), 1435–1442.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature Methods*, *19*(6), 679–682.

Mitra, M., Lato, S. M., Ynalvez, R. A., Xiao, Y., & Moroney, J. V. (2004). Identification of a new chloroplast carbonic anhydrase in Chlamydomonas reinhardtii. *Plant Physiology*, *135*(1), 173–182.

Mitrea, D. M., & Kriwacki, R. W. (2016). Phase separation in biology; functional organization of a higher order. *Cell Communication and Signaling: CCS*, *14*, 1.

Miura, K., Yamano, T., Yoshioka, S., Kohinata, T., Inoue, Y., Taniguchi, F., Asamizu, E., Nakamura, Y., Tabata, S., Yamato, K. T., Ohyama, K., & Fukuzawa, H. (2004). Expression profiling-based identification of CO2-responsive genes regulated by CCM1 controlling a carbon-concentrating mechanism in Chlamydomonas reinhardtii. *Plant Physiology*, *135*(3), 1595–1607.

Miyachi, S., Tsuzuki, M., Maruyama, I., Gantar, M., Miyachi, S., & Matsushima, H. (1986). Effects of co2concentration during growth on the intracellular structure of Chlorella and Scenedesmus (Chlorophyta). *Journal of Phycology*, *22*(3), 313–319.

Miyamura, S., Hori, T., Ohya, T., Tohma, T., & Ikawa, T. (1996). Co-localization of chloroplast DNA and ribulose-1,5-bisphosphate carboxylase/oxygenase in the so-called pyrenoid of the

siphonous green alga Caulerpa lentillifera (Caulerpales, Chlorophyta). *Phycologia*, *35*(2), 156–160.
Montañez, I. P., & Poulsen, C. J. (2013). *The Late Paleozoic Ice Age: An Evolving Paradigm*. https://doi.org/10.1146/annurev.earth.031208.100118
Montresor, M., Procaccini, G., & Stoecker, D. K. (1999). Polarella glacialis, gen. Nov., Sp. Nov. (Dinophyceae): Suessiaceae are still alive! *Journal of Phycology*, *35*(1), 186–197.
Moose, S. P., & Mumm, R. H. (2008). Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiology*, *147*(3), 969–977.
Morita, E., Abe T Fau-Tsuzuki, M., Tsuzuki M Fau-Fujiwara, S., Fujiwara S Fau-Sato, N., Sato N Fau-Hirata, A., Hirata A Fau-Sonoike, K., Sonoike K Fau-Nozaki, H., & Nozaki, H. (1998). Presence of the CO2-concentrating mechanism in some species of the pyrenoid-less free-living algal genus Chloromonas (Volvocales, Chlorophyta). *Planta*, *204*(3), 269–276.
Mueller-Cajar, O., Stotz, M., Wendler, P., Hartl, F. U., Bracher, A., & Hayer-Hartl, M. (2011). Structure and function of the AAA+ protein CbbX, a red-type Rubisco activase. *Nature*, *479*(7372), 194–199.
Mukherjee, A., Lau, C. S., Walker, C. E., Rai, A. K., Prejean, C. I., Yates, G., Emrich-Mills, T., Lemoine, S. G., Vinyard, D. J., Mackinder, L. C. M., & Moroney, J. V. (2019). Thylakoid localized bestrophin-like proteins are essential for the CO2 concentrating mechanism of Chlamydomonas reinhardtii. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(34), 16915–16920.
Munakata, H., Nakada, T., Nakahigashi, K., Nozaki, H., & Tomita, M. (2016). Phylogenetic Position and Molecular Chronology of a Colonial Green Flagellate, Stephanosphaera pluvialis (Volvocales, Chlorophyceae), among Unicellular Algae. *The Journal of Eukaryotic Microbiology*, *63*(3), 340–348.
Mustardy, L., Cunningham, F. X., & Gantt, E. (1990). Localization and quantitation of chloroplast enzymes and light-harvesting components using immunocytochemical methods. *Plant Physiol.*, *94*(1), 334–340.
Nagasato, C., & Motomura, T. (2002). New Pyrenoid Formation in the Brown Alga, Scytosiphon lomentaria (Scytosiphonales, Phaeophyceae). *J. Phycol.*, *38*(4), 800–806.
Nam, S. W., Go, D., Son, M., & Shin, W. (2013). Ultrastructure of the flagellar apparatus in Rhinomonas reticulata var. atrorosea (Cryptophyceae, Cryptophyta). *Algae*, *28*(4), 331–341.
Nassoury, N., Fritz, L., & Morse, D. (2001). Circadian changes in ribulose-1,5-bisphosphate carboxylase/oxygenase distribution inside individual chloroplasts can account for the rhythm in dinoflagellate carbon fixation. *The Plant Cell*, *13*(4), 923–934.
Nawaly, H., Tanaka, A., Toyoshima, Y., Tsuji, Y., & Matsuda, Y. (2023). Localization and characterization θ carbonic anhydrases in Thalassiosira pseudonana. *Photosynthesis Research*, *156*(2), 217–229.
Nelson, W. A., & Ryan, K. G. (1988). Porphyridium purpureum (Bory) Drew et Ross (Porphyridiales, Rhodophyceae)—first record of a marine unicellular red alga in New Zealand. *Journal of the Royal Society of New Zealand*, *18*(1), 127–128.
Němcová, Y., & Kalina, T. (2000). Cell wall development, microfibril and pyrenoid structure in type strains of Chlorella vulgaris, C. kessleri, C. sorokiniana compared with C. luteoviridis (Trebouxiophyceae, Chlorophyta). *Algological Studies/Archiv Für Hydrobiologie, Supplement Volumes*, 95–105.
Neupert, J., Gallaher, S. D., Lu, Y., Strenkert, D., Segal, N., Barahimipour, R., Fitz-Gibbon, S. T., Schroda, M., Merchant, S. S., & Bock, R. (2020). An epigenetic gene silencing pathway selectively acting on transgenic DNA in the green alga Chlamydomonas. *Nature Communications*, *11*(1), 6269.
Newman, A. M., & Cooper, J. B. (2007). XSTREAM: a practical algorithm for identification and architecture modeling of tandem repeats in protein sequences. *BMC Bioinformatics*, *8*, 382.
Newman, J., Branden, C. I., & Jones, T. A. (1993). Structure determination and refinement of ribulose 1,5-bisphosphate carboxylase/oxygenase from Synechococcus PCC6301. *Acta Crystallographica. Section D, Biological Crystallography*, *49*(Pt 6), 548–560.

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 α-carboxysomes. *Nature Communications*, *13*(1), 4299.

Nievergelt, A. P., Diener, D. R., Bogdanova, A., Brown, T., & Pigino, G. (2022). *Efficient Precision Editing of Endogenous Chlamydomonas reinhardtii Genes with Crispr/Cas*. https://doi.org/10.2139/ssrn.4278953

Nobel, P. S. (1991). Achievable productivities of certain CAM plants: basis for high values compared with C3 and C4 plants. *The New Phytologist*, *119*(2), 183–205.

Nozaki, H., Ott, F. D., & Coleman, A. W. (2006). Morphology, molecular phylogeny and taxonomy of two new species of Pleodorina (volvoceae, Chlorophyceae). *Journal of Phycology*, *42*(5), 1072–1080.

Oborník, M., Modrý, D., Lukeš, M., Cernotíková-Stříbrná, E., Cihlář, J., Tesařová, M., Kotabová, E., Vancová, M., Prášil, O., & Lukeš, J. (2012). Morphology, ultrastructure and life cycle of Vitrella brassicaformis n. sp., n. gen., a novel chromerid from the Great Barrier Reef. *Protist*, *163*(2), 306–323.

Oborník, M., Vancová, M., Lai, D.-H., Janouškovec, J., Keeling, P. J., & Lukeš, J. (2011). Morphology and ultrastructure of multiple life cycle stages of the photosynthetic relative of apicomplexa, Chromera velia. *Protist*, *162*(1), 115–130.

Ogren, W. L. (1984). Photorespiration: Pathways, regulation, and modification. *Annual Review of Plant Physiology*, *35*(1), 415–442.

Oh, Z. G., Ang, W. S. L., Poh, C. W., Lai, S.-K., Sze, S. K., Li, H.-Y., Bhushan, S., Wunder, T., & Mueller-Cajar, O. (2023). A linker protein from a red-type pyrenoid phase separates with Rubisco via oligomerizing sticker motifs. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(25), e2304833120.

Oh, Z. G., Askey, B., & Gunn, L. H. (2022). Red Rubiscos and opportunities for engineering green plants. *Journal of Experimental Botany*. https://doi.org/10.1093/jxb/erac349

Ohad, I., Siekevitz, P., & Palade, G. E. (1967). Biogenesis of chloroplast membranes. I. Plastid dedifferentiation in a dark-grown algal mutant (Chlamydomonas reinhardi). *The Journal of Cell Biology*, *35*(3), 521–552.

Ohiwa, T. (1976). Observations on chloroplast growth and pyrenoid formation in Spirogyra. A study by means of uncoiled picture of chloroplast. *Bot Mag Tokyo*, *89*(4), 259–266.

O'Kelly, C. J., Sieracki, M. E., Thier, E. C., & Hobson, I. C. (2003). A transient bloom of Ostreococcus (Chlorophyta, Prasinophyceae) in west neck bay, Long Island, New York. *Journal of Phycology*, *39*(5), 850–854.

Oleksienko, A. A., Kot, Y. G., & Komaristaya, V. P. (2020). DNA-Specific DAPI Staining of the Pyrenoid Matrix During its Fission in Dunaliella salina (Dunal) Teodoresco (Chlorophyta). *Current Microbiology*, *77*(11), 3450–3459.

Oltrogge, L. M., Chaijarasphong, T., Chen, A. W., Bolin, E. R., Marqusee, S., & Savage, D. F. (2020). Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation. *Nature Structural & Molecular Biology*, *27*(3), 281–287.

Osafune, T., Sumida, S., Ehara, T., & Hase, E. (1989). Three-Dimensional Distribution of Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase in Chloroplasts of Actively Photosynthesizing Cell of Euglena gracilis. *J. Electron Microsc.*, *38*(5), 399–402.

Osafune, T., Yokota, A., Sumida, S., & Hase, E. (1990). Immunogold Localization of Ribulose-1,5-Bisphosphate Carboxylase with Reference to Pyrenoid Morphology in Chloroplasts of Synchronized Euglena gracilis Cells. *Plant Physiology*, *92*(3), 802–808.

Ota, S., Ueda, K., & Ishida, K.-I. (2005). Lotharella vacuolata sp. nov., a new species of chlorarachniophyte algae, and time-lapse video observations on its unique post-cell division behavior. *Phycological Res.*, *53*(4), 275–286.

Ott, C. M., Smith, B. D., Portis, A. R., Jr, & Spreitzer, R. J. (2000). Activase region on chloroplast ribulose-1,5-bisphosphate carboxylase/oxygenase. Nonconservative substitution in the large subunit alters species specificity of protein interaction. *The Journal of Biological Chemistry*, *275*(34), 26241–26244.

Owen, I., & Shewmaker, F. (2019). The Role of Post-Translational Modifications in the Phase Transitions of Intrinsically Disordered Proteins. *International Journal of Molecular Sciences*, *20*(21). https://doi.org/10.3390/ijms20215501

Pace, C. N., & Scholtz, J. M. (1998). A helix propensity scale based on experimental studies of peptides and proteins. *Biophysical Journal*, *75*(1), 422–427.

Palmqvist, K. (1993). Photosynthetic CO2-use efficiency in lichens and their isolated photobionts: the possible role of a CO2-concentrating mechanism. *Planta*, *191*(1), 48–56.

Patron, N. J., Orzaez, D., Marillonnet, S., Warzecha, H., Matthewman, C., Youles, M., Raitskin, O., Leveau, A., Farré, G., Rogers, C., Smith, A., Hibberd, J., Webb, A. A. R., Locke, J., Schornack, S., Ajioka, J., Baulcombe, D. C., Zipfel, C., Kamoun, S., Jones, J. D. G., Kuhn, H., Robatzek, S., Van Esse, H. P., Sanders, D., Oldroyd, G., Martin, C., Field, R., O'Connor, S., Fox, S., Wulff, B., Miller, B., Breakspear, A., Radhakrishnan, G., Delaux, P.-M., Loqué, D., Granell, A., Tissier, A., Shih, P., Brutnell, T. P., Quick, W. P., Rischer, H., Fraser, P. D., Aharoni, A., Raines, C., South, P. F., Ané, J.-M., Hamberger, B. R., Langdale, J., Stougaard, J., Bouwmeester, H., Udvardi, M., Murray, J. A. H., Ntoukakis, V., Schäfer, P., Denby, K., Edwards, K. J., Osbourn, A., & Haseloff, J. (2015). Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *The New Phytologist*, *208*(1), 13–19.

Patrone, L. M., Broadwater, S. T., & Scott, J. L. (1991). Ultrastructure of vegetative and dividing cells of the unicellular red algae Rhodella violacea and Rhodella maculata. *J. Phycol.*, *27*(6), 742–753.

Peng, Y., Yan, H., Guo, L., Deng, C., Wang, C., Wang, Y., Kang, L., Zhou, P., Yu, K., Dong, X., Liu, X., Sun, Z., Peng, Y., Zhao, J., Deng, D., Xu, Y., Li, Y., Jiang, Q., Li, Y., Wei, L., Wang, J., Ma, J., Hao, M., Li, W., Kang, H., Peng, Z., Liu, D., Jia, J., Zheng, Y., Ma, T., Wei, Y., Lu, F., & Ren, C. (2022). Reference genome assemblies reveal the origin and evolution of allohexaploid oat. *Nature Genetics*, *54*(8), 1248–1258.

Peona, V., Blom, M. P. K., Xu, L., Burri, R., Sullivan, S., Bunikis, I., Liachko, I., Haryoko, T., Jønsson, K. A., Zhou, Q., Irestedt, M., & Suh, A. (2021). Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *Molecular Ecology Resources*, *21*(1), 263–286.

Peran, I., & Mittag, T. (2020). Molecular structure in biomolecular condensates. *Current Opinion in Structural Biology*, *60*, 17–26.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science: A Publication of the Protein Society*, *30*(1), 70–82.

Pichard, S. L., Campbell, L., & Paul, J. H. (1997). Diversity of the ribulose bisphosphate carboxylase/oxygenase form I gene (rbcL) in natural phytoplankton communities. *Applied and Environmental Microbiology*, *63*(9), 3600–3606.

Polle, J. E. W., Roth, R., Ben-Amotz, A., & Goodenough, U. (2020). Ultrastructure of the green alga Dunaliella salina strain CCAP19/18 (Chlorophyta) as investigated by quick-freeze deep-etch electron microscopy. *Algal Research*, *49*, 101953.

Possmayer, M., Gupta, R. K., Szyszka-Mroz, B., Maxwell, D. P., Lachance, M.-A., Hüner, N. P. A., & Smith, D. R. (2016). Resolving the phylogenetic relationship between Chlamydomonas sp. UWO 241 and Chlamydomonas raudensis sag 49.72 (Chlorophyceae) with nuclear and plastid DNA sequences. *Journal of Phycology*, *52*(2), 305–310.

Price, D. C., Goodenough, U. W., Roth, R., Lee, J.-H., Kariyawasam, T., Mutwil, M., Ferrari, C., Facchinelli, F., Ball, S. G., Cenci, U., Chan, C. X., Wagner, N. E., Yoon, H. S., Weber, A. P. M., & Bhattacharya, D. (2019). Analysis of an improved Cyanophora paradoxa genome assembly. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, *26*(4), 287–299.

Price, G. D., & Badger, M. R. (1989). Expression of Human Carbonic Anhydrase in the Cyanobacterium Synechococcus PCC7942 Creates a High CO2-Requiring Phenotype

Evidence for a Central Role for Carboxysomes in the CO2 Concentrating Mechanism. *Plant Physiol.*, *91*(2), 505–513.
Pritišanac, I., Vernon, R. M., Moses, A. M., & Forman Kay, J. D. (2019). Entropy and Information within Intrinsically Disordered Protein Regions. *Entropy* , *21*(7). https://doi.org/10.3390/e21070662
Pröschold, T., & Darienko, T. (2018). *The distinctions between Chlamydomonas incerta Pascher and Chlamydomonas globosa J.W. Snow and their taxonomic consequences*. notulaealgarum.org. https://www.notulaealgarum.org/documents/Notulae%20algarum%20No.%2056.pdf
Pröschold, T., Darienko, T., Krienitz, L., & Coleman, A. W. (2018). Chlamydomonas schloesseri sp. nov. (Chlamydophyceae, Chlorophyta) revealed by morphology, autolysin cross experiments, and multiple gene analyses. *Phytotaxa*, *362*(1), 21.
Pröschold, T., Darienko, T., Silva, P. C., Reisser, W., & Krienitz, L. (2011). The systematics of Zoochlorella revisited employing an integrative approach. *Environmental Microbiology*, *13*(2), 350–364.
Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, *14*(3), 290–296.
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., Förster, 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. *Journal of Experimental Botany*, *68*(14), 3717–3737.
Ramazanov, Z., Rawat, M., Henk, M. C., Mason, C. B., Matthews, S. W., & Moroney, J. V. (1994). The induction of the CO2-concentrating mechanism is correlated with the formation of the starch sheath around the pyrenoid of Chlamydomonas reinhardtii. *Planta*, *195*(2), 210–216.
Raven, J. A. (2013). Rubisco: still the most abundant protein of Earth? *New Phytol.*, *198*(1), 1–3.
Raven, J. A., Beardall, J., & Sánchez-Baracaldo, P. (2017). The possible evolution and future of CO2-concentrating mechanisms. *Journal of Experimental Botany*, *68*(14), 3701–3716.
Raven, J. A., Giordano, M., Beardall, J., & Maberly, S. C. (2012). Algal evolution in relation to atmospheric CO2: carboxylases, carbon-concentrating mechanisms and carbon oxidation cycles. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1588), 493–507.
Rawat, P., Boehning, M., Hummel, B., Aprile-Garcia, F., Pandit, A. S., Eisenhardt, N., Khavaran, A., Niskanen, E., Vos, S. M., Palvimo, J. J., Pichler, A., Cramer, P., & Sawarkar, R. (2021). Stress-induced nuclear condensation of NELF drives transcriptional downregulation. *Molecular Cell*, *81*(5), 1013-1026.e11.
Retallack, B., & Butler, R. D. (1970). The development and structure of pyrenoids in Bulbochaete hiloensis. *Journal of Cell Science*, *6*(1), 229–241.
Reynolds, E. S. (1963). The use of lead citrate at high pH as an electron-opaque stain in electron microscopy. *The Journal of Cell Biology*, *17*(1), 208–212.
Richardson, L. (2007). *Beautiful Soup Documentation*. media.readthedocs.org. https://media.readthedocs.org/pdf/beautiful-soup-4/latest/beautiful-soup-4.pdf
Rohl, C. A., Fiori, W., & Baldwin, R. L. (1999). Alanine is helix-stabilizing in both template-nucleated and standard peptide helices. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(7), 3682–3687.
Rohou, A., & Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, *192*(2), 216–221.
Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., & Dunker, A. K. (2001). Sequence complexity of disordered protein. *Proteins*, *42*(1), 38–48.
Rosowski, K. A., Sai, T., Vidal-Henriquez, E., Zwicker, D., Style, R. W., & Dufresne, E. R. (2020). Elastic ripening and inhibition of liquid-liquid phase separation. *Nature Physics*, *16*(4), 422–425.

Safi, C., Zebib, B., Merah, O., Pontalier, P.-Y., & Vaca-Garcia, C. (2014). Morphology, composition, production, processing and applications of Chlorella vulgaris: A review. *Renewable and Sustainable Energy Reviews*, *35*, 265–278.
Sage, R. F. (2004). The evolution of C4 photosynthesis. *The New Phytologist*, *161*(2), 341–370.
Sage, R. F., Gilman, I. S., Smith, J. A. C., Silvera, K., & Edwards, E. J. (2023). Atmospheric CO2 decline and the timing of CAM plant evolution. *Annals of Botany*. https://doi.org/10.1093/aob/mcad122
Salisbury, J. L., & Floyd, G. L. (1978). Molecular, enzymatic and ultrastructure characterization of the scaly green monad Micromonas squamata. *Journal of Phycology*, *14*(3), 362–368.
Salomé, P. A., & Merchant, S. S. (2019). A Series of Fortunate Events: Introducing Chlamydomonas as a Reference Organism. *Plant Cell*, *31*(8), 1682–1707.
Santos, L. M. A., & Leedale, G. F. (1995). Some notes on the ultrastructure of small azoosporic members of the algal class Eustigmatophyceae. *Nova Hedwigia*, *60*(1), 219–226.
Särkinen, T., Bohs, L., Olmstead, R. G., & Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evolutionary Biology*, *13*, 214.
Satagopan, S., & Spreitzer, R. J. (2004). Substitutions at the Asp-473 Latch Residue of Chlamydomonas Ribulosebisphosphate Carboxylase/Oxygenase Cause Decreases in Carboxylation Efficiency and CO2/O2 Specificity *. *The Journal of Biological Chemistry*, *279*(14), 14240–14244.
Sato, M., Mogi, Y., Nishikawa, T., Miyamura, S., Nagumo, T., & Kawano, S. (2009). The dynamic surface of dividing cyanelles and ultrastructure of the region directly below the surface in Cyanophora paradoxa. *Planta*, *229*(4), 781.
Satoh, H., Okada, M., Nakayama, K., & Murata, T. (1985). Purification of Ribulose 5-phosphate Kinase and Minor Polypeptides of Pyrenoid from the Green Alga Bryopsis maxima. *Plant & Cell Physiology*, *26*(5), 931–940.
Schaefer, C. (2023). Competitive Binding of Force-Extensible Sticker-and-Spacer “EPYC1” Polypeptides to a Patchy Colloidal “Rubisco” Protein. In *arXiv [cond-mat.soft]*. arXiv. http://arxiv.org/abs/2301.05681
Scheres, S. H. W. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, *180*(3), 519–530.
Schimper, A. W. F. (1885). Untersuchungen über die Chlorophyllkörper und die ihnen homologen Gebilde. *Jahrb. Wiss. Bot*, *16*, 1–247.
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. *Nature Methods*, *9*(7), 676–682.
Schmitz, F. (1882). *Die chromatophoren der algen: Vergleichende untersuchungen über bau und entwicklung der chlorophyllkörper und der analogen farbstoffkörper der algen*. M. Cohen & Sohn (F. Cohen).
Schornstein, K. L., & Scott, J. (1982). Ultrastructure of cell division in the unicellular red alga Porphyridium purpureum. *Canadian Journal of Botany. Journal Canadien de Botanique*, *60*(1), 85–97.
Schroda, M. (2019). Good News for Nuclear Transgene Expression in Chlamydomonas. *Cells* , *8*(12). https://doi.org/10.3390/cells8121534
Schuette, S., & Renzaglia, K. S. (2010). Development of multicellular spores in the hornwort genus Dendroceros (Dendrocerotaceae, Anthocerotophyta) and the occurrence of endospory in Bryophytes. *Nova Hedwigia*, *91*(3–4), 301–316.
Schulz, K. G., Riebesell, U., Rost, B., Thoms, S., & Zeebe, R. E. (2006). Determination of the rate constants for the carbon dioxide to bicarbonate inter-conversion in pH-buffered seawater systems. *Marine Chemistry*, *100*(1), 53–65.
Scott, J. L., Baca, B., Ott, F. D., & West, J. A. (2006). Light and Electron Microscopic Observations on Erythrolobus coxiae gen. et sp. nov.(Porphyridiophyceae, Rhodophyta) from Texas USA. *Algae*, *21*(4), 407–416.

Scott, J., Yang, E.-C., West, J. A., Yokoyama, A., Kim, H.-J., De Goer, S. L., O'Kelly, C. J., Orlova, E., Kim, S.-Y., Park, J.-K., & Others. (2011). On the genus Rhodella, the emended orders Dixoniellales and Rhodellales with a new order Glaucosphaerales (Rhodellophyceae, Rhodophyta). *Algae* , *26*(4), 277–288.

Sharkey, T. D. (1988). Estimating the rate of photorespiration in leaves. *Physiologia Plantarum*, *73*(1), 147–152.

Sharma, P. K., Goud, V. V., Yamamoto, Y., & Sahoo, L. (2021). Efficient Agrobacterium tumefaciens-mediated stable genetic transformation of green microalgae, Chlorella sorokiniana. *3 Biotech*, *11*(4), 196.

Shi, B., Li, W., Song, Y., Wang, Z., Ju, R., Ulman, A., Hu, J., Palomba, F., Zhao, Y., Le, J. P., Jarrard, W., Dimoff, D., Digman, M. A., Gratton, E., Zang, C., & Jiang, H. (2021). UTX condensation underlies its tumour-suppressive activity. *Nature*, *597*(7878), 726–731.

Shih, P. M., Occhialini, A., Cameron, J. C., Andralojc, P. J., Parry, M. A. J., & Kerfeld, C. A. (2016). Biochemical characterization of predicted Precambrian RuBisCO. *Nature Communications*, *7*, 10382.

Shimamura, D., Yamano, T., Niikawa, Y., Hu, D., & Fukuzawa, H. (2023). A pyrenoid-localized protein SAGA1 is necessary for Ca2+-binding protein CAS-dependent expression of nuclear genes encoding inorganic carbon transporters in Chlamydomonas reinhardtii. *Photosynthesis Research*, *156*(2), 181–192.

Shinzato, C., Mungpakdee, S., Satoh, N., & Shoguchi, E. (2014). A genomic approach to coral-dinoflagellate symbiosis: studies of Acropora digitifera and Symbiodinium minutum. *Frontiers in Microbiology*, *5*, 336.

Shivhare, D., & Mueller-Cajar, O. (2017). In Vitro Characterization of Thermostable CAM Rubisco Activase Reveals a Rubisco Interacting Surface Loop. *Plant Physiology*, *174*(3), 1505–1516.

Sieburth, J. M., Johnson, P. W., & Hargraves, P. E. (1988). Ultrastructure and ecology of Aureococcus anophageferens gen. Et Sp. Nov. (Chrysophyceae): The dominant picoplankter during a bloom in Narragansett bay, Rhode Island, summer 1985. *Journal of Phycology*, *24*(3), 416–425.

Sims, P. A., Mann, D. G., & Medlin, L. K. (2006). Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia*, *45*(4), 361–402.

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. *Biochimica et Biophysica Acta*, *1817*(8), 1248–1255.

Smith, G. M. (1918). Cytological Studies in the Protococcales. III. Cell Structure and Autospore Formation in Tetraedron minimum (A. Br.), Hansg. *Annals of Botany*, *32*(128), 459–464.

Soli, A. L., & Byrne, R. H. (2002). CO2 system hydration and dehydration kinetics and the equilibrium CO2/H2CO3 ratio in aqueous NaCl solution. *Marine Chemistry*, *78*(2), 65–73.

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. (1993). Genetic Dissection of Rubisco Structure and Function. *Annual Review of Plant Physiology and Plant Molecular Biology*, *44*(1), 411–434.

Spreitzer, R. J., & Salvucci, M. E. (2002). Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annual Review of Plant Biology*, *53*, 449–475.

Stein, J. R. (1958). A morphological study of Astrephomene gubernaculifera and Volvulina steinii. *American Journal of Botany*, *45*(5), 388–397.

Studer, R. A., Christin, P.-A., Williams, M. A., & Orengo, C. A. (2014). Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proceedings of the National Academy of Sciences*, *111*(6), 2223–2228.

Su, Q., Schumann, P., Schild, C., & Boschetti, A. (1999). A processing intermediate of a stromal chloroplast import protein in Chlamydomonas. *Biochemical Journal*, *344 Pt 2*, 391–395.

Subrahmanyan, R. (1945). On the cell-division and mitosis in some South Indian diatoms. *Proceedings of the Indian Academy of Sciences. Mathematical Sciences*. https://www.ias.ac.in/article/fulltext/secb/022/06/0331-0354

Sun, C. N. (1962). Submicroscopic structure and development of the chloroplast and pyrenoid in Anthoceros laevis. *Protoplasma*, *55*(1), 89–98.
Sun, X., Shen, J., Bai, F., & Xu, N. (2016). Transcriptome profiling of the microalga Chlorella pyrenoidosa in response to different carbon dioxide concentrations. *Marine Genomics*, *29*, 81–87.
Sunaga, Y., Maeda, Y., Yabuuchi, T., Muto, M., Yoshino, T., & Tanaka, T. (2015). Chloroplast-targeting protein expression in the oleaginous diatom Fistulifera solaris JPCC DA0580 toward metabolic engineering. *Journal of Bioscience and Bioengineering*, *119*(1), 28–34.
Suzuki, R., Ito, N., Uno, Y., Nishii, I., Kagiwada, S., Okada, S., & Noguchi, T. (2013). Transformation of lipid bodies related to hydrocarbon accumulation in a green alga, Botryococcus braunii (Race B). *PloS One*, *8*(12), e81626.
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. *Journal of Experimental Botany*, *59*(7), 1515–1524.
Tachibana, M., Allen, A. E., Kikutani, S., Endo, Y., Bowler, C., & Matsuda, Y. (2011a). Localization of putative carbonic anhydrases in two marine diatoms, Phaeodactylum tricornutum and Thalassiosira pseudonana. *Photosynth. Res.*, *109*(1–3), 205–221.
Tachibana, M., Allen, A. E., Kikutani, S., Endo, Y., Bowler, C., & Matsuda, Y. (2011b). Localization of putative carbonic anhydrases in two marine diatoms, Phaeodactylum tricornutum and Thalassiosira pseudonana. *Photosynthesis Research*, *109*(1–3), 205–221.
Tanaka, A., Nagasato, C., Uwai, S., Motomura, T., & Kawai, H. (2007). Re-examination of ultrastructures of the stellate chloroplast organization in brown algae: Structure and development of pyrenoids. *Phycological Research*, *55*(3), 203–213.
Tcherkez, G. G. B., Farquhar, G. D., & Andrews, T. J. (2006). Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(19), 7246–7251.
Thomas, P. D., & Dill, K. A. (1996). An iterative method for extracting energy-like quantities from protein structures. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(21), 11628–11633.
Toyokawa, C., Yamano, T., & Fukuzawa, H. (2020). Pyrenoid Starch Sheath Is Required for LCIB Localization and the CO2-Concentrating Mechanism in Green Algae. *Plant Physiology*. https://doi.org/10.1104/pp.19.01587
Treves, H., Raanan, H., Kedem, I., Murik, O., Keren, N., Zer, H., Berkowicz, S. M., Giordano, M., Norici, A., Shotland, Y., Ohad, I., & Kaplan, A. (2016). The mechanisms whereby the green alga Chlorella ohadii, isolated from desert soil crust, exhibits unparalleled photodamage resistance. *The New Phytologist*, *210*(4), 1229–1243.
Tubb, M. R., Smith, L. E., & Davidson, W. S. (2009). Purification of recombinant apolipoproteins A-I and A-IV and efficient affinity tag cleavage by tobacco etch virus protease. *Journal of Lipid Research*, *50*(7), 1497–1504.
Turkina, M. V., Blanco-Rivero, A., Vainonen, J. P., Vener, A. V., & Villarejo, A. (2006). CO2 limitation induces specific redox-dependent protein phosphorylation inChlamydomonas reinhardtii. *Proteomics*, *6*(9), 2693–2704.
Uebel, C. J., & Phillips, C. M. (2019). Phase-separated protein dynamics are affected by fluorescent tag choice. *MicroPublication Biology*, *2019*. https://www.ncbi.nlm.nih.gov/pubmed/32440657
Ueda, K. (1963). The Pyrenoid of Chlorogonium elongatum. In H. Tamiya (Ed.), *Studies on Microalgae and Photosynthetic Bacteria: A Collection of Papers Dedicated to Hiroshi Tamiya on the Occasion of His 60th Birthday* (p. 636). Japanese Society of Plant Physiologists, University of Tokyo Press.
Uwizeye, C., Decelle, J., Jouneau, P.-H., Flori, S., Gallet, B., Keck, J.-B., Bo, D. D., Moriscot, C., Seydoux, C., Chevalier, F., Schieber, N. L., Templin, R., Allorent, G., Courtois, F., Curien, G., Schwab, Y., Schoehn, G., Zeeman, S. C., Falconet, D., & Finazzi, G. (2021). Morphological bases of phytoplankton energy management and physiological responses unveiled by 3D subcellular imaging. *Nature Communications*, *12*(1), 1049.

Valegård, K., Andralojc, P. J., Haslam, R. P., Pearce, F. G., Eriksen, G. K., Madgwick, P. J., Kristoffersen, A. K., van Lun, M., Klein, U., Eilertsen, H. C., Parry, M. A. J., & Andersson, I. (2018). Structural and functional analyses of Rubisco from arctic diatom species reveal unusual posttranslational modifications. *The Journal of Biological Chemistry*, *293*(34), 13033–13043.

van Baren, M. J., Bachy, C., Reistetter, E. N., Purvine, S. O., Grimwood, J., Sudek, S., Yu, H., Poirier, C., Deerinck, T. J., Kuo, A., Grigoriev, I. V., Wong, C.-H., Smith, R. D., Callister, S. J., Wei, C.-L., Schmutz, J., & Worden, A. Z. (2016). Evidence-based green algal genomics reveals marine diversity and ancestral characteristics of land plants. *BMC Genomics*, *17*, 267.

Van Thinh, L., Griffiths, D. J., & Winsor, H. (1986). Ultrastructure of Symbiodinium microadriaticum (Dinophyceae) symbiotic with Zoanthus sp. (Zoanthidea). *Phycologia*, *25*(2), 178–184.

van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E., Melquiond, A. S. J., van Dijk, M., de Vries, S. J., & Bonvin, A. M. J. J. (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology*, *428*(4), 720–725.

Vandenbosch, K. A. (1990). *Immunogold labeling In: Hall JL, Hawes C (Eds.) Electron microscopy of plant cells*. San Diego CA, USA. Academic Press.

Vaucher, J.-P. (1803). *Histoire des conferves d'eau douce, contenant leurs différens modes de reproduction, et la description de leurs principales espèces, suivie de l'histoire des trémelles et des ulves d'eau douce. Par Jean-Pierre Vaucher*.

Vaughn, K. C., Campbell, E. O., Hasegawa, J., Owen, H. A., & Renzaglia, K. S. (1990). The pyrenoid is the site of ribulose 1,5-bisphosphate carboxylase/oxygenase accumulation in the hornwort (Bryophyta: Anthocerotae) chloroplast. *Protoplasma*, *156*(3), 117–129.

Villarejo, A., Martinez, F., del Pino Plumed, M., & Ramazanov, Z. (1996). The induction of the CO2 concentrating mechanism in a starch-less mutant of Chlamydomonas reinhardtii. *Physiologia Plantarum*, *98*(4), 798–802.

Villarejo, A., Orus, M. I., & Martinez, F. (1997). Regulation of the CO2-concentrating mechanism in Chlorella vulgaris UAM 101 by glucose. *Physiologia Plantarum*, *99*(2), 293–301.

Villarejo, A., Orús, M. I., Ramazanov, Z., & Martínez, F. (1998). A 38-kilodalton low-CO2-inducible polypeptide is associated with the pyrenoid in Chlorella vulgaris. *Planta*, *206*(3), 416–425.

Villarreal, J. C., & Renner, S. S. (2012). Hornwort pyrenoids, carbon-concentrating structures, evolved and were lost at least five times during the last 100 million years. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(46), 18873–18878.

Viola, R., Nyvall, P., & Pedersén, M. (2001). The unique features of starch metabolism in red algae. *Proceedings. Biological Sciences / The Royal Society*, *268*(1474), 1417–1422.

Vitlin Gruber, A., & Feiz, L. (2018). Rubisco Assembly in the Chloroplast. *Frontiers in Molecular Biosciences*, *5*, 24.

Vítová, M., Hendrychova, J., Čížková, M., Cepak, V., Umen, J. G., Zachleder, V., & Bišová, K. (2008). Accumulation, activity and localization of cell cycle regulatory proteins and the chloroplast division protein FtsZ in the alga Scenedesmus quadricauda under inhibition of nuclear DNA replication. *Plant & Cell Physiology*, *49*(12), 1805–1817.

Vladimirova, M. G., Markelova, A. G., & Semenenko, V. E. (1982). Use of the cytoimmunofluorescent method to clarify localization of ribulose bisphosphate carboxylase in pyrenoids of unicellular algae. *Fiziol Rast*, *29*(5), 725–734.

von Quillfeldt, C. H. (2001). *Identification of Some Easily Confused Common Diatom Species in Arctic Spring Blooms*. *44*(4), 375–389.

Wagner, R. (1835). Einige bemerkungen und fragen über das keimbläschen (vesicular germinativa). *Müller's Archiv Anat Physiol Wissenschaft Med*, *268*, 373–377.

Wai, C. M., VanBuren, R., Zhang, J., Huang, L., Miao, W., Edger, P. P., Yim, W. C., Priest, H. D., Meyers, B. C., Mockler, T., Smith, J. A. C., Cushman, J. C., & Ming, R. (2017). Temporal and spatial transcriptomic and microRNA dynamics of CAM photosynthesis in pineapple. *The Plant Journal: For Cell and Molecular Biology*, *92*(1), 19–30.

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

Walne, P. L., & Arnott, H. J. (1967). The comparative ultrastructure and possible function of eyespots: Euglena granulata and Chlamydomonas eugametos. *Planta*, *77*(4), 325–353.
Walsh, I., Martin, A. J. M., Di Domenico, T., & Tosatto, S. C. E. (2012). ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics* , *28*(4), 503–509.
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 β-carboxysome biogenesis. *Nature*, *566*(7742), 131–135.
Wang, L., Patena, W., Van Baalen, K. A., Xie, Y., Singer, E. R., Gavrilenko, S., Warren-Williams, M., Han, L., Harrigan, H. R., Hartz, L. D., Chen, V., Ton, V. T. N. P., Kyin, S., Shwe, H. H., Cahn, M. H., Wilson, A. T., Onishi, M., Hu, J., Schnell, D. J., McWhite, C. D., & Jonikas, M. C. (2023). A chloroplast protein atlas reveals punctate structures and spatial organization of biosynthetic pathways. *Cell*, *186*(16), 3499-3518.e14.
Wang, Y., Liu, X., Gao, H., Zhang, H.-M., Guo, A.-Y., Xu, J., & Xu, X. (2020). Early Stage Adaptation of a Mesophilic Green Alga to Antarctica: Systematic Increases in Abundance of Enzymes and LEA Proteins. *Molecular Biology and Evolution*, *37*(3), 849–863.
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.
Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* , *25*(9), 1189–1191.
Weber, C. I. (1965). PYRENOID FORMATION INDUCED IN CLOSTERIUM MONILIFERUM BY CENTRIFUGATION. *Experimental Cell Research*, *38*, 507–510.
Weber, E., Engler, C., Gruetzner, R., Werner, S., & Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PloS One*, *6*(2), e16765.
Wei, M.-T., Elbaum-Garfinkle, S., Holehouse, A. S., Chen, C. C.-H., Feric, M., Arnold, C. B., Priestley, R. D., Pappu, R. V., & Brangwynne, C. P. (2017). Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nature Chemistry*, *9*(11), 1118–1125.
Weiss, D., Schneider, G., Niemann, B., Guttmann, P., Rudolph, D., & Schmahl, G. (2000). Computed tomography of cryogenic biological specimens based on X-ray microscopic images. *Ultramicroscopy*, *84*(3–4), 185–197.
Werdan, K., Heldt, H. W., & Milovancev, M. (1975). The role of pH in the regulation of carbon fixation in the chloroplast stroma. Studies on CO2 fixation in the light and dark. *Biochimica et Biophysica Acta*, *396*(2), 276–292.
Wertheim, J. O., Sanderson, M. J., Worobey, M., & Bjork, A. (2010). Relaxed molecular clocks, the bias-variance trade-off, and the quality of phylogenetic inference. *Systematic Biology*, *59*(1), 1–8.
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. *Proceedings of the National Academy of Sciences of the United States of America*, *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.
Wichard, T., Charrier, B., Mineur, F., Bothwell, J. H., Clerck, O. D., & Coates, J. C. (2015). The green seaweed Ulva: a model system to study morphogenesis. *Frontiers in Plant Science*, *6*, 72.
Wietrzynski, W., Schaffer, M., Tegunov, D., Albert, S., Kanazawa, A., Plitzko, J. M., Baumeister, W., & Engel, B. D. (2020). Charting the native architecture of Chlamydomonas thylakoid membranes with single-molecule precision. *Elife*, *9*, e53740.
Worden, A. Z., Lee, J.-H., Mock, T., Rouzé, P., Simmons, M. P., Aerts, A. L., Allen, A. E., Cuvelier, M. L., Derelle, E., Everett, M. V., Foulon, E., Grimwood, J., Gundlach, H., Henrissat, B., Napoli, C., McDonald, S. M., Parker, M. S., Rombauts, S., Salamov, A., Von Dassow, P., Badger, J. H., Coutinho, P. M., Demir, E., Dubchak, I., Gentemann, C., Eikrem, W., Gready, J. E., John, U., Lanier, W., Lindquist, E. A., Lucas, S., Mayer, K. F. X., Moreau, H., Not, F., Otillar, R., Panaud, O., Pangilinan, J., Paulsen, I., Piegu, B., Poliakov, A., Robbens, S.,

Schmutz, J., Toulza, E., Wyss, T., Zelensky, A., Zhou, K., Armbrust, E. V., Bhattacharya, D., Goodenough, U. W., Van de Peer, Y., & Grigoriev, I. V. (2009). Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes Micromonas. *Science*, *324*(5924), 268–272.

Wu, A., Hammer, G. L., Doherty, A., von Caemmerer, S., & Farquhar, G. D. (2019). Quantifying impacts of enhancing photosynthesis on crop yield. *Nature Plants*, *5*(4), 380–388.

Wu, T., Li, L., Jiang, X., Yang, Y., Song, Y., Chen, L., Xu, X., Shen, Y., & Gu, Y. (2019). Sequencing and comparative analysis of three Chlorella genomes provide insights into strain-specific adaptation to wastewater. *Scientific Reports*, *9*(1), 9514.

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. *Nature Communications*, *9*(1), 5076.

Xia, L.-Y., Jiang, Y.-L., Kong, W.-W., Sun, H., Li, W.-F., Chen, Y., & Zhou, C.-Z. (2020). Molecular basis for the assembly of RuBisCO assisted by the chaperone Raf1. *Nature Plants*, *6*(6), 708–717.

Xu, C., Jiao, C., Sun, H., Cai, X., Wang, X., Ge, C., Zheng, Y., Liu, W., Sun, X., Xu, Y., Deng, J., Zhang, Z., Huang, S., Dai, S., Mou, B., Wang, Q., Fei, Z., & Wang, Q. (2017). Draft genome of spinach and transcriptome diversity of 120 Spinacia accessions. *Nature Communications*, *8*, 15275.

Yaman, M., & Radek, R. (2005). Helicosporidium infection of the great European spruce bark beetle, Dendroctonus micans (Coleoptera: Scolytidae). *European Journal of Protistology*, *41*(3), 203–207.

Yamano, T., & Fukuzawa, H. (2020). Transformation of the Model Microalga Chlamydomonas reinhardtii Without Cell-Wall Removal. In S. Li, L. Chang, & J. Teissie (Eds.), *Electroporation Protocols: Microorganism, Mammalian System, and Nanodevice* (pp. 155–161). Springer US.

Yamano, T., Sato, E., Iguchi, H., Fukuda, Y., & Fukuzawa, H. (2015). Characterization of cooperative bicarbonate uptake into chloroplast stroma in the green alga *Chlamydomonas reinhardtii*. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(23), 7315–7320.

Yamano, T., Tsujikawa, T., Hatano, K., Ozawa, S.-I., Takahashi, Y., & Fukuzawa, H. (2010). Light and Low-CO2-Dependent LCIB–LCIC Complex Localization in the Chloroplast Supports the Carbon-Concentrating Mechanism in Chlamydomonas reinhardtii. *Plant & Cell Physiology*, *51*(9), 1453–1468.

Yang, F., Wei, Z., Long, C., & Long, L. (2023). Toxicological effects of oxybenzone on the growth and bacterial composition of Symbiodiniaceae. *Environmental Pollution* , *317*, 120807.

Yang, Y. W., Lai, K. N., Tai, P. Y., & Li, W. H. (1999). Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages. *Journal of Molecular Evolution*, *48*(5), 597–604.

Yang, Z., Ma, X., Wang, Q., Tian, X., Sun, J., Zhang, Z., Xiao, S., De Clerck, O., Leliaert, F., & Zhong, B. (2023). Phylotranscriptomics unveil a Paleoproterozoic-Mesoproterozoic origin and deep relationships of the Viridiplantae. *Nature Communications*, *14*(1), 5542.

Yoshitomi, T., Kaminaga, S., Sato, N., Toyoshima, M., Moriyama, T., & Yoshimoto, K. (2020). Formation of Spherical Palmelloid Colony with Enhanced Lipid Accumulation by Gel Encapsulation of Chlamydomonas debaryana NIES-2212. *Plant & Cell Physiology*, *61*(1), 158–168.

Young, J. N., Heureux, A. M. C., Sharwood, R. E., Rickaby, R. E. M., Morel, F. M. M., & Whitney, S. M. (2016). Large variation in the Rubisco kinetics of diatoms reveals diversity among their carbon-concentrating mechanisms. *Journal of Experimental Botany*, *67*(11), 3445–3456.

Zang, K., Wang, H., Hartl, F. U., & Hayer-Hartl, M. (2021). Scaffolding protein CcmM directs multiprotein phase separation in β-carboxysome biogenesis. *Nature Structural & Molecular Biology*, *28*(11), 909–922.

Zeng, L., Zhang, N., Zhang, Q., Endress, P. K., Huang, J., & Ma, H. (2017). Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. *The New Phytologist*, *214*(3), 1338–1354.

Zeskind, B. J., Jordan, C. D., Timp, W., Trapani, L., Waller, G., Horodincu, V., Ehrlich, D. J., & Matsudaira, P. (2007). Nucleic acid and protein mass mapping by live-cell deep-ultraviolet microscopy. *Nature Methods*, *4*(7), 567–569.

Zhan, Y., Marchand, C. H., Maes, A., Mauries, A., Sun, Y., Dhaliwal, J. S., Uniacke, J., Arragain, S., Jiang, H., Gold, N. D., Martin, V. J. J., Lemaire, S. D., & Zerges, W. (2018). Pyrenoid functions revealed by proteomics in Chlamydomonas reinhardtii. *PloS One*, *13*(2), e0185039.

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *Journal of Structural Biology*, *193*(1), 1–12.

Zheng, S. Q., Palovcak, E., Armache, J.-P., Verba, K. A., Cheng, Y., & Agard, D. A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*, *14*(4), 331–332.

Zhou, Y., Gunn, L. H., Birch, R., Andersson, I., & Whitney, S. M. (2023). Grafting Rhodobacter sphaeroides with red algae Rubisco to accelerate catalysis and plant growth. *Nature Plants*, *9*(6), 978–986.

Zhu, X.-G., Long, S. P., & Ort, D. R. (2008). What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Curr. Opin. Biotechnol.*, *19*(2), 153–159.

Zimmerman, S. B., & Trach, S. O. (1991). Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *Journal of Molecular Biology*, *222*(3), 599–620.

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. N., & Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology*, *430*(15), 2237–2243.

Zinssmeister, C., Keupp, H., Tischendorf, G., Kaulbars, F., & Gottschling, M. (2013). Ultrastructure of calcareous dinophytes (Thoracosphaeraceae, Peridiniales) with a focus on vacuolar crystal-like particles. *PloS One*, *8*(1), e54038.

Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E., & Scheres, S. H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *ELife*, *7*. https://doi.org/10.7554/eLife.42166

8. APPENDIX

8.1. CsLinker sequences

- Where underlined amino acids (*e.g.,* MQAL) indicate the predicted chloroplast transit peptide.
- And bold nucleotide sequences (*e.g.,* **ATGCGC**) indicate introns.

>**CsLinker**_protein_CSI2_123000012064
MQALSARVTLGRPATARVAKKQQRPLTVAMATAQLSAEQKAFLERKARESRSPAPVETRGRSAPRPRS
ASPVGRPSSRTPAPASYSAPAASYSSAPANGLSAEQREFMERKARESRSPAPVETRGRSAPRPRSASP
RPTSRTPAPASYSAPAASYTAPAAAPAGGLSAEQREFLERKARESRSPAPVETRGRSAPRPRSASPRP
TSRTPAPASYSAPAASYSAPAAAPAGGLSAEQREFMERKARESRSFSAPVETRGRSASARPSARPSSR
TPTPVSNSGVRSAMSSGSSAGLSAEQREFLERKARESRSPAPAAASQSSSYGSAASGLSAEQREFLER
KARESRSFSAPAPTRGRSSTPRRR

>**CsLinker**_fullCDS_CSI2_123000012064
ATGCAGGCCCTGAGCGCACGCGTGACCCTCGGCCGCCCGGCCACCGCCCGGGTGGCCAAGAAGCAGCA
GCGCCCCCTGACTGTGGCCATGGCCACTGCCCAGCTGAGCGCCGAGCAGAAGGCATTCCTGGAGCGCA
AGGCGCGCGAGTCCCGCAGCCCCGCCCCCGTGGAGACCCGCGGCCGCAGCGCACCCCGCCCCCGCTCC
GCTTCCCCTGTGGGCCGCCCCTCCAGCCGCACCCCCGCCCCCGCGTCCTACAGCGCCCCCGCTGCCTC
CTACAGCAGCGCCCCCGCTAACGGCCTGAGCGCTGAGCAGCGCGAGTTCATGGAGCGCAAGGCCCGTG
AGAGCCGCAGCCCCGCCCCCGTGGAGACCCGCGGCCGCAGCGCGCCCCGCCCCCGCTCCGCGTCCCCC
CGCCCCACCAGCCGCACCCCGGCCCCCGCCTCCTACAGCGCCCCCGCCGCTTCCTACACCGCCCCCGC
CGCTGCCCCCGCTGGCGGCCTGAGCGCCGAGCAGCGCGAGTTCCTGGAGCGCAAGGCTCGTGAGAGCC
GCAGCCCTGCCCCCGTGGAGACCCGCGGCCGCAGCGCGCCCCGCCCCCGCTCCGCGTCCCCCCGCCCC
ACCAGCCGCACCCCGGCCCCCGCGTCCTACAGCGCTCCTGCTGCCTCCTACAGCGCCCCCGCCGCTGC
CCCCGCTGGCGGCCTGAGCGCTGAGCAGCGCGAGTTCATGGAGCGGAAGGCGCGCGAGTCCCGCAGCT
TCAGCGCCCCCGTGGAGACCCGCGGCCGCAGCGCGTCTGCCCGCCCTTCGGCCCGCCCCTCCAGCCGC
ACCCCGACCCCCGTGTCCAACAGCGGCGTGCGCTCCGCCATGAGCTCTGGCTCCTCCGCGGGCCTGTC
CGCTGAGCAGCGCGAGTTCCTGGAGCGCAAGGCCCGTGAGAGCCGGTCCCCCGCCCCCGCCGCCGCCT
CCCAGTCGTCCAGCTACGGCTCCGCGGCGTCTGGCCTGAGCGCCGAGCAGCGCGAGTTCCTGGAGCGC
AAGGCGCGCGAGTCCCGCAGCTTCAGCGCCCCCGCTCCCACCCGCGGCCGCAGCTCCACCCCCCGCCG
CCGCTAA

>CsLinker_fullgene_CSI2_123000012064
CTCCAGAATGGGGTAGGCTCGGGGAGGGCGCGGCGGCGTGCGAGGAGATGCTGGGTGTGCTGTGCCAC
ATGGGTGGGAGACTTCTGCCCGGCTCCCTCATCTGCCCGCGCCGCACCGCCCCGATCACCACTCCATT
CTTTCTGGGCACTGTAGGGCTTGCTGGCCGCCACTCTTGCACTCCTAGCAAGCCAAGCGTAACG**ATGC**
AGGCCCTGAGCGCACGCGTGACCCTCGGCCGCCCGGCCACCGCCCGGGTGGCCAAGAAGCAGCAGGTG
CGCCCGGGCAGTGCGGCAATGCGCCAGCGGGGCGTCCAGGCGGCCGGCTGGGGCTGGGTATAGTAGTG
GGCCGGGCCAGCCCCTCGGTGCAGCGCCCAAGGGCCGGCGGCAGGCAGCGCATCGCGTTTGTGAGACA
GTTCGACGCCTGATCACGGCGGCAAAGAGGCGGTCGGACACAGCTTTGGGGGCACCCGCGCTGCCGCC
GCCCCGAGGAGTCTACCCCGTCGCTGACCGTCTGCCGCCCGCATCCCCGCCGCCTGCAG**CGCCCCCTG**
ACTGTGGCCATGGCCACTGCCCAGCTGAGCGCCGAGCAGAAGGCATTCCTGGAGCGCAAGGCGCGCGA
GTCCCGCAGCCCCGCCCCCGTGGAGGTGCGATTCAAGGCCTGATTCTTTTGGCGGTGCAGCACCCAGC
GCATGCACGGCGATTCCGCACATGCGCAGCCACCTGCAGCCCACAGCTGCATGCATGCACAGCTTTTG
ACGTCTGCATCCCTCCCATTATTGTGCCATGTGCAG**ACCCGCGGCCGCAGCGCACCCCGCCCCCGCTC**
CGCTTCCCCTGTGGGCCGCCCCTCCAGCCGCACCCCCGCCCCCGCGTCCTACAGCGCCCCCGCTGCCT
CCTACAGCAGCGCCCCCGCTAACGGCCTGAGCGCTGAGCAGCGCGAGTTCATGGAGCGCAAGGCCCGT
GAGAGCCGCAGCCCCGCCCCCGTGGAGGTGAGAGGCTGAGGACAGTGGCTGGCATGCACCCTGGATCA
GCCCAGCACACGGTGCAAGGGTCGGTGCTCCGTGCAGTTCCTTCTGGCACCTGTGGGCAGCAGCAATG
CACATCCAGCATTGCCTGGATGCGTGCAGCAAGCAGAAATTGCAATGGGGCATGCGTGGGAGATACTA
CTCTGCCAGCGGCATCTCCCTGCATCAGCCCCTATCCGCTCGGGCAGCATGCCAGCCACCCACCCTGG
CTGAATTTAACCTGCCGTTCCCGCCATCAATCCCTCACTCCTGCCCGTTTCCACTGCCTTTTCTCCCT
GCAG**ACCCGCGGCCGCAGCGCGCCCCGCCCCCGCTCCGCGTCCCCCCGCCCCACCAGCCGCACCCCGG**

CCCCCGCCTCCTACAGCGCCCCCGCCGCTTCCTACACCGCCCCCGCCGCTGCCCCCGCTGGCGGCCTG
AGCGCCGAGCAGCGCGAGTTCCTGGAGCGCAAGGCTCGTGAGAGCCGCAGCCCTGCCCCCGTGGAGGT
GCGCCTGGGTCTTGGCTGGGTGCCAGCACATTTTTTTGAGGGATCAGCCTGAGGGCTGGCCAGGCCGT
GTTTGCGCCCCTTGAAGATGTACCGCACTTGCAGCCTGCACTCTTGCTTCGCCAAAGCGTCCTGAAAT
AGCCTGCAAATCTCACCGTTCCTTGCCTCTTGTGCGTCCCTCTACCCTGCAG**ACCCGCGGCCGCAGCG**
CGCCCCGCCCCCGCTCCGCGTCCCCCCGCCCCACCAGCCGCACCCCGGCCCCCGCGTCCTACAGCGCT
CCTGCTGCCTCCTACAGCGCCCCCGCCGCTGCCCCCGCTGGCGGCCTGAGCGCTGAGCAGCGCGAGTT
CATGGAGCGGAAGGCGCGCGAGTCCCGCAGCTTCAGCGCCCCCGTGGAGGTGAGAGAAAGTTCTGGGG
TTGGGTTCGATTCTCTGCGGCGCATCTGCGTGGCATTTGCAAGCAAGATGCAAGATTTCTCGCATGCA
AGGCCAATCAGAGGCGTAGCCATCACGCCTAGCACACATCCTGACCCATGTTGCAACTTCCCACCTCC
TCCCTTTGCCATCTGCAG**ACCCGCGGCCGCAGCGCGTCTGCCCGCCCTTCGGCCCGCCCCTCCAGCCG**
CACCCCGACCCCCGTGTCCAACAGCGGCGTGCGCTCCGCCATGAGCTCTGGCTCCTCCGCGGGCCTGT
CCGCTGAGCAGCGCGAGTTCCTGGAGCGCAAGGCCCGTGAGGTGAGACCGAGCTGCATGTTCTCTCAA
GGGAGCAATGGGGTGGAATGAGCAGAACGCCCTGCCTGGGGTGGAACTCAGCTGCATGCCACACGCGA
GTCCCGACACGCAGCCAGCGGGCATGTTCCCCTGCATGCCTCAAATCCAGCAACTGGCCTTTCACTTA
CTTACGACCCCGCCCAACCTGCCAATTTTCCCTTCCTGCAG**AGCCGGTCCCCCGCCCCCGCCGCCGCC**
TCCCAGTCGTCCAGCTACGGCTCCGCGGCGTCTGGCCTGAGCGCCGAGCAGCGCGAGTTCCTGGAGCG
CAAGGCGCGCGAGTCCCGCAGCTTCAGCGCCCCCGCTCCCACCCGCGGCCGCAGCTCCACCCCCCGCC
GCCGCTAAGCAGCTGAGCTAAGCAGCAGCTTTGGCCCCTGTGGCACTACATATGCCGGGCCGTCCGGC
CCGCAGCGGCGGCAGCCCCTGCCCAACGCAGCAGCGGCTGCGCCACAGCTGCAGTGCCGGCGGCGCGC
CGGCCTGTGTGTGTGATGTTGTGAAATGTTGACAGTGCGTCCCGGCAAGCTTCCCTCTGCTCTGCCCC
CGTC

8.2. Other linker sequences

>**EPYC1**_protein_Cre10.g436550
MATISSMRVGAASRVVVSGRVKTVKVAARGSWRESSTATVQASRASSATNRVSPTRSVLPANWRQELE
SLRNGNGSSSAASSAPAPARSSSASWRDAAPASSAPARSSSASKKAVTPSRSALPSNWKQELESLRSS
SPAPASSAPAPARSSSASWRDAAPASSAPARSSSSKKAVTPSRSALPSNWKQELESLRSSSPAPASSA
PAPARSSSASWRDAAPASSAPARSSSASKKAVTPSRSALPSNWKQELESLRSNSPAPASSAPAPARSS
SASWRDAPASSSSSSADKAGTNPWTGKSKPEIKRTALPADWRKGL

>**PYCO1**_protein_Phatr_J49957
MKISVKSATLALLMVPTTGFLHAHPARSVETAFHAPTHTTWKTGARWGRTPRSGSSGFSPQSAGSYAG
RSAFGRDSTYGSSTSVGSATAPTAPGYSSMPAKVYANAGPNQKYSMTKWSPQNGASVNGGSPSAYSSS
NGVPAGGNGAIGTGYNPSAQSNNAYQSSAPATGSAAPTYSSMPGQAYAGSGPPKNYSMVKWSPRGGSS
RASNGGGATGGSIGTGYNPAQPASAASSNDAYTAPSTGSGANGAPSYSSMPGQAYAGSGPRKNYSMVK
WSPRGGSSRASNGGGGGSPTYSSLPGQAYSGSGPGKNYSMVKWSPQGGSSVSSNRGGGASAGSLGTGY
NPAQPTTAGGASSNESYSAPSTGGSSSGSTPSYSSMPGQAYAGSGPSKNYSMVKWSPRGGSSRASNGG
GATGGSIGTGYNPAQAAPAASSYGSSPSSSNSNAYSAPTGNAAPSYSSMPGQTNTGSGPTKRYAMVKW
SPRGGSVVSFNRVGGGGGSMGTGYNPAQAAAPPAAAAASYSPAPPAPASSGYGNLASEWASMNTEKAS
PAPASSAYGGGAPSPAPAPAPAPRSGNAYGNLAAEWGSMNQGDSSAAF

> **Porphyridium**_linker_protein_KAA8498780
MNDSVESMMMSGVAFMPGSAVPLGGGRYIVKGPASRTVRYDKPTRDPNPSFIDPNDPKAKQKAIHTAP
KSFADYLAERARKEGTAPPAAAAAAPAQNGNGYKPSSVQVAPSAPKPAGGAMSPYEEYMAKRNGGDSA
SAASPAPPAPAAKAGGAMSPYEEYLQKRRAQESGSVPAYTAPAAASPLPSQSRAAPAAAKELSPYEEY
MRKRASQSQAPAAVPQPAATSAAPVKKEAELSPYEEYMRKRNGGGSVTGSTAPPAPKPYTPPAAPKAP
ASSGSATLYEEYMRARQQKESGRSASSPAASYSPPAASYTPPASPASPQPVTGNLFEDYMKRRAEKEY
ARPATGTAPAPSAPRAFAPPAAPSGGSLFESYMRERAAKEAGSSQGTNAPGTRAAPAPTKSSASETSL
FDSYMRERAAKEGGAGSSSSSSSKPVSPTAAPAAAPVAAPKSAPKLAAVSGGSLFESYMAERKKKEAA
GGGTTSASSTPSVPRAAPPASSSTSSSSSSSPSFSAPSFSKPSFSTPSMPKLFDFLILEKAAKSFASE
MQKSMTPSTSSTSKPYKASSSSTAGDGSGVAEIHQALLLILVAMAIGVLAVYTSSGLSDFRP

> **Ulva**_linker_protein_UM120_0037
MLSLRAPVSSARRAVVLNARRPQPDYAARRSGGREARRSGAATNRQPIQRRASVDYSSRRSGGEEENN
RRRSGGYSSPTPSRSSSYSPPQRNSSPAPANRLVSQAVLKRRSVDYAARRSGGREAQRSGSSPPARRA
SSSSSYSAPERNTSPAPANRLISQAVLKRRSVDYAARRSGGREAQRSGSSGRPSSSSSYSSPQRSSSP
APADRLVSQAVLKRRSVDYAARRSGGRAGSNDSRRRSSSFSSPTPSRSSSPAPAPANRLVSPAVRKRR
SVDYSARRSGSQGRRNSRR

8.3. PMST and PMTT sequences

>Chlorella_sorokiniana_UTEX1230_**PMST**_ CSI2_123000007310
MRAAVAAPSAARATMPRLAPVSARPSARPAIRRAPVRRASYHTAVVQASHTSDGADTMAVTFKLNRKV
NFGQELVLVGSTDSLGGWELARAPHMTWNEGDLWTATIELPAGTHVEYKFALWDPHSPPVWENCPNRA
FSVPAATVVTCAWDEPGATTCEPALEDGSSDSIQDVPAPAFRPATVVASYEANGHVTVEGRTTTAVKN
GKASATDGAYSAEQLDFLKRKLETTHAAPSAVRQPAAYPNGHANGHANGSAEYTPEQLEFLRRSGKLN
GSASPAPSQPARGASPAAASSAGGAQYTPEQLAFLQRSGKSPSPAPAVRRSSSPRPPAVRRSSSPPPP
AARVNGAAAGSSAQYSPEQLAFLQRSGKTPSPPPAAARQQSAQPARGSASPAPSAAATQYTPEQLEFL
RRSGKLAGSSSSPRPAAQRSSASPAPGQYSPEQLEFLRRSGKL

>Chlorella_sorokiniana_UTEX1230_**PMTT**_ CSI2_123000003237
MQALAQRPSRPAGVLGRSARRLRRPARPGAAQARAQRSKTEPEADHGAGGWRSLLGAAALAVILKEAP
MAIAAPADQALASPAAETAETRDGGSTLFTALAAGGVVGVALVAARSFAKGDDSEKAAMEAQVAAARA
RLPPPERPDPARSPSPPPAPAGSASPPPAPAGSASPPPAPAGSASPAPAPVRSPSPASAARLAVAAER
AEELQDQARQMRTMLEEVSRSLAAVRQARQVAQEQLEAAAEAGLPLQLTQEELEAAELRAQQEAEAEA
AWVEADARFQDVLDDLEAEQRSVNQTQGVEADRLKARFLGILDELHAERERAAILQASHYKWLEEERE
KAAAAASAAARGRSPSPGPARSLSPAPPRPASPAPAAAAPGSAAAQFGAGRAASPMPSPAPAAAAGGL
SPEQAAFLERKRTASTSPAPPAVRPRSSSPRPSSRQPTASPRPTSRQPSMSPPPSAAAAAEPAGLSAE
QLEFMERKRAASTSPSPAPSPAAAAGSAASAGLSAEQLEFMERKRAASSSPAPAAVRPRSSSPRPSSR
QPTVSPRPSSRQPTVSPRPTSRQASMSPPPASAAAAAPAGLSPEQAAFLERKRTASTSPPPATSPSPA
AAPAGLSAEQAAFLERKRAASTSPPPMAAEQGGNGAAPMQRPSFRPSSRSPSPVARRAAQSQPPAASY
SPPPPSQQQAWEPAQQAQHTQQAAAPAGRQRTWQDPEFAAQTLAAFPEKSVASVEEARCLIEKGGYIF
LDVRPPIELEEEGRFAGATNIPLVHAERVFDAASGKRELEEWTNEDFVDQVKWRFPDLNTPLLVGDAD
GARLAVEALELLDDAGYTCLVGLKGGFAAWTRVFDSKGARRGASETVTA*

>Micractinium_conductrix_SAG 241.80_**PMTT**_rna-C2E20_4272
MNVAVVSAFNSPAVQLRQPRCVRVAGGGTAAAASGRAGVCRAAQQQAQPAAAQQQQQDQKLAAMATSR
RRVLALPAGLAAAAALAAGAGPAAAIVIPPPGYRYHVDKLDGYSFFYPADWSVVTTSGNDVQYRNPFN
VEENLFVNVSSPSSSKYESVADLGSPMDAAKKLQSQFLEEFMSTRLGVKRTAEVVNAEERTGADGQLY
YDIATRVKSFASRNQLAVTQAEIEEGVVLEWDRVYLTVLGVAGRRLYEFRLQAANSVYEADPERLLTV
EPMQALQQVASARPLVAGRPRARLPSGRRPQVLVAALSPEQQAFLDRKAREARSSSPVPAPTRGRSAS
SRPASVRPTSRPAPAAAPAPRYSSSAAAPAGSLSPEQQEFLARRAQEGRSASPAPVPTRGRSGSVRPS
SRTPSPVRPAASGHSAPAAAPAAGLSPEQQEFLARRAQEGRSASPAGAPTRGRSASARPASVRPTSRP
APVAAPAPSYSSSAAPSAGLTPEQQEFLARRAQEARSASPAPAPTRGRSGSVRPSSRTPSPVRPAAPA
APAASPTRGRSGSVRPASLRPSSRTPAPVRPASAAAPSSYAAPASSSAAAPAGGLSPEQQEFLARRQQ
ESRSPPPAAAPTRGRSGSVRPSSRTPSPVRSAAASAPPPAQAEPRVGDVDFARAALERTLQTTPAVHP
RDQPQTRAPPAAAPAPTGPSTGRWPDPAFFQATLQAFPAKAVCDAEEARCLMDAGYAVLDVRSELESG
SKGKLRGSVMIPCVFQRQQFDAAAGQMVIDEQDNPGFLAQVARKFPRKDAPIIVMCGNGKQFSIDVLE
ALEEEGYVNLVGMQGGYTNFTRTFDVKLNPRDRRPVPRRQLKKLI

>Chlorella_vulgaris_211/11P_**PMTT**_KAI3429225
MPSLPKPTPLQRERLLLLLLGASAAAVLPAIWRRLRSLRRKRGLPPVPTPAAAARSTTSSAGGGADAY
ETRKAVDEYLQFHFGRPEDIIPYDIGPKEALRFTEQLAALCERHCSALRDFTGEGGEATAVDIGCAVG
GASFHLARAFPHVLGIDFSQHFVNAANTMRERGWMRYTAAEEGELSKENTAVVPEDIDRSRVRFQQAD
ACNLPGNLGPVDAVLAANLLCRLPDPTLFLGRMHSFIKRDGVLVLVSPYSWLPGWTPRDKWLGGFLDQ
EGKPVWTADVLKKLLGEHFTLVEEAELPFVIREHRRKFQWGVSHAMAPVVSRRADATRAPRHQRLRVQ
VQASLSAEQQAFLERRRSESRSPAAAPTRGRSTVLRSPVRPASSTPARGRSSAPAPAQSNGLTAEQQA
FLERRRSESRSPAPAPTRGRSTVPRSPVRPSSRAPSARPAASSPASYSPPSSAPAGGLSAEQKAFMDR
RRSESRPAASAPTRGRSTVPRRPITSSSAPRRASSSAPSASYSAPSSAPAGGLSAEQQAFLERHRSES
RSAAPAPTRGRSTVPRSPVRPTSRAPSARPAASSPTSYSPPSSAPAGGLSAEQQAFMDRRRSESRSPA
PAPTRGRSAVTRSPVRPTSRAPAAPAARTPSVRPSARPTSRPASGRPSSRPAAHRSGASSSPASSSGL
SAEQQAFLDRKRSGGR

8.4. CsLinker purification products

- Where bold italics (*e.g.,* ***ENLYFQ***) indicate scar amino acids left following cleavage with TEV protease.
- And bold red (*e.g.,* **DD**) indicate mutated amino acids.

>pLM872_CsLinker_protein_product
SNAMATAQLSAEQKAFLERKARESRSPAPVETRGRSAPRPRSASPVGRPSSRTPAPASYSAPAASYSS
APANGLSAEQREFMERKARESRSPAPVETRGRSAPRPRSASPRPTSRTPAPASYSAPAASYTAPAAAP
AGGLSAEQREFLERKARESRSPAPVETRGRSAPRPRSASPRPTSRTPAPASYSAPAASYSAPAAAPAG
GLSAEQREFMERKARESRSFSAPVETRGRSASARPSARPSSRTPTPVSNSGVRSAMSSGSSAGLSAEQ
REFLERKARESRSPAPAAASQSSSYGSAASGLSAEQREFLERKARESRSFSAPAPTRGRSSTPRRR***EN***
LYFQ

>pLM712_mEGFP-CsLinker_protein_product
MGSSHHHHHHGSSVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPV
PWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRI
ELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDG
PVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGIEENLYFQSNAMATAQLSAE
QKAFLERKARESRSPAPVETRGRSAPRPRSASPVGRPSSRTPAPASYSAPAASYSSAPANGLSAEQRE
FMERKARESRSPAPVETRGRSAPRPRSASPRPTSRTPAPASYSAPAASYTAPAAAPAGGLSAEQREFL
ERKARESRSPAPVETRGRSAPRPRSASPRPTSRTPAPASYSAPAASYSAPAAAPAGGLSAEQREFMER
KARESRSFSAPVETRGRSASARPSARPSSRTPTPVSNSGVRSAMSSGSSAGLSAEQREFLERKARESR
SPAPAAASQSSSYGSAASGLSAEQREFLERKARESRSFSAPAPTRGRSSTPRRR

>pLM1286_α3_protein_product
STPAPASYSAPAASYTAPAAAPAGGLSAEQREFLERKARESRSPAPVETRGRSAPRPRSASPRPTSR***E***
NLYFQ

>pLM1287_α3-α4_protein_product
STPAPASYSAPAASYTAPAAAPAGGLSAEQREFLERKARESRSPAPVETRGRSAPRPRSASPRPTSRT
PAPASYSAPAASYSAPAAAPAGGLSAEQREFMERKARESRSFSAPVETRGRSASARPSARPSSR***ENLY***
FQ

>pLM1454_α3-α4-hydrophobic-SDM_protein_product
STPAPASYSAPAASYTAPAAAPAGGLSAEQRE**DD**ERKARESRSPAPVETRGRSAPRPRSASPRPTSRT
PAPASYSAPAASYSAPAAAPAGGLSAEQRE**DD**ERKARESRSFSAPVETRGRSASARPSARPSSR***ENLY***
FQ

>pLM1455_α3-α4-electrostatic-SDM_protein_product
STPAPASYSAPAASYTAPAAAPAGGLSA**A**QREFLE**A**KARESRSPAPVETRGRSAPRPRSASPRPTSRT
PAPASYSAPAASYSAPAAAPAGGLSA**A**QREFME**A**KARESRSFSAPVETRGRSASARPSARPSSR***ENLY***
FQ

>pLM1461_α3-α4-partial-electrostatic-SDM_protein_product
STPAPASYSAPAASYTAPAAAPAGGLSA**A**QREFLE**A**KARESRSPAPVETRGRSAPRPRSASPRPTSRT
PAPASYSAPAASYSAPAAAPAGGLSA**A**QREFMERKARESRSFSAPVETRGRSASARPSARPSSR***ENLY***
FQ

8.5. Motifs from Chlorella linkers

- Used for the creation of the sequence logo in Fig. 2-15.

```
AEQKAFLERKARES
AEQREFMERKARES
AEQREFMERKARES
AEQREFLERKARES
AEQREFLERKAXES
AEQKAFLDRKARES
AEQVAFMERKARES
AEQREFMERKARES
KEQQEFMARKAQES
PEQEAFMQRKAAEQ
AEQEAFMQRKAAEV
AEQEAFMQRKAAEQ
AEQAAFVQRKAAEQ
KEQAQFLERKRAGG
EEQRQFLERKARES
AEQAAFLERKARES
KEQEEFMQRKARES
SEQLAFLERKKVES
AEQAAFVQRKAAEQ
PEQEAFMQRKAAEV
AEQAAFVQRKAVEQ
AEQAAFVQRKAAEQ
KEQAEFLERKRAGG
AEQKAFLERKARES
AEQREFMERKARES
AEQREFLERKARES
AEQREFMERKARES
AEQREFLERKARES
AEQREFLERKARES
PEQQAFLDRKRQSG
AEQQAFLARKAQES
AEQQEFLARKAQES
AEQQAFLARKRSES
AEQQAFLARKVQES
AEQQAFLARKASGQ
AEQQAFLDRKRNET
AEQQAFLERRRSES
AEQQAFLDRRRSES
AEQQAFLERRRSES
AEQQAFLDRRRSES
AEQQAFLDRKRNGG
ADQQAFLERKRRES
ADQQEFLERKRRET
AEQREFLERKRRES
AEQREFLERKRRES
AEQQEFLRRKASES
AEQQEFLRRKASEP
AEQRAFLERKAAET
AEQMAFLERKRSGG
```

8.6. RbcL Sequences of

>**Chlorella_**RbcL
MAPQTETRAGAGFKAGVKDYRLTYYTPDYQPKDTDILAAFRMTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYDIEPVPGEENQYIAYIAYPLDLFEEGSVTNLFTSIVGNVFGFKALRALRLE
DLRIPPAYVKTFQGPPHGIQVERDKLNKYGRGLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFVAEAIYKSQAETGEIKGHYLNATAATAEEMLKRAECAKDLGVPIIMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQRNHGIHFRVLAKALRLSGGDHLHSGTVVGKLEGERE
VTLGFVDLMRDDYIEKDRSRGIYFTQDWVSLPGTMPVASGGIHVWHMPALVEIFGDDACLQFGGGTLG
HPWGNAPGAAANRVALEACTQARNEGRDLAREGGDVIRAACKWSPELAAACEVWKEIKFEFETIDTL

>**Chlamydomonas_**RbcL
MVPQTETKAGAGFKAGVKDYRLTYYTPDYVVRDTDILAAFRMTPQLGVPPEECGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYDIEPVPGEDNQYIAYVAYPIDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPPAYVKTFVGPPHGIQVERDKLNKYGRGLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFVAEAIYKAQAETGEVKGHYLNATAGTCEEMMKRAVCAKELGVPIIMHDYLTG
GFTANTSLAIYCRDNGLLLHIHRAMHAVIDRQRNHGIHFRVLAKALRMSGGDHLHSGTVVGKLEGERE
VTLGFVDLMRDDYVEKDRSRGIYFTQDWCSMPGVMPVASGGIHVWHMPALVEIFGDDACLQFGGGTLG
HPWGNAPGAAANRVALEACTQARNEGRDLAREGGDVIRSACKWSPELAAACEVWKEIKFEFDTIDKL

>**Chlamydomonas_D86H**_RbcL
MVPQTETKAGAGFKAGVKDYRLTYYTPDYVVRDTDILAAFRMTPQLGVPPEECGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYDIEPVPGEDNQYIAYVAYPIDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPPAYVKTFVGPPHGIQVERDKLNKYGRGLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFVAEAIYKAQAETGEVKGHYLNATAGTCEEMMKRAVCAKELGVPIIMHDYLTG
GFTANTSLAIYCRDNGLLLHIHRAMHAVIDRQRNHGIHFRVLAKALRMSGGDHLHSGTVVGKLEGERE
VTLGFVDLMRDDYVEKDRSRGIYFTQDWCSMPGVMPVASGGIHVWHMPALVEIFGDDACLQFGGGTLG
HPWGNAPGAAANRVALEACTQARNEGRDLAREGGDVIRSACKWSPELAAACEVWKEIKFEFDTIDKL

>**Ulva**_RbcL
MAPQTETKAGTGFKAGVKDYRLTYYTPDYQVKDTDILAAFRMTPQPGVPAEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYDIEPLGEDDQYIAYIAYPLDLFEEGSVTNLFTSIVGNVFGFKALRALRLED
LRIPPAYVKTFQGPPHGIQVERDKLNKYGRGLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDEN
VNSQPFMRWRDRFLFVAEAIYKSQSETGEVKGHYLNATAGTCEEMMERGQFAKDLGVPIVMHDYITGG
FTANTTLAHFCRASGLLLHIHRAMHAVIDRQRNHGIHFRVLAKILRMSGGDHLHSGTVVGKLEGEREI
TLGFVDLMRDDYIEKDRSRGIYFTQDWVSLPGTMPVASGGIHVWHMPALVEIFGDDACLQFGGGTLGH
PWGNAPGAAANRVALEACTQARNEGRDLASEGGDVIRAACKWSPELAAACEVWKEIKFEFDAIDTL

>**Fern**_RbcL
MSPQTETKASVGFKAGVKDYRLTYYTPEYVTLDTDILAAFRMTPQPGVPAEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYDIEPVAGEENQYIAYVAYPLDLFEEGSVTNMLTSIVGNVFGFKALRALRLE
DLRIPPAYSKTFLGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFVAEALFKSQAETGEIKGHYLNATAGTCEEMMKRAIFARELGAPIVMHDYLTG
GFTANTSLAFYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKGLRMSGGDHIHAGTVVGKLEGERE
VTLGFVDLLRDDYIEKDRSRGIYFTQDWVSMPGVFPVASGGIHVWHMPALTEIFGDDAVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAREGNEIIREASKWSPELAAACE

8.7. Consensus RbcL Sequences

>**Monocot**_RbcL_n=31,639
MSPQTETKASVGFKAGVKDYKLTYYTPDYETKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYHIEPVVGEENQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPPAYSKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFCAEAIYKAQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHAGTVVGKLEGERE
MTLGFVDLLRDDFIEKDRSRGIFFTQDWVSMPGVIPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAREGNEIIREACKWSPELAAACEVWKAIKFEFEPVDKL

>**Eudicot**_RbcL_n=82,622
MSPQTETKASVGFKAGVKDYKLTYYTPEYETKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYHIEPVAGEENQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPPAYVKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFCAEAIYKAQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHAGTVVGKLEGERE
ITLGFVDLLRDDFIEKDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAREGNEIIREASKWSPELAAACEVWKEIKFEFEAMDTL

>**Gynosperm**_RbcL_n=4,785
MSPKTETKASVGFKAGVKDYRLTYYTPEYQTKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYDIEPVPGEETQFIAYVAYPLDLFEEGSVTNLFTSIVGNVFGFKALRALRLE
DLRIPPAYSKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFVFCAEAIYKAQAETGEIKGHYLNATAGTCEEMMKRAVFARELGVPIVMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQRNHGMHFRVLAKALRMSGGDHIHAGTVVGKLEGERD
VTLGFVDLLRDDFIEKDRSRGIYFTQDWVSMPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAREGNEVIREASKWSPELAAACEIWKEIKFEFDTIDRLD

>**Fern**_RbcL_n=11,785
MSPQTETKAGVGFKAGVKDYRLTYYTPEYKTKDTDILAAFRMTPQPGVPAEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYDIEPVAGEENQYIAYVAYPLDLFEEGSVTNLFTSIVGNVFGFKALRALRLE
DLRIPPAYSKTFIGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFVAEALFKSQAETGEIKGHYLNATAGTCEEMMKRAVFARELGAPIVMHDYLTG
GFTANTSLAFYCRDNGLLLHIHRAMHAVIDRQRNHGMHFRVLAKALRMSGGDHIHAGTVVGKLEGERE
VTLGFVDLLRDDYIEKDRSRGIYFTQDWVSMPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAREGNEIIREASKWSPELAAACEIWKAIKFEFDTIDTL

>**Lycophyte**_RbcL_n=1,383
MSPQTETKASVGFKAGVKDYRLTYYTPDYKTKDTDILAAFRMTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYDIEPVAGEKDQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPPAYSKTFIGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFVFVAEALYKAQAETGEIKGHYLNATAGTCEEMMKRAEFARELGVPIVMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGIHFRVLAKALRMSGGDHIHAGTVVGKLEGERQ
VTLGFVDLLRDDYIEKDRSRGIYFTQDWVSMPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLALEGNEIIREASKWSPELAAACEVWKEIKFEFETIDTL

>**Bryophyte**_RbcL_n=2,615
MSPQTETKAGVGFKAGVKDYRLTYYTPDYQTKETDILAAFRMTPQPGVPAEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYDIEAVPGEENQYIAYVAYPLDLFEEGSVTNLFTSIVGNVFGFKALRALRLE
DLRIPPAYSKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFVAEAIYKSQAETGEIKGHYLNATAGTCEEMMKRAQFARELGMPIVMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRLSGGDHIHAGTVVGKLEGERQ
VTLGFVDLLRDDYIEKDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAREGNEIIREAAKWSPELAAACEVWKEIKFEFETIDTV

>**Charophyte**_RbcL_n=1,347
MSPQTETKAGAGFKAGVKDYRLTYYTPDYETKDTDILAAFRMTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYDIEPVAGEENQYIAYVAYPLDLFEEGSVTNLFTSIVGNVFGFKALRALRLE
DLRIPPAYVKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFVAEAIYKAQAETGEIKGHYLNATAGTCEEMMKRAEYAKELGVPIIMHDYLTG
GFTANTSLSHYCRDNGLLLHIHRAMHAVIDRQKNHGIHFRVLAKALRMSGGDHIHSGTVVGKLEGERQ
VTLGFVDLLRDDYIEKDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAREGNEVIREAAKWSPELAAACEVWKEIKFEFXTIDTL

>**Other_Chlorophyte**_RbcL_n=149
MAPQTETKTGTGFKAGVKDYRLTYYTPDYQVKETDILAAFRMTPQPGVPPEECGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYDIEPVPGEDNQYICYVAYPIDLFEEGSVTNLFTSIVGNVFGFKALRALRLE
DLRIPVAYCKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFVAEAIYKSQAETGEIKGHYLNATAGTCEEMLKRAECAKELGVPIVMHDYLTG
GFTANTSLAIYCRDNGLLLHIHRAMHAVIDRQRNHGIHFRVLAKALRMSGGDHLHSGTVVGKLEGERE
VTLGFVDLMRDAYIEKDRSRGIYFTQDWCGLPGVMPVASGGIHVWHMPALVEIFGDDACLQFGGGTLG
HPWGNAPGAAANRVALEACVQARNEGRDLAREGGDXIRAACKWSPELAAACEVWKEIKFEFDTVDTL

>**Chlorophyceae**_RbcL_n=1,536
MVPQTETKAGAGFKAGVKDYRLTYYTPDYVVKDTDILAAFRMTPQPGVPPEECGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYDIEPVPGEDNQYIAYVAYPIDLFEEGSVTNLFTSIVGNVFGFKALRALRLE
DLRIPPAYVKTFQGPPHGIQVERDKLNKYGRGLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFVAEAIYKAQAETGEIKGHYLNATAGTCEEMLKRAQCAKELGVPIIMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQRNHGIHFRVLAKALRMSGGDHLHSGTVVGKLEGERE
VTLGFVDLMRDDYIEKDRSRGIYFTQDWCSMPGVMPVASGGIHVWHMPALVEIFGDDACLQFGGGTLG
HPWGNAPGAVANRVALEACTQARNEGRDLAREGGDVIRSACKWSPELAAACEVWKEIKFEFDTIDKL

>**Ulvophyceae**_RbcL_n=3,492
MAPQTETKAGAGFKAGVKDYRLTYYTPDYQVKDTDILAAFRMTPQPGVPAEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYDIEPVGEDNQYIAYVAYPLDLFEEGSVTNLFTSIVGNVFGFKALRALRLED
LRIPPAYVKTFQGPPHGIQVERDKLNKYGRGLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDEN
VNSQPFMRWRDRFLFVAEAIYKSQAETGEIKGHYLNATAGTCEEMMERGQFAKDLGVPIIMHDYITGG
FTANTSLAHFCRASGLLLHIHRAMHAVIDRQRNHGIHFRVLAKILRMSGGDHLHSGTVVGKLEGEREI
TLGFVDLMRDDYIEKDRSRGIYFTQDWVSLPGTMPVASGGIHVWHMPALVEIFGDDACLQFGGGTLGH
PWGNAPGAAANRVALEACTQARNEGRDLASEGGDVIRAACKWSPELAAACEVWKEIKFEFDTIDTL

>**Trebouxiophyceae**_RbcL_n=1,328
MAPQTETKAGAGFKAGVKDYRLTYYTPDYQVKETDILAAFRMTPQPGVPPEECGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYDIEPVAGEENQYIAYVAYPLDLFEEGSVTNLFTSIVGNVFGFKALRALRLE
DLRIPPAYVKTFQGPPHGIQVERDKLNKYGRSLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFVAEAIYKSQAETGEIKGHYLNATAGTCEEMLKRAECAKDLGVPIVMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQRNHGIHFRVLAKALRLSGGDHLHSGTVVGKLEGERE
VTLGFVDLMRDDYIEKDRSRGIYFTQDWVSLPGVMPVASGGIHVWHMPALVEIFGDDACLQFGGGTLG
HPWGNAPGAAANRVALEACTQARNEGRDLAREGGDVIRAACKWSPELAAACEVWKEIKFEFETIDTL

8.8. C_3 Crop RbcL Sequences

>**Wheat**_RbcL
MSPQTETKAGVGFKAGVKDYKLTYYTPEYETKDTDILAAFRVSPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYHIEPVAGEDSQWICYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPPTYSKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRACYECLRGGLDFTKDDE
NVNSQPFMRWRDRFVFCAEAIYKSQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTG
GFTANTTLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHSGTVVGKLEGERE
MTLGFVDLLRDDFIEKDRARGIFFTQDWVSMPGVIPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAAANRVALEACVQARNEGRDLAREGNEIIRAACKWSPELAAACEVWKAIKFEFEPVDTID
K

>**Barley**_RbcL
MSPQTETKAGVGFQAGVKDYKLTYYTPEYETKDTDILAAFRVSPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYHIEPVAGEDSQWICYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPPTYSKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRACYECLRGGLDFTKDDE
NVNSQPFMRWRDRFVFCAEAIYKSQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTG
GFTANTTLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHSGTVVGKLEGERE
MTLGFVDLLRDDFIEKDRARGIFFTQDWVSMPGVIPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAAANRVALEACVQARNEGRDLAREGNEIIRAACKWSPELAAACEVWKAIKFEFEPVDTID
KKV

>**Oat**_RbcL
MSPQTETKASVGFQAGVKDYKLTYYTPEYETKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYHIEPVAGEDNQWICYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPPAYTKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRACYECLRGGLDFTKDDE
NVNSQPFMRWRDRFVFCAEAIYKAQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYITG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHSGTVVGKLEGERE
MTLGFVDLLRDDFIEKDRARGIFFTQDWVSMPGVIPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAAANRVALEACVQARNEGRDLAREGNEIIREACKWSPELAAACEVWKAIKFEFEPVDTID
E

>**Rice**_RbcL
MSPQTETKASVGFKAGVKDYKLTYYTPEYETKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYHIEPVVGEDNQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPPTYSKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRACYECLRGGLDFTKDDE
NVNSQPFMRWRDRFVFCAEAIYKSQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHAGTVVGKLEGERE
MTLGFVDLLRDDFIEKDRARGIFFTQDWVSMPGVIPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAAANRVALEACVQARNEGRDLAREGNEIIRSACKWSPELAAACEIWKAIKFEFEPVDKLD
S

>**Plantain**_RbcL
MSCREGLMSPQTETKASVGFKAGVKDYKLNYYTPDYEVKDTDILAAFRVTPQPGVPPEEAGAAVAAES
STGTWTTVWTDGLTSLDRYKGRCYHIEAVVGEENQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKA
LRALRLEDLRIPTSYSKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGL
DFTKDDENVNSQPFMRWRDRFLFCTEALFKAQAETGEIKGHYLNATAGTCEEMMKRAICARELGVPIV
MHDYLTGGFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHAGTVVG
KLEGEREMTLGFVDLLRDDYIEKDRSRGIFFTQDWVSMPGVLPVASGGIHVWHMPALTEIFGDDSVLQ
FGGGTLGHPWGNAPGAVANRVALEACVQARNEGRDLAREGNEIIREASKWSPELAAACEVWKEIKFEF
EPVDKLDKEKK

>**Onion**_RbcL
MSPQTETKASVGFKAGVKDYRLTYYTPDYETKDTDILAAFRVTPQPGVPAEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYHIEAVVGEENQYIAYIAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPPAYSKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFIFCAEAIYKAQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIIMHDYITG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHAGTVVGKLEGERE
MTLGFVDLLRDDFIEKDRSRGIFFTQDWVSMPGVIPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAREGNEIIREACKWSPELAAACEVWKEIKFEFEPVDKID
KQK

>**Yam**_RbcL
MSPQTETKASVGFKAGVKDYKLTYYTPDYETKDTDILAAFRVSPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYHIESVVGEENQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPTSYSKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFCAEAIYKAQAETGEIKGHYLNATAGTCEEMIKRAVFARELGAPIVMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHAGTVVGKLEGERE
MTLGFVDLLRDDFIEKDRSRGIFFTQDWVSMPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAREGNEIIREACKWSPELAAACEVWKEIKFEFKPVDTLD
L

>**Grape**_RbcL
SVGFKAGVKDYKLTYYTPEYETKPTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTTVWTDGLTSL
DRYKGRCYHIEPVAGEESQFIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPPAYT
KTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMR
WRDRFLFCAEAIFKSQAETGEIKGHYLNATAGTCEEMMKRAIFARELGVPIVMHDYLTGGFTANTSLA
QYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHAGTVVGKLEGEREITLGFVDLL
RDDFVEKDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLGHPWGNAPGA
VANRVALEACVQARNEGRDLAREGNEIIRAASKWSPELAAACEVWKEIKFEFPAMDTL

>**Soy**_RbcL
MSPQTETKASVGFKAGVKDYKLTYYTPDYETKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYGLEPVAGEENQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPTAYIKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFCAEAIFKSQAETGEIKGHYLNATAGTCEEMMKRAVFARELGVPIVMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRLSGGDHVHAGTVVGKLEGERE
ITLGFVDLLRDDFVEKDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAREGNEIIREASKWSPELAAACEVWKEIKFEFEAMDTL

>**Alfalfa**_RbcL
MSPQTETKATVGFKAGVKDYRLTYYTPDYETKDTDILAAFRVSPQPGVPAEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYHIEPVAGEETQFIAYVAYPLDLFEEGSVNYMFTSIVGNVFGFKALRALRLE
DLRIPAAYVKTFQGPPQGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFCAEAIYKAQAETGEIKGHYLNATAGTCEEMMKRAVFARELGVPIVMHDYLTV
GFTANTTLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHAGTVVGKLEGERD
ITLGFVDLLRDDFIEKDRSRGIFFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAREGNEIIREATKWSPELAAACEVWKEIKFEFPAMDN

>**Hemp**_RbcL
MSPQTETKASVGFKAGVKDYKLTYYTPEYQTKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYHIEPVAGEENQFIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPTSYTKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFCAEAIYKSQSETGEIKGHYLNATAGTCEEMMKRAIFARELGVPIVMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGIHFRVLAKALRMSGGDHIHSGTVVGKLEGERE
ITLGFVDLLRDDFIEKDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAREGNEIIREACKWSPELAAACEVWKEIKFEFEAMDTL

>**Apple**_RbcL
MSPQTETKASVGFKAGVKDYKLTYYTPDYETKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYHIEPVAGEESQFIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPVAYVKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFCAEAIYKAQAETGEIKGHYLNATAGTCEDMMKRAVFARELGVPIVMHDYLTG
GFTANTTLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHAGTVVGKLEGERE
ITLGFVDLLRDDFVEKDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAREGNEIIREASKWSPELAAACEIWKEIKFEFEAMDTL

>**Cassava**_RbcL
MSPQTETKASVGFKAGVKDYKLTYYTPDYQTKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYGLEPVPGEENQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRVPPAYSKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFCAEAIYKAQAETGEIKGHYLNATAGTCEEMIKRAVCARELGVPIVMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRLSGGDHIHAGTVVGKLEGERD
ITLGFVDLLRDDFIEKDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAREGNDIIREASKWSPELAAACEVWKEIKFEFAAVDTLD
K

>**Cotton**_RbcL
MSPQTETKASVGFKAGVKEYKLTYYTPEYEVKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYDIEPVPGEEDQYICYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRVPTAYIKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFCAEAIFKSQAETGEIKGHYLNATAGTCEEMIKRAMCARELGVPIVMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHAGTVVGKLEGERD
ITLGFVDLLRDDVIEKDRSRGIYFTQDWVSMPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAREGNEIIREASKWSPELAAACEVWKAIKFEFDAVDKLD
KPAS

>**Rapeseed**_RbcL
MSPQTETKASVGFKAGVKEYKLNYYTPEYETKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYHIEPVPGEETQFIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALAALRLE
DLRIPPAYTKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFCAEAIYKSQAETGEIKGHYLNATAGTCEEMMKRAIFARELGVPIVMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRLSGGDHVHAGTVVGKLEGDRE
STLGFVDLLRDDYVEKDRSRGIFFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAVEGNEIIREACKWSPELAAACEVWKEITFNFPTIDKLD
GQD

>**Arabidopsis**_RbcL
MSPQTETKASVGFKAGVKEYKLTYYTPEYETKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYHIEPVPGEETQFIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALAALRLE
DLRIPPAYTKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFCAEAIYKSQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTG
GFTANTSLSHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRLSGGDHIHAGTVVGKLEGDRE
STLGFVDLLRDDYVEKDRSRGIFFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAVEGNEIIREACKWSPELAAACEVWKEITFNFPTIDKL

>**Beet**_RbcL
MSPQTETKASVGFKAGVKDYKLTYYTPEYETLDTDILAAFRVSPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYHIEPVAGEENQYICYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPVAYVKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDNE
NVNSQPFMRWRDRFLFCAEALYKAQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTG
GFTANTTLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRLSGGDHIHSGTVVGKLEGERD

ITLGFVDLLRDDYTEKDRSRGIYFTQSWVSTPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRNLATEGNAIIREASKWSPELAAACEVWKEIKFEFPAMDTV

>**Spinach**_RbcL
MSPQTETKASVEFKAGVKDYKLTYYTPEYETLDTDILAAFRVSPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTNLDRYKGRCYHIEPVAGEENQYICYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPVAYVKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFCAEALYKAQAETGEIKGHYLNATAGTCEDMMKRAVFARELGVPIVMHDYLTG
GFTANTTLSHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRLSGGDHIHSGTVVGKLEGERD
ITLGFVDLLRDDYTEKDRSRGIYFTQSWVSTPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAREGNTIIREATKWSPELAAACEVWKEIKFEFPAMDTV

>**Rubber**_RbcL
VGFKAGVKDYKLNYYTPDYETKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTAVWTDGLTSLD
RYKGRCYHIEPVAGEENQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPPAYSK
TFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMRW
RDRFLFCAEALYKAQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTGGFTANTSLAH
YCRDNGLLLHIHRAMHAVIDRQKNHGIHFRVLAKALRMSGGDHIHAGTVVGKLEGERDITLGFVDLLR
DDFVEKDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLGHPWGNAPGAV
ANRVALEACVRARNEGRDLAREGNEIIREAAKWSPELAAACEVWKAIKFEFPAMDTL

>**Coffee**_RbcL
MSPQTETKASVGFKAGVKEYKLTYYTPEYETKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTA
VWTDGLTSLDRYKGRCYHIEPVPGEENQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRVPPAYIKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFCFCAEALFKAQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKGLRMSGGDHIHAGTVVGKLEGERD
ITLGFVDLLRDDFIEKDRSRGIYFTQDWVSLPGVIPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVKARNEGRDLAAEGNEIIREASKWSPELAAACEVWKEIRFNFEAMDKLD
KEKDL

>**Sweet_potato**_RbcL
MSPQTETKASVGFKAGVKDYKLTYYTPEYQTKDTDILAASRVTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYRIERVIGEKDQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRISTAYIKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFCAEALYKAQAETGEIKGHYLNATAGTCEEMMKRAIFARELGVPIVMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRLSGGDHIHAGTVVGKLEGERE
ITLGFVDLLRDDFVEQDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVQARNEGRDLAREGNEIIREACKWSPELSAACEVWKEIRFEFKPVDTLD
PGTA

>**Pepper**_RbcL
SVGFKAGVKEYKLTYYTPEYQTKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTTVWTDGLTSL
DRYKGRCYRIERVVGEKDQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRVPTAYI
KTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMR
WRDRFLFCAEALFKAQTETGEIKGHYLNATAGTCEEMMKRAIFARELGVPIVMHDYLTGGFTANTSLA
HYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHAGTVVGKLEGERDITLGFVDLL
RDDFVEQDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLGHPWGNAPGA
VANRVALEACVKPRNEGRDLAREGNEIIREACKWSPELAAACEVWKEIVFNFAAVDVLDK

>**Aubergine**_RbcL
MSPQTETKASVGFKAGVKEYKLTYYTPEYQTKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYRIERVVGEKDQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPPAYVKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFCFCAEALFKAQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTG

GFTANTTLAHYCRDNGLLLHIHRAMHAVIDRQKNHGIHFRVLAKALRMSGGDHIHAGTVVGKLEGERD
ITLGFVDLLRDDFVEQDRSRGIYFTQDWVSLPGVIPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVKARNEGRDLAQEGNEIIREASKWSPELAAACEVWKEIVFNFAAVDVLD
K

>**Potato**_RbcL
MSPQTETKASVGFKAGVKEYKLTYYTPEYQTKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYRIERVVGEKDQYIAYVAYPLDLFEEGSVTNMLTSIVGNVFGFKALRALRLE
DLRIPVAYVKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFCAEALFKAQAETGEIKGHYLNATAGTCEEMMKRAVFARELGTPIVMHDYLTG
GFTANTTLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHSGTVVGKLEGERD
ITLGFVDLLRDDFIEQDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVKARNEGRDLAREGNEIIREAAKWSPELAAACEVWKEIVFNFAAMDVLD
K

>**Tomato**_RbcL
MSPQTETKASVGFKAGVKEYKLTYYTPEYQTKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYRIERVVGEKDQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPPAYVKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFCAEALFKAQTETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTG
GFTANTTLAHYCRDNGLLLHIHRAMHAVIDRQKNHGIHFRVLAKALRMSGGDHIHSGTVVGKLEGERD
ITLGFVDLLRDDFVEQDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVKARNEGRDLAREGNEIIREACKWSPELAAACEVWKEIVFNFAAVDVLD
K

>**Tobacco**_RbcL
MSPQTETKASVGFKAGVKEYKLTYYTPEYQTKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTT
VWTDGLTSLDRYKGRCYRIERVVGEKDQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLE
DLRIPPAYVKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDE
NVNSQPFMRWRDRFLFCAEALYKAQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTG
GFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGIHFRVLAKALRMSGGDHIHSGTVVGKLEGERD
ITLGFVDLLRDDFVEQDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLG
HPWGNAPGAVANRVALEACVKARNEGRDLAQEGNEIIREACKWSPELAAACEVWKEIVFNFAAVDVLD
K

8.9. SPR Sensorgrams and raw data

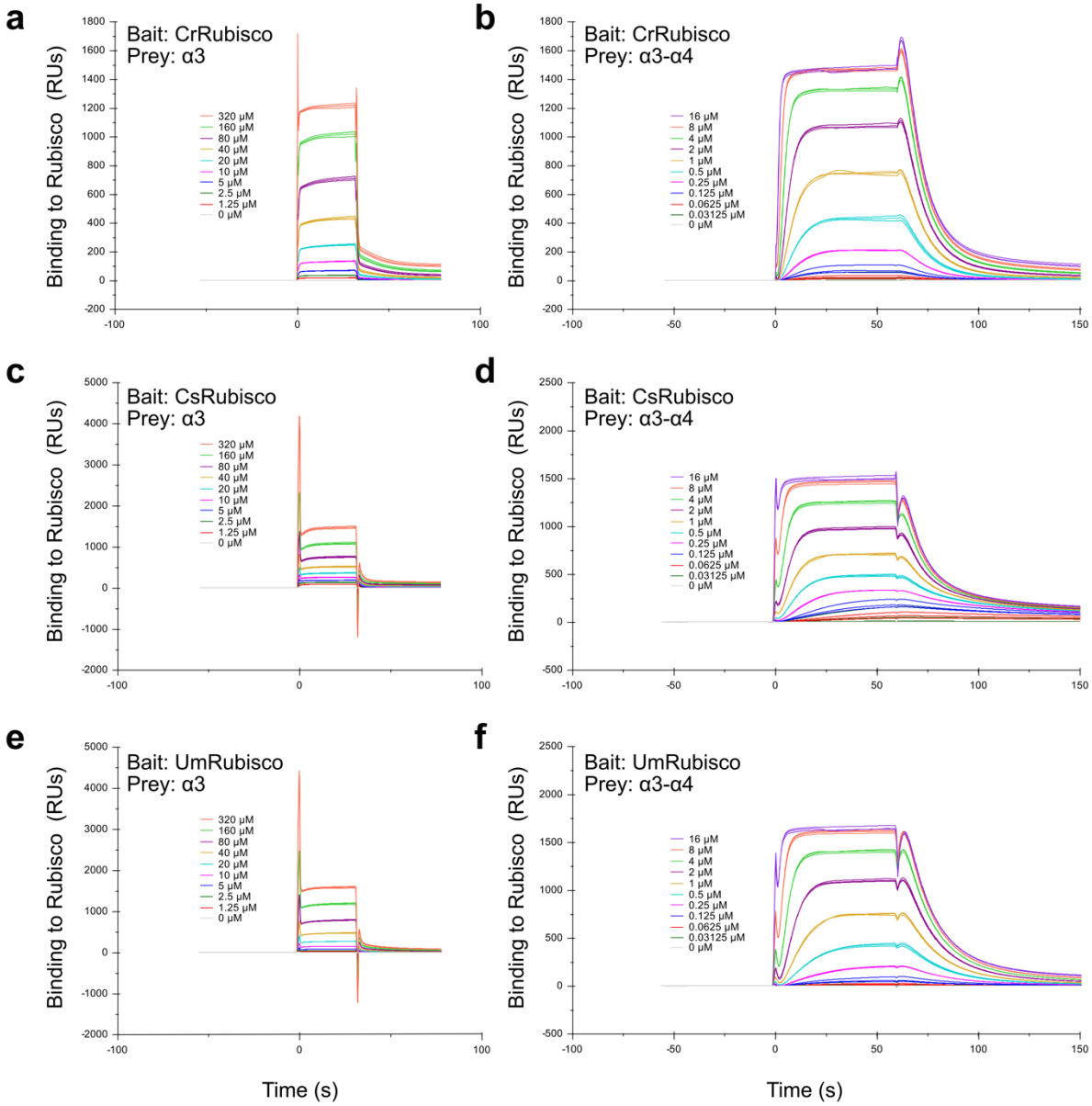


Figure 8-1 | Sensorgrams from SPR experiments.
(a) Raw sensorgrams from the titration of α3 against Chlamydomonas Rubisco. (b) Raw sensorgrams from the titration of α3-α4 against Chlamydomonas Rubisco. (c) Raw sensorgrams from the titration of α3 against Chlorella Rubisco. (d) Raw sensorgrams from the titration of α3-α4 against Chlorella Rubisco. (e) Raw sensorgrams from the titration of α3 against Ulva Rubisco. (f) Raw sensorgrams from the titration of α3-α4 against Ulva Rubisco.

Table 8-1: Reference-corrected data from SPR experiments

Bait	Prey	Prey concentration (μM)	Binding to Rubisco (RUs)
Chlamydomonas	α3	0	-0.3
		1.25	17.2
		2.5	34.8
		5	69.1
		10	133.8
		20	249.7
		40	440.7
		80	716.3
		160	1026.6
		320	1226.2
		0	-0.2
		1.25	16
		2.5	32.5
		5	67.7
		10	131.1
		20	245.1
		40	431
		80	704.2
		160	1013.9
		320	1213.4
		0	-0.2
		1.25	15.3
		2.5	31.9
		5	66.1
		10	129.7
		20	241.9
		40	423.4
		80	693.7
		160	995.6
		320	1198.6
		0	-0.3
		1.25	17.2
		2.5	34.8
		5	69.1
		10	133.8
		20	249.7
		40	440.7
		80	716.3
		160	1026.6
		320	1226.2
Chlorella	α3	0	0.5
		1.25	86.6
		2.5	130.1
		5	185.4
		10	260.5
		20	368
		40	526.9
		80	766.9
		160	1100.8
		320	1491.3
		0	0.7
		1.25	82.6
		2.5	124.6
		5	179.7
		10	252.1
		20	356.1

		40	511.2
		80	746.1
		160	1076.6
		320	1463
		0	0.9
		1.25	81.1
		2.5	122.4
		5	176.4
		10	248.4
		20	350.6
		40	502.2
		80	733.6
		160	1055.7
		320	1440.7
		0	0.5
		1.25	86.6
		2.5	130.1
		5	185.4
		10	260.5
		20	368
		40	526.9
		80	766.9
		160	1100.8
		320	1491.3
Ulva	α3	0	-0.5
		1.25	18.7
		2.5	38.1
		5	75.1
		10	145
		20	268.5
		40	474.9
		80	790
		160	1195.5
		320	1597.1
		0	0
		1.25	18.6
		2.5	37.5
		5	75.4
		10	144.4
		20	266.9
		40	469.7
		80	781
		160	1183.9
		320	1581
		0	0.3
		1.25	18.5
		2.5	37.2
		5	74
		10	143.3
		20	263.8
		40	462.2
		80	769.8
		160	1162.5
		320	1558.7
		0	-0.5
		1.25	18.7
		2.5	38.1
		5	75.1
		10	145
		20	268.5

		40	474.9
		80	790
		160	1195.5
		320	1597.1
Chlamydomonas	α3-α4	0	-0.5
		0.03125	32.5
		0.0625	56
		0.125	105.8
		0.25	205.6
		0.5	446.4
		1	755.6
		2	1093
		4	1342.8
		8	1477.7
		16	1493.7
		0	0.2
		0.03125	19.8
		0.0625	14.2
		0.125	53.8
		0.25	208.8
		0.5	429.8
		1	745
		2	1073.4
		4	1331.9
		8	1468.8
		16	1469.9
		0	0.3
		0.03125	11.5
		0.0625	1.1
		0.125	65.2
		0.25	210
		0.5	413.4
		1	729.4
		2	1064.6
		4	1316.7
		8	1457.5
		16	1468.1
Chlorella	α3-α4	0	0.3
		0.03125	93.8
		0.0625	150.9
		0.125	233.3
		0.25	331
		0.5	496.8
		1	718.2
		2	996.3
		4	1269.1
		8	1474.4
		16	1528
		0	0.9
		0.03125	60.2
		0.0625	45
		0.125	151.1
		0.25	331.7
		0.5	486.3
		1	707.8
		2	979.1
		4	1256.7
		8	1461.2
		16	1496.6
		0	1.2

		0.03125	36.7
		0.0625	8.6
		0.125	173.1
		0.25	332.7
		0.5	475.2
		1	698.5
		2	970.6
		4	1237.5
		8	1440.9
		16	1488.1
Ulva	α3-α4	0	-0.3
		0.03125	23.5
		0.0625	43.3
		0.125	91
		0.25	196.2
		0.5	440.2
		1	756.7
		2	1117
		4	1421.9
		8	1623.8
		16	1671.2
		0	0.2
		0.03125	14.5
		0.0625	10.6
		0.125	43.7
		0.25	199.6
		0.5	428.6
		1	747.1
		2	1098.5
		4	1410
		8	1611.8
		16	1639
		0	0.2
		0.03125	8.2
		0.0625	1.4
		0.125	54.5
		0.25	203.6
		0.5	412.6
		1	737.1
		2	1089.8
		4	1390.8
		8	1593.8
		16	1634.8