

# A promiscuous mechanism to phase separate eukaryotic carbon fixation in the green lineage

James Barrett<sup>1,2</sup>, Mihris I.S. Naduthodi<sup>1,2</sup>, Yuwei Mao<sup>3,4</sup>, Clément Dégut<sup>1</sup>, Sabina Musiał<sup>1,2</sup>, Aidan Salter<sup>1,2</sup>, Mark C. Leake<sup>1,5</sup>, Michael J. Plevin<sup>1,6</sup>, Alistair J. McCormick<sup>3,4</sup>, James N. Blaza<sup>6,7</sup>, Luke C.M. Mackinder<sup>1,2\*</sup>

## Affiliations

<sup>1</sup>Department of Biology, University of York; York, YO10 5DD, UK.

<sup>2</sup>Centre for Novel Agricultural Products (CNAP), Department of Biology, University of York; York, YO10 5DD, UK.

<sup>3</sup>Institute of Molecular Plant Sciences, School of Biological Sciences, University of Edinburgh; Edinburgh, EH9 3BF, UK.

<sup>4</sup>Centre for Engineering Biology, University of Edinburgh; Edinburgh, EH9 3BF, UK.

<sup>5</sup>School of Physics, Engineering and Technology, University of York; York, YO10 5DD, UK.

<sup>6</sup>York Structural Biology Laboratory, The University of York; York, YO10 5DD, UK.

<sup>7</sup>Department of Chemistry, York Structural Biology Laboratory; York, YO10 5DD, UK.

\*Corresponding author. Email: [luke.mackinder@york.ac.uk](mailto:luke.mackinder@york.ac.uk)

## Abstract

CO<sub>2</sub> fixation is commonly limited by inefficiency of the CO<sub>2</sub>-fixing enzyme Rubisco. Eukaryotic algae concentrate and fix CO<sub>2</sub> in phase-separated condensates called pyrenoids, which complete up to one-third of global CO<sub>2</sub> fixation. Condensation of Rubisco in pyrenoids is dependent on interaction with disordered linker proteins that show little conservation between species. We developed a sequence-independent bioinformatic pipeline to identify linker proteins in green algae. We report the linker from *Chlorella* and demonstrate that it binds a conserved site on the Rubisco large subunit. We show the *Chlorella* linker phase separates *Chlamydomonas* Rubisco and that despite their separation by ~800 million years of evolution, the *Chlorella* linker can support the formation of a functional pyrenoid in *Chlamydomonas*. This cross-species reactivity extends to plants, with the *Chlorella* linker able to drive condensation of some native plant Rubiscos *in vitro* and *in planta*. Our results represent an exciting frontier for pyrenoid engineering in plants, which is modelled to increase crop yields.

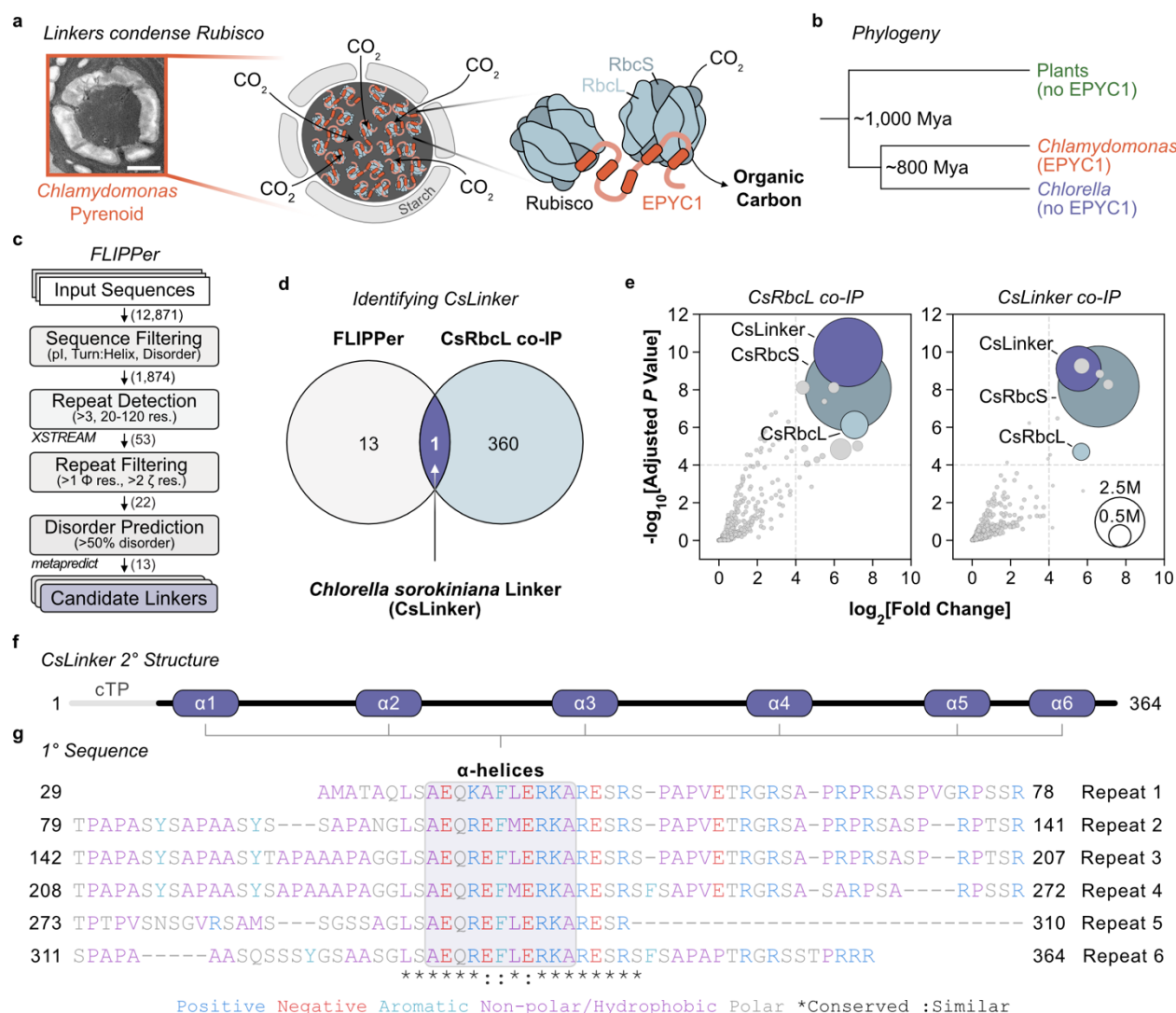
As the primary gateway between atmospheric carbon dioxide (CO<sub>2</sub>) and organic carbon, ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) fixes approximately 400 gigatons of CO<sub>2</sub> annually<sup>1</sup>. Despite this huge global productivity, Rubisco as an enzyme is catalytically slow<sup>2</sup>. Its Archean origin<sup>3</sup> and exceptionally slow evolutionary trajectory<sup>4</sup>, has meant Rubisco has failed to significantly overcome the supposed trade-off between specificity for its substrate (CO<sub>2</sub> or O<sub>2</sub>) and catalytic rate<sup>4,5</sup>. These shortfalls mean Rubisco is often limiting for photosynthesis in most plants (i.e. C<sub>3</sub> plants)<sup>6-11</sup>. Accordingly, C<sub>3</sub> plants compensate by producing and maintaining large amounts of Rubisco<sup>12,13</sup>. By mass, Rubiscos in aquatic phototrophs (algae and cyanobacteria) are ~20 times more efficient<sup>14</sup>. These Rubiscos benefit from operating in biophysical CO<sub>2</sub>-concentrating mechanisms (CCMs) that increase the CO<sub>2</sub>:O<sub>2</sub> ratio at their active site. A large proportion of aquatic CO<sub>2</sub>-fixation occurs in the pyrenoid<sup>15</sup>, a sub-compartment of the chloroplast found in most eukaryotic algae and some basal land plants that is the centerpiece of their biophysical CCMs<sup>16</sup>. Pyrenoid formation is underpinned by biomolecular condensation of Rubisco by disordered linker proteins<sup>15,17</sup> (Fig. 1a). Engineering pyrenoid-based CCMs in crop plants that operate C<sub>3</sub> photosynthesis is a promising avenue to increase their primary productivity and reduce nitrogen and water usage<sup>18,19</sup>. While significant progress has been made in the characterization and transfer of pyrenoid components from the model alga *Chlamydomonas reinhardtii* to the model C<sub>3</sub> plant *Arabidopsis thaliana*<sup>15,20-24</sup>, our knowledge of pyrenoids from other species remains limited. By characterizing pyrenoids from other species, we hope to expand the toolbox available for plant pyrenoid engineering and gain insight into the commonalties and differences between pyrenoid assembly.

Here, we identify and characterize CsLinker, a pyrenoid linker protein from the green alga *Chlorella* which is an ancient relative of *Chlamydomonas*. Using biochemical and structural approaches, we demonstrate that CsLinker is functionally analogous to linker proteins from other organisms, despite low sequence identity. Crucially, and in contrast to the *Chlamydomonas* linker EPYC1, we show that CsLinker binds to the Rubisco large subunit (RbcL). The high conservation of the binding interface on the RbcL supports functional cross-reactivity of *Chlamydomonas* Rubisco and CsLinker as well as CsLinker-mediated condensation of native plant Rubiscos both *in vitro* and *in planta*. These findings represent a significant advance towards the engineering of synthetic pyrenoids and overcome a major hurdle for the future engineering of pyrenoid-based CCMs in plants.

# Results

## A Fast Linker Identification Pipeline for Pyrenoids (FLIPPer) identifies the pyrenoid linker in *Chlorella sorokiniana*

The *Chlamydomonas* pyrenoid linker protein EPYC1 is not conserved outside closely related species<sup>15</sup> (Fig. 1b). To identify functional analogues in other pyrenoid-containing species of the green lineage, we developed a sequence homology-independent Fast Linker Identification Pipeline for Pyrenoids (FLIPPer). FLIPPer searches for proteins with features key to the function of EPYC1, namely: i) largely disordered, with ii) repeating structured elements, that are iii) spaced at a feasible length scale for cross-linking Rubiscos (20-120 residues; ~2-8 nm<sup>25,26</sup>) (Fig. 1c). Amongst our analyses, we identified 13 proteins with these features in the genome of the unicellular Trebouxiophyte *Chlorella sorokiniana* UTEX1230<sup>27</sup> (*Chlorella* hereafter) (Extended Data Fig. 1), which diverged from *Chlamydomonas* ~800 million years ago (Mya)<sup>28</sup> (Fig. 1b). To validate the identity of the *Chlorella* linker, we completed co-immunoprecipitation (co-IP) experiments on *Chlorella* cells grown in low CO<sub>2</sub>, using an antibody specific to the Rubisco large subunit (RbcL) (Fig. 1e, Extended Data Fig. 2a,b and Supplementary Table 1). Of the 360 proteins enriched relative to control co-IPs (*Chlamydomonas* BST2 antibody), only 1 was shared amongst the FLIPPer outputs (Fig. 1d). This protein, CSI2\_123000012064, was the most significantly enriched in the RbcL co-IP experiment alongside both subunits of the *Chlorella* Rubisco holoenzyme (CsRbcL and CsRbcS) (Fig. 1e, left). We called this protein CsLinker. We subsequently completed reciprocal co-IPs using a CsLinker antibody (Extended Data Fig. 2c,d), which enriched both CsRbcL and CsRbcS, indicating *ex vivo* complex formation between CsLinker and *Chlorella* Rubisco (Fig. 1e, right). We confirmed the CsLinker gene sequence, gene model and predicted chloroplast transit peptide (cTP) by PCR and sequencing and through mapping of peptides identified in mass spectrometry experiments (Extended Data Fig. 3). We further confirmed the low CO<sub>2</sub> inducibility of CsLinker through analysis of available RNA-seq data, and by western blotting (Extended Data Fig. 4 and Supplementary Table 2). Whilst AlphaFold 2 modelling demonstrated CsLinker and EPYC1 share clear structural analogy, their primary sequences share little similarity (Fig. 1g, 25% identity; Extended Data Fig. 5), reflecting their independent origins.



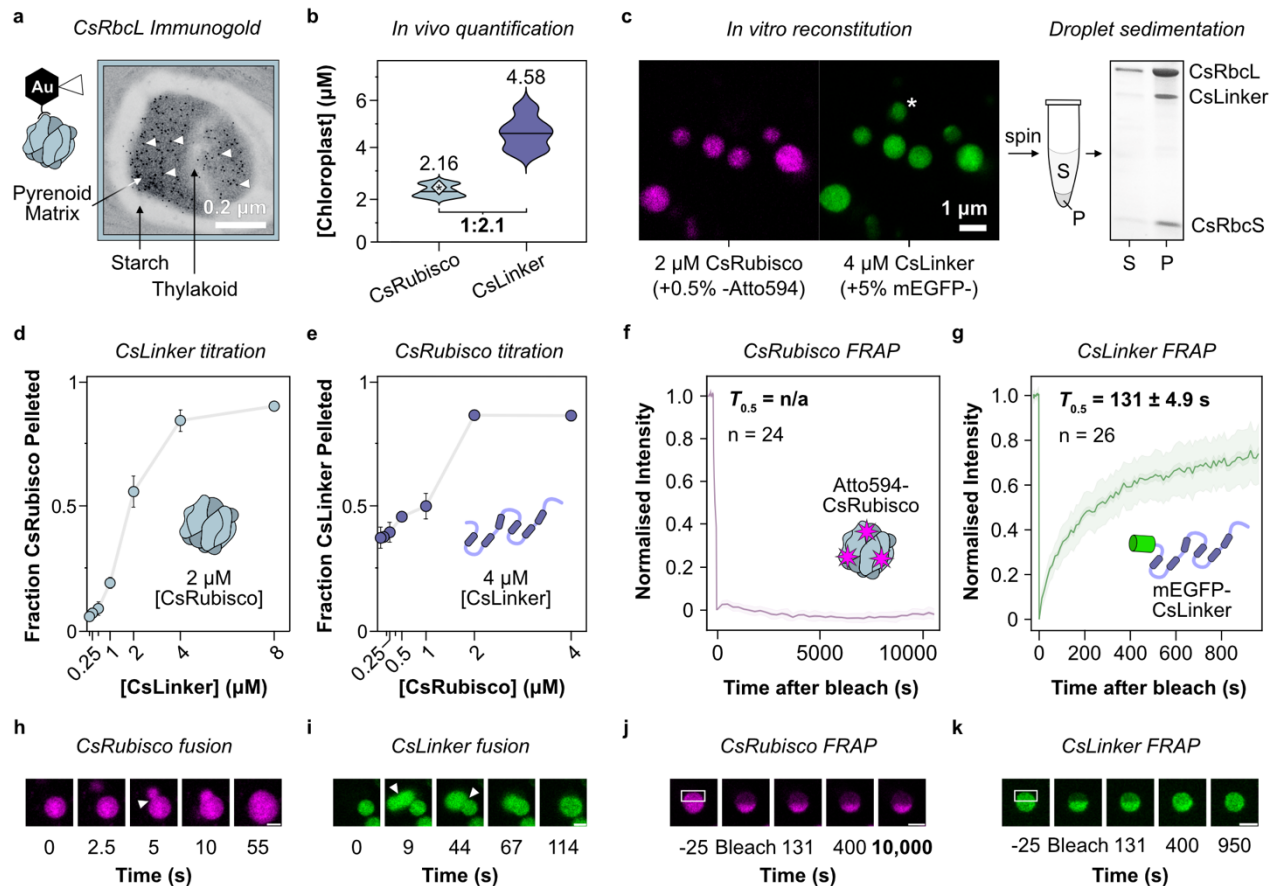
**Fig. 1 | Identification of the *Chlorella sorokiniana* linker protein (CsLinker).** **a**, Transmission electron micrograph (TEM) of the *Chlamydomonas reinhardtii* pyrenoid, with adjacent schematic of Rubisco condensation in the pyrenoid by interaction of EPYC1 helices with the Rubisco small subunits (RbcSs). Condensed Rubisco fixes CO<sub>2</sub> to organic carbon. Scale bar = 1 μm. **b**, Phylogeny of *Chlamydomonas*, *Chlorella* and plants. Estimated divergence points from a time-calibrated phylogeny<sup>28</sup>. **c**, Schematic representation of the Fast Linker Identification Pipeline for Pyrenoids (FLIPPer) used to identify candidate linkers that share features with EPYC1. Where relevant, the program used is indicated. The number of sequences remaining after each filtering step of the *Chlorella sorokiniana* UTEX1230 genome is indicated. Abbreviations: pl = isoelectric point, res. = residue, Φ = hydrophobic, ζ = electrostatic. **d**, Venn diagram demonstrating identification of CsLinker from FLIPPer and CsRbcL co-immunoprecipitation (co-IP). **e**, Reciprocal co-IP experiments performed using antibodies raised to the Rubisco large subunit (left) and CsLinker (right). Dashed lines indicate arbitrary significance thresholds ( $-\log_{10}[\text{Adjusted } P \text{ Value}] > 4$ ,  $\log_2[\text{Fold Change}] > 4$ ), above which points are sized according to their summed abundance (M = millions) from three replicates as per the inset key. **f**, Predicted secondary structure of CsLinker from AlphaFold modelling (Extended Data Fig. 1). The predicted chloroplast transit peptide (cTP) and α-helices (α1-6) are indicated. **g**, Primary sequence alignment of the six repeat regions of CsLinker, colored by residue property.

# CsLinker is a *bona fide* pyrenoid linker protein

Having demonstrated *ex vivo* interaction between CsLinker and CsRubisco (Fig. 1e), we sought to confirm both were abundant components of the *Chlorella* pyrenoid *in vivo*. Using the RbcL antibody, we completed immunoelectron microscopy and observed CsRubisco almost exclusively localized in the pyrenoid ( $96.2 \pm 3.5\%$  S.D.,  $n=6$ ) (Fig. 2a, Extended Data Fig. 6 and Supplementary Table 3). We observed no immuno-labelling using the CsLinker antibody, likely due to poor antibody antigenicity (Extended Data Fig. 6c). To determine the abundance of CsLinker and CsRubisco *in vivo*, we completed absolute quantification mass spectrometry using standard curves of recombinant CsLinker and CsRubisco purified from *Chlorella*. Accounting for the chloroplast volume<sup>29</sup>, we measured the concentration of CsLinker and CsRubisco to be  $4.58 \pm 0.53\ \mu\text{M}$  S.D. and  $2.16 \pm 0.04\ \mu\text{M}$  S.D. respectively (Fig. 2b, Extended Data Fig. 7 and Supplementary Tables 4-7), demonstrating CsLinker is highly abundant in the chloroplast. We validated the Rubisco quantification by western blotting ( $2.42\ \mu\text{M}$ ; asterisk in Fig. 2b, Extended Data Fig. 2b and Supplementary Table 8). Alongside the localization of CsRubisco in the pyrenoid (Fig. 2a), the two-fold abundance of CsLinker over CsRubisco *in vivo* (Fig. 2b) and their *ex vivo* interaction (Fig. 1e) indicate CsLinker likely interacts with CsRubisco as an abundant component of the *Chlorella* pyrenoid *in vivo*.

The *Chlamydomonas* pyrenoid is a liquid-liquid phase-separated (LLPS) biomolecular condensate *in vivo*<sup>26</sup>, which is essential to Rubisco packaging and pyrenoid function<sup>15</sup>. *In vitro*, reconstituted pyrenoids formed by mixing purified linker EPYC1 and *Chlamydomonas* Rubisco demonstrate similar properties to *in vivo*<sup>30</sup>. Accordingly, we sought to understand whether mixing CsLinker and CsRubisco gives rise to similar emergent properties *in vitro*. When mixed at the chloroplast concentrations we measured (Fig. 2b), we observed demixing into micron-scale droplets that was dependent on, and incorporated both, CsLinker and CsRubisco (Fig. 2c, Extended Data Fig. 8a,b). To assess the relative occupancy of the components within droplets, we completed reciprocal titration in droplet sedimentation assays (Fig. 2c, right). By separately fixing both components at their chloroplast concentrations, we observed requirement of an approximately two-fold excess of CsLinker to fully demix both components (Fig. 2d,e, Extended Data Fig. 8c,d). This observation agreed with the approximately two-fold abundance of CsLinker we measured *in vivo* (Fig. 2b), and previous observations in *Chlamydomonas*<sup>30</sup>. Using the same ratio, we observed a critical global concentration ( $0.3\text{-}0.5\ \mu\text{M}$ ) and salt-dependency for droplet formation (Extended Data Fig. 9); both key indicators of LLPS<sup>31</sup>. To further understand the droplet properties, we completed fluorescence recovery after photobleaching (FRAP) experiments, which allowed us to monitor the mobility of labelled CsLinker and CsRubisco. Strikingly, we observed largely different mobilities for CsRubisco and CsLinker in droplets. Whilst mEGFP-CsLinker exhibited exchange with the dilute phase in whole FRAP experiments ( $T_{0.5} = 201 \pm 14.3\ \text{s}$  S.E.M.; Extended Data Fig. 10a and Supplementary Table 9) and internal mixing in half-FRAP experiments ( $T_{0.5} = 131 \pm 4.89\ \text{s}$  S.E.M.; Fig. 2g,k), Atto594-CsRubisco demonstrated no mixing or exchange over hour timescales (Fig. 2f,j). However, both CsLinker and CsRubisco appeared to undergo internal rearrangement over second timescales upon droplet fusion (Fig. 2h,i), suggesting CsRubisco is not immobilized in droplets, akin to observations of the *Phaeodactylum tricornutum* pyrenoid reconstitution<sup>17</sup>. To ensure our observations were not set-up specific, we confirmed the mobility of *Chlamydomonas* Rubisco (CrRubisco) and EPYC1 in the *Chlamydomonas* pyrenoid reconstitution using the same strategy. Consistent with previous *in vitro*<sup>30</sup> and *in vivo*<sup>26</sup> observations, both components were highly mobile (EPYC1-mEGFP half-FRAP  $T_{0.5} = 22 \pm 6.3\ \text{s}$  S.E.M., Atto594-CrRubisco half-FRAP  $T_{0.5} = 55 \pm 5.1\ \text{s}$  S.E.M.; Extended Data Fig. 10d,e). Taken together these results demonstrate the ability of CsLinker to phase separate *Chlorella* Rubisco *in vitro*, and alongside our *ex vivo* and *in vivo* observations, indicate this process likely underpins pyrenoid formation *in vivo*, analogous to EPYC1 in *Chlamydomonas*.





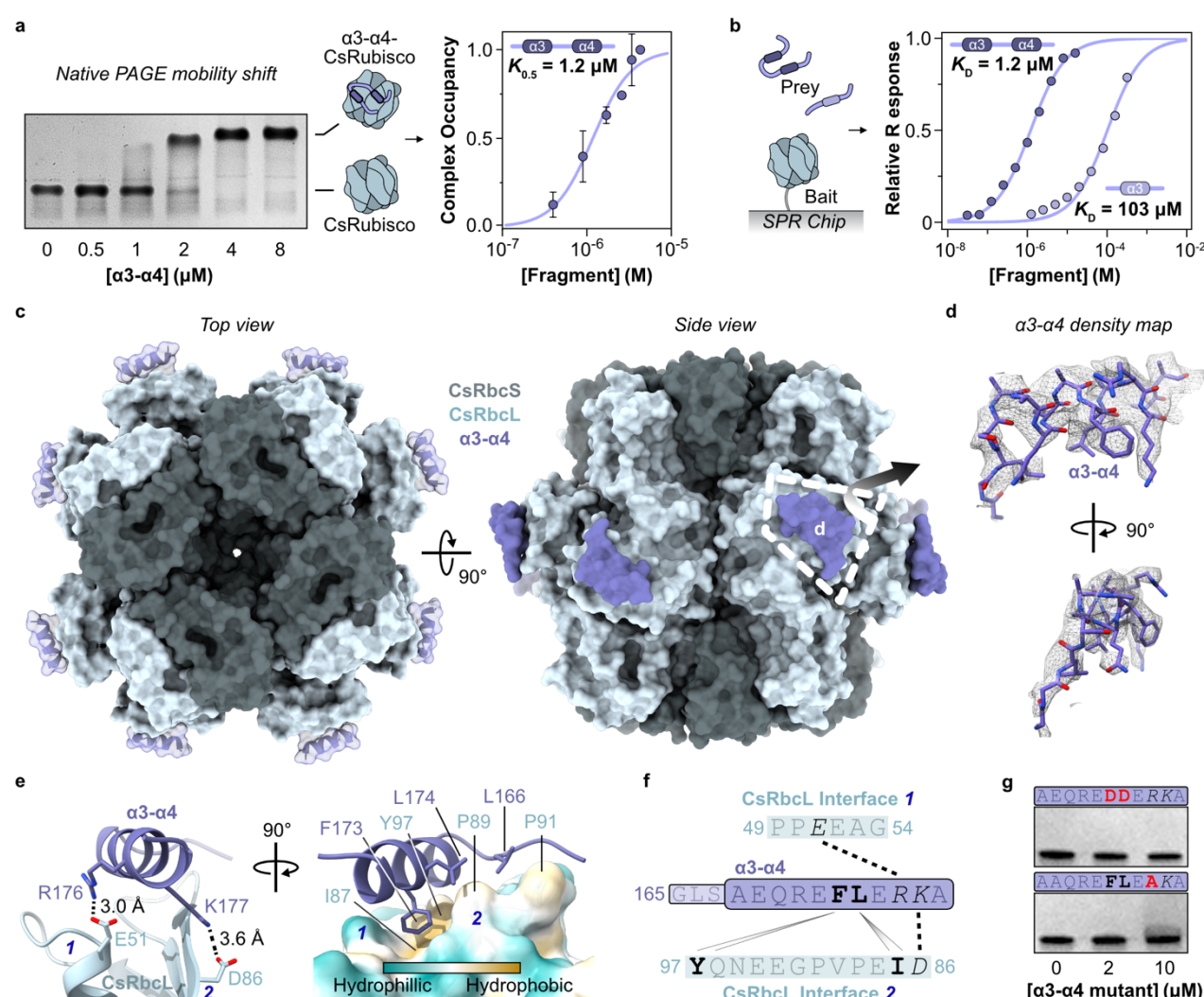
**Fig. 2 | CsLinker phase separates Rubisco at physiological conditions.** **a**, Representative immunogold TEM of the *Chlorella* pyrenoid after primary incubation with the RbcL antibody. A subset of gold nanoparticles are indicated by white arrowheads. **b**, Violin plot from absolute quantification of *CsRubisco* holoenzyme (derived from *CsRbcL*) and *CsLinker* *in vivo* ( $n = 3$ ). The ratio between *CsRubisco* and *CsLinker* is indicated below. (See Extended Data Fig. 7 for details). The asterisk point indicates the independent quantification from western blotting (Extended Data Fig. 2b). **c**, Confocal fluorescence microscopy image of droplets in the *in vitro* reconstitution of the *Chlorella* pyrenoid formed at the measured chloroplast concentrations of *CsRubisco* and *CsLinker* ( $2 \mu\text{M}$  and  $4 \mu\text{M}$ , respectively) (left panel). Asterisk indicates a droplet that settled between imaging of the two channels. Atto594-*CsRubisco* and mEGFP-*CsLinker* were incorporated at 0.5% and 5% molar concentrations respectively. Droplets can be sedimented by centrifugation and the composition of the pellet (P), relative to the supernatant (S) analyzed by SDS-PAGE (right panel). This experiment format was used to generate datapoints in d and e. **d**, Titration droplet sedimentation assays with *CsRubisco* fixed at the measured chloroplast concentration. **e**, Titration droplet sedimentation assays with fixed *CsLinker*. Where visible, error bars represent the S.D. from two replicates. **f**, Average full-scale normalized half-FRAP recovery curve of Atto594-*CsRubisco* in droplets formed as in c. The mean, S.E.M and S.D. are represented by the line, the smaller shaded region, and the larger shaded region respectively. **g**, Average full-scale normalized half-FRAP recovery curve of mEGFP-*CsLinker* in droplets with the same composition. **h**, Time series of 0.05% (molar ratio) Atto594-*CsRubisco* labelled droplets formed as in c, undergoing fusion (arrow) and relaxation. **i**, 5% mEGFP-*CsLinker* labelled droplets undergoing consecutive fusions. **j**, Time series of a representative *CsRubisco* FRAP experiment, as quantified in f. The white box indicates the region bleached. **k**, Representative *CsLinker* FRAP experiment, as quantified in g. Scale bars for panels h-k =  $1 \mu\text{m}$ .

## CsLinker binds to the Rubisco large subunit (RbcL)

Previously characterized Rubisco-condensing linker proteins from algal pyrenoids<sup>17,32</sup> and bacterial carboxysomes<sup>33,34</sup> utilize structured regions to bind different regions on Rubisco. We hypothesized that the predicted  $\alpha$ -helical regions in CsLinker may bind to Rubisco and that this interaction would involve a previously uncharacterized interface. To characterize the interaction, we produced fragments of CsLinker encompassing entire repeat sequences centered on predicted  $\alpha$ -helix 3 ( $\alpha 3$ ) and  $\alpha$ -helices 3 and 4 ( $\alpha 3$ - $\alpha 4$ ) (Fig. 1f,g, Extended Data Fig. 11a). Using 2D NMR spectroscopy, we confirmed the only stable structure in the fragments to be  $\alpha$ -helices of approximately 10 residues, in line with structural predictions (Fig. 1f, Extended Data Fig. 5a, 11). The similarity of the NMR spectra of  $\alpha 3$  and  $\alpha 3$ - $\alpha 4$  indicated the individual repeat regions have similar overall properties (Extended Data Fig. 11d), consistent with their similar residue composition and the lack of stable tertiary structure. We next tested the ability of the fragments to interact with CsRubisco. In line with the dependency of LLPS on multivalent interactions for cross-linking, the single helix  $\alpha 3$  fragment was unable to phase-separate CsRubisco, but did demonstrate concentration-dependent mobility shift of CsRubisco in native PAGE experiments (Extended Data Fig. 12a-c). Surprisingly, the  $\alpha 3$ - $\alpha 4$  double repeat fragment was also unable to induce LLPS but did demonstrate formation of a stable higher-order complex with CsRubisco that had a half-occupancy ( $K_{0.5}$ ) of  $\sim 1.2 \mu\text{M}$  ( $\text{CI}_{95}$ :  $0.93$ - $1.42 \mu\text{M}$ ) (Fig. 3a, Extended Data Fig. 12d-h and Supplementary Table 10). To measure the affinity of the fragments for CsRubisco, we used surface plasmon resonance (SPR) experiments in which CsRubisco was immobilized as bait, and the  $\alpha 3$  and  $\alpha 3$ - $\alpha 4$  fragments were used as prey (Fig. 3b, left). We measured a similar affinity ( $K_D$ ) of the  $\alpha 3$ - $\alpha 4$  fragment for CsRubisco ( $1.21 \mu\text{M}$  [ $\text{CI}_{95}$ :  $1.10$ - $1.33 \mu\text{M}$ ]) as we observed for the  $K_{0.5}$  of the  $\alpha 3$ - $\alpha 4$ -CsRubisco complex by native PAGE (Fig. 3a, Extended Data Fig. 13 and Supplementary Table 11). The  $K_D$  of the  $\alpha 3$  fragment was  $\sim 100$ -fold higher ( $103 \mu\text{M}$  [ $\text{CI}_{95}$ :  $80$ - $133 \mu\text{M}$ ]) than  $\alpha 3$ - $\alpha 4$ , consistent with cooperative binding of the two  $\alpha$ -helical regions in  $\alpha 3$ - $\alpha 4$  to the same Rubisco rather than other Rubiscos in solution. This explained the lack of LLPS, and indicated the higher-order complex observed by native PAGE likely consists of single Rubiscos bound by  $\alpha 3$ - $\alpha 4$ .

To determine where CsLinker binds CsRubisco, we elucidated the 3D structure of the  $\alpha 3$ - $\alpha 4$ -CsRubisco complex using cryogenic electron microscopy (cryo-EM) (Supplementary Table 12). We prepared samples of  $\alpha 3$ - $\alpha 4$  and CsRubisco such that the solution concentration of CsLinker repeats was comparable to the value we measured in the chloroplast ( $32 \mu\text{M}$  in this experiment,  $27.5 \mu\text{M}$  *in vivo*; Fig. 2b). This concentration also saturated the  $\alpha 3$ - $\alpha 4$ -CsRubisco complex (Fig. 3a). The  $2.4 \text{ \AA}$  map and model we obtained with D4 reconstruction (PDB: 8Q04) was highly similar to a previously determined cryo-EM structure of *Chlamydomonas* Rubisco<sup>32</sup> (Ca RMSD of RbcL =  $1.106 \text{ \AA}$  and RbcS =  $0.937 \text{ \AA}$ , Extended Data Fig. 14k). During processing we observed low-resolution density additional to the core subunits at the equatorial region of CsRubisco in the RbcL (Extended Data Fig. 14d). We hypothesized this additional density corresponded to the helical regions of  $\alpha 3$ - $\alpha 4$  and that sub-stoichiometric binding meant many binding sites were unoccupied, giving rise to poorly resolved density where the linker binds. Using a symmetry expansion approach<sup>35</sup>, with a soft featureless mask around the additional density (Extended Data Fig. 15a), we found  $\sim 23\%$  of the 8 CsRbcL sites to be occupied by additional density in the symmetry expanded sub-particles (Extended Data Fig. 15c,f). By focusing on the sub-particles bound by  $\alpha 3$ - $\alpha 4$ , we obtained a map of the additional density which possessed clearly helical nature (Fig. 3d, Extended Data Fig. 15e; PDB: 8Q05). We built the helical region of  $\alpha 3$  into the density and numbered residues according to their position in the full-length protein (Fig. 1g, Fig. 3e, Extended Data Fig. 15h,i). Each CsLinker binding site is contained within the N-terminal region of a single CsRbcL subunit, utilizing two salt bridges on a hydrophobic interface (Fig. 3e,f and Supplementary Table 13). The two salt bridges are between Arg176 of  $\alpha 3$ - $\alpha 4$  and Glu51 of the CsRbcL, and between Lys177 of  $\alpha 3$ - $\alpha 4$  and Asp86 of the third beta-sheet ( $\beta\text{C}$ ) in CsRbcL (Extended Data Fig. 15j). Phe173 of  $\alpha 3$ - $\alpha 4$  dominates the hydrophobic interaction in a pocket formed by Ile87 and Tyr97 of the CsRbcL. As the  $\alpha 3$ - $\alpha 4$  fragment contains a single residue difference between  $\alpha 3$  and  $\alpha 4$  (Fig. 1g), the residue at helix position 7 could be either a Leu (174)

or Met (240). The map resolution did not allow distinction between the two, though when Leu174 was built into this density, it was also positioned to contribute to the hydrophobic interaction. We validated the binding interface by site-directed mutagenesis (SDM) of the  $\alpha 3$ - $\alpha 4$  fragment; selecting substitutions that disrupted the hydrophobic and salt bridge interfaces. In both cases, disruption of  $\alpha 3$ - $\alpha 4$ -CsRubisco complex formation by native PAGE was observed (Fig. 3g, Extended Data Fig. 16c).



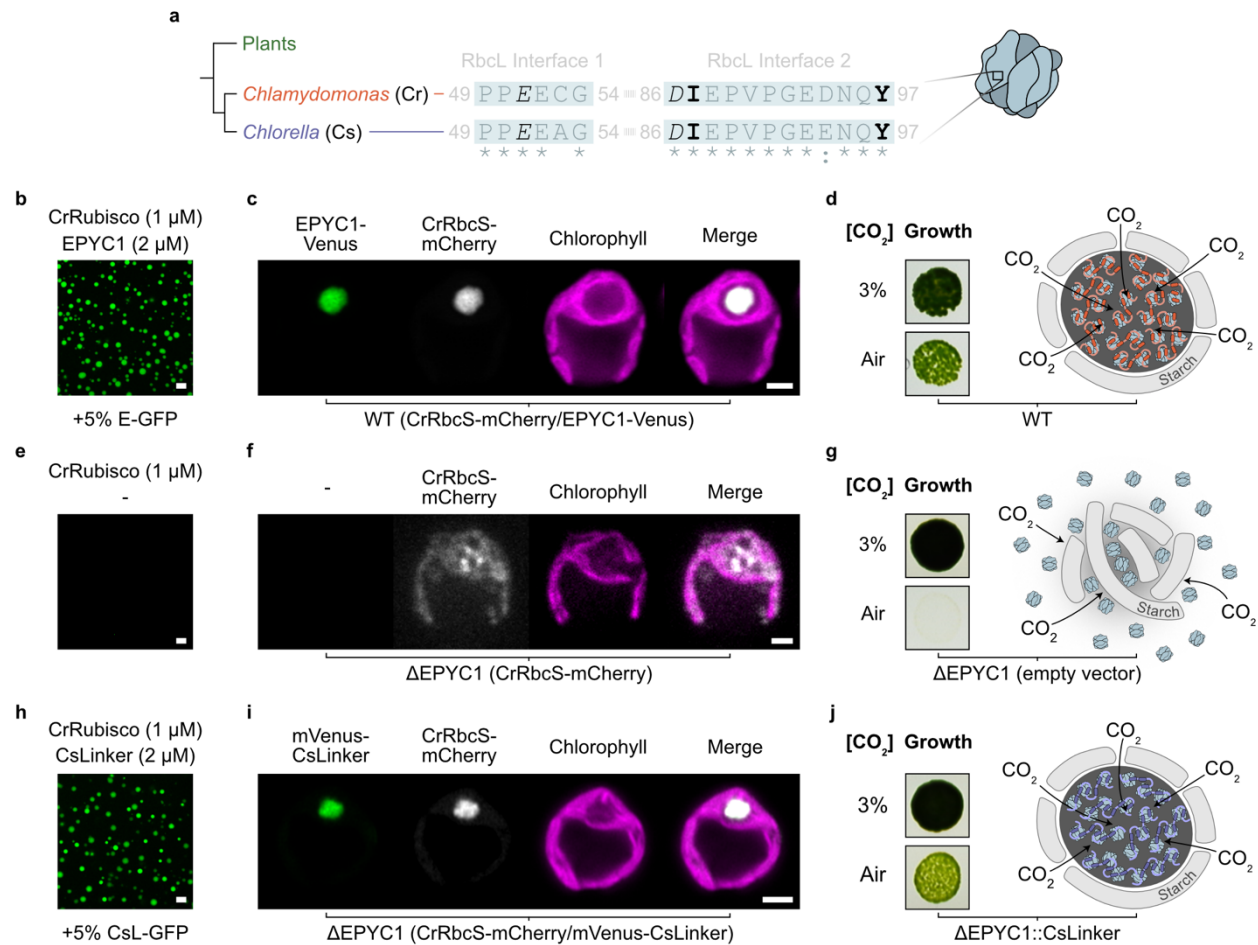
**Fig. 3 | CsLinker binds to the Rubisco large subunit (RbcL).** **a**,  $\alpha 3$ - $\alpha 4$  forms a stable complex with CsRubisco. Representative native PAGE gel shift assay (left) demonstrating formation of higher order  $\alpha 3$ - $\alpha 4$ -CsRubisco complex. Quantification of replicate data (right) taken from Extended data Fig. 12h;  $n=2$ , error bars = S.D. **b**, Schematic of surface plasmon resonance (SPR) experiments (left) used to determine binding affinity of  $\alpha 3$  and  $\alpha 3$ - $\alpha 4$  for CsRubisco (right). SPR response normalized to  $B_{\text{max}}$  value obtained from fit of raw data;  $n=3$ , error bars = S.D. **c**, Top (left) and side (right) views of surface representations of cryo-EM determined structure of the  $\alpha 3$ - $\alpha 4$ -CsRubisco complex. The modelled  $\alpha$ -helix of  $\alpha 3$  is superimposed on each CsRbcL based on the coordinates built into one axis of the C1 map, as shown in **d**. **d**, Density map of the  $\alpha 3$ - $\alpha 4$  region in the C1 complex map, carved with a radius of 2 from the built coordinates, at a contour level of 0.033. **e**, Molecular interactions at the interface. Shortest range electrostatic interactions highlighted by PDBePISA (left) and residues contributing to the hydrophobic interface (right). The surface of CsRbcL is colored according to hydrophobicity. Residues are numbered according to their position in the full length CsLinker. **f**, Map of the interactions between the  $\alpha$ -helices of  $\alpha 3$ - $\alpha 4$  and the two interfaces on CsRbcL. Dashes indicate salt bridges, with wedges representing significant contributions to the hydrophobic interface, with italicised and bolded residues contributing the same interfaces, respectively. **g**, Native PAGE gel shift assays showing that mutation of  $\alpha 3$ - $\alpha 4$  disrupts binding to CsRubisco. The sequence of the  $\alpha$ -helices in each fragment is provided above each image.



# CsLinker complements pyrenoid formation in *Chlamydomonas*

The Rubisco large subunit (RbcL) is highly conserved across the green lineage<sup>36</sup>, in contrast to the small subunit (RbcS) which shows much higher sequence variation<sup>37</sup> (Extended Data Fig. 17). Given the interaction of CsLinker with the RbcL (Fig. 3c), we wondered whether this affords CsLinker increased cross-reactivity for other Rubiscos relative to EPYC1, which binds the RbcS<sup>32,38</sup>. We first considered cross-reactivity of CsLinker between *Chlorella* and *Chlamydomonas* Rubiscos, where the CsLinker interaction interfaces are almost totally conserved (Fig. 4a). CsLinker phase separated *Chlamydomonas* Rubisco (CrRubisco) with a similar efficiency as its cognate CsRubisco (Fig. 4h), and with a similar efficiency as EPYC1 for its cognate CrRubisco (Fig. 4b, Extended Data Fig. 21a,d). By contrast, EPYC1 was unable to demix CsRubisco in the reciprocal experiment (Extended Data Fig. 21a,c), in line with the lack of conservation of EPYC1-interacting residues in CsRbcS (Extended Data Fig. 17). SPR experiments with  $\alpha 3$  and  $\alpha 3$ - $\alpha 4$  showed comparable  $K_D$  values for CrRubisco as the cognate interaction ( $108\ \mu\text{M}$  [ $\text{CI}_{95}$ :  $102$ – $115\ \mu\text{M}$ ] and  $1.30\ \mu\text{M}$  [ $\text{CI}_{95}$ :  $1.16$ – $1.46\ \mu\text{M}$ ] respectively; Extended Data Fig. 13). A similar  $\alpha 3$ - $\alpha 4$ -CrRubisco complex was also observed by native PAGE, albeit with apparently less stability (Extended Data Fig. 18c). Together, these data are consistent with CsLinker binding the same conserved interface in both Rubiscos.

LLPS of Rubisco in the *Chlamydomonas* pyrenoid is essential to its function, allowing growth at atmospheric  $\text{CO}_2$  levels<sup>15</sup> (Fig. 4d,g). Given the *in vitro* cross-reactivity of CsLinker and the functional analogy with EPYC1, we next considered whether CsLinker could replace EPYC1 in the pyrenoid of *Chlamydomonas*. We utilized a previously characterized *Chlamydomonas* strain which lacks EPYC1 ( $\Delta\text{EPYC1}$ ) and accordingly does not form pyrenoids. This strain has Rubisco distributed throughout the chloroplast and exhibits reduced growth under ambient  $\text{CO}_2$  conditions (air)<sup>15</sup> (Fig. 4f,g). We expressed mVenus-tagged CsLinker in the chloroplast of  $\Delta\text{EPYC1}$ , and subsequently co-expressed mCherry-tagged CrRbcS in the resulting strain to create a  $\Delta\text{EPYC1}$  (CrRbcS-mCherry/mVenus-CsLinker) line. In line with our observations *in vitro* (Fig. 4h), mVenus-CsLinker expression led to the formation of a micron-scale condensate at the canonical pyrenoid position that contained both CrRbcS and CsLinker (Fig. 4i). Although Rubisco partitioning and condensate size was reduced compared with WT, it was significantly increased over the background  $\Delta\text{EPYC1}$  strain, suggesting *in vivo* condensation of CrRubisco by CsLinker (Extended Data Fig. 19). To avoid any phenotypic impact of tagging either CsLinker or CrRubisco, we expressed untagged CsLinker in the  $\Delta\text{EPYC1}$  background and confirmed expression by western blotting (Extended Data Fig. 19i). In line with the visual recovery of Rubisco condensation by mVenus-CsLinker, introduction of untagged CsLinker restored the growth of the resulting  $\Delta\text{EPYC1}::\text{CsLinker}$  strain in air to almost wild-type levels (Fig. 4j and Extended Data Fig. 20). The functional complementation of EPYC1 by CsLinker presents a compelling example of functional LLPS-driven organelle assembly complementation by a protein with little sequence similarity and a different binding interface, across an approximately 800 My evolutionary gap.



**Fig. 4 | CsLinker can functionally replace EPYC1 in the *Chlamydomonas* pyrenoid.** **a**, Alignment of the CsLinker-binding interface sequences from *Chlorella* (Cs) RbcL and the equivalent region of the *Chlamydomonas* (Cr) RbcL. Interacting residues are shown in black and stylized by interaction type according to Fig. 3f. Conserved residues are indicated by asterisks. **b**, Confocal fluorescence microscopy image of droplets formed with *Chlamydomonas* Rubisco (CrRubisco) and EPYC1, in which 5% (molar ratio) of the EPYC1 was GFP-tagged (E-GFP). Scale bar = 5  $\mu$ m. **c**, Confocal fluorescence microscopy image of WT *Chlamydomonas* (CC-4533) expressing EPYC1-Venus and CrRbcS-mCherry. Scale bar = 2  $\mu$ m. **d**, Growth phenotype of WT *Chlamydomonas* grown on TP minimal media under elevated (3%) and ambient (Air) levels of CO<sub>2</sub> (left). Schematic representation of the pyrenoid (right). **e**, Confocal fluorescence microscopy image of CrRubisco alone, scale bar = 5  $\mu$ m. **f**, Confocal fluorescence microscopy image of  $\Delta$ EPYC1 *Chlamydomonas* strain expressing CrRbcS-mCherry. Scale bar = 2  $\mu$ m. **g**, Growth phenotype of  $\Delta$ EPYC1 *Chlamydomonas* strain. Schematic representation of pyrenoid region (right). **h**, Droplets formed with *Chlamydomonas* Rubisco (CrRubisco) and CsLinker, in which 5% (molar ratio) of the CsLinker was GFP-tagged (CsL-GFP). Scale bar = 5  $\mu$ m. **i**, Confocal fluorescence microscopy image of  $\Delta$ EPYC1 *Chlamydomonas* strain expressing CrRbcS-mCherry and mVenus-CsLinker. Scale bar = 2  $\mu$ m. **j**, Growth phenotype of  $\Delta$ EPYC1 *Chlamydomonas* strain complemented with untagged CsLinker. Schematic representation of pyrenoid (right).

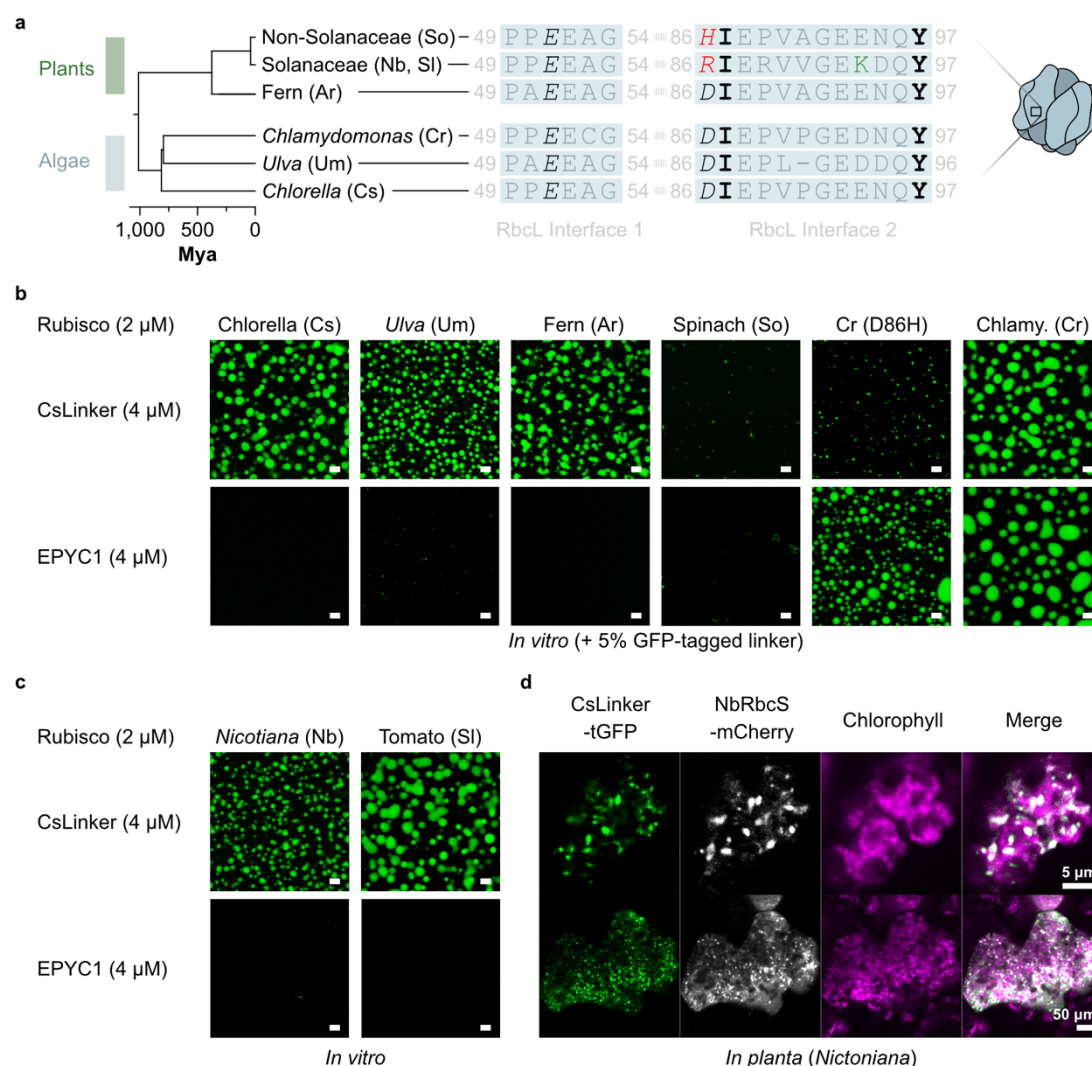
## CsLinker condenses native plant Rubisco *in vitro* and *in planta*

Encouraged by the functional cross-reactivity of CsLinker we observed with *Chlamydomonas* Rubisco, we sought to understand the extent of cross-reactivity for Rubiscos in the green lineage. We next demonstrated CsLinker was able to demix Rubisco from the multicellular Ulvophyte seaweed, *Ulva mutabilis* (Um), which retains all four CsLinker-interacting residues (Fig. 5a). Notably, the efficiency of phase separation was lower (Extended Data Fig. 21), SPR experiments demonstrated higher  $K_D$  values of the  $\alpha 3$  and  $\alpha 3$ - $\alpha 4$  fragments for UmRubisco (162  $\mu$ M [ $CI_{95}$ : 158–167  $\mu$ M] and 1.56  $\mu$ M [ $CI_{95}$ : 1.40–1.72  $\mu$ M] respectively; Extended Data Fig. 13d), and native PAGE assays showed little gel-shift (Extended Data Fig. 18f). We attribute the reduced affinity and concomitant phase separation efficiency to the reduction of the *Ulva* RbcL interface 2 by one residue (Fig. 5a). This change would disrupt a potential hydrogen bond network with the Gln residue of the CsLinker helices, though the resolution in this region of the  $\alpha 3$ - $\alpha 4$ -CsRubisco complex map did not allow distinctive assignment of this network (Extended Data Fig. 15k, 22c).

Motivated by the wider goal of engineering pyrenoids in  $C_3$  angiosperm crop plants to enhance  $CO_2$ -fixation<sup>18,19</sup>, we next wondered whether the cross-reactivity of CsLinker extends to plant Rubiscos. We analyzed RbcL sequences from major plant groups for conservation of CsLinker-interacting residues and found early diverged, non-flowering ferns were phylogenetically the closest plant group to angiosperms to contain all four key residues (Fig. 5a, Extended Data Fig. 22a). Accordingly, Rubisco from the fern *Adiantum raddianum* (Ar) demonstrated phase-separation with CsLinker *in vitro* (Fig. 5b), with a similar efficiency to the cognate pairing (Extended Data Fig. 21). Most angiosperm crops have a substitution of Asp86 for His86 in their RbcLs (Extended Data Fig. 22b). We tested the effect of this substitution (D86H) using Rubisco purified from Spinach (*Spinacia oleracea*; So), which has RbcL interfaces representative of the consensus angiosperm sequence (Extended Data Fig. 22a,b). We observed sub-micron aggregates of Spinach Rubisco at the same concentrations used for the cognate CsLinker interaction, but never observed droplet formation (Fig. 5b and Extended Data Fig. 21). To confirm this effect was predominated by the D86H substitution, we made the same substitution in the RbcL of CrRubisco that we previously demonstrated phase separates with CsLinker (Fig. 4h, Fig. 5b). The D86H mutation disrupted droplet formation of CrRubisco, which instead demonstrated similar behavior to the Spinach Rubisco (Fig. 5b). Notably, the D86H mutated CrRubisco still phase-separated with its cognate linker EPYC1 (Fig. 5b). Complex formation of D86H CrRubisco with  $\alpha 3$ - $\alpha 4$  was also disrupted, as assessed by native PAGE (Extended Data Fig. 18i). These data suggest substitution of the D86 residue severely impacts the affinity and phase separation propensity of CsLinker for Rubiscos of most angiosperm plants which possess a His in this position, including most of the key  $C_3$  crop plants (rice, wheat, soybean) (Extended Data Fig. 22b).

Amongst angiosperm RbcL sequences, there is some variation in the sequences of the CsLinker-interacting interfaces (Extended Data Fig. 22b). RbcLs of the nightshade family (Solanaceae), which contains some of the most widely consumed plants (potato, tomato, eggplant, pepper, tobacco), possess a distinct sequence composition at interface 2, due to co-evolution with their specific Rubisco activase which also binds this region<sup>39</sup>. Solanaceae RbcLs possess an Arg at position 86 whilst retaining the other three key interacting residues, but also display other unique sequence features in interface 2 (Fig. 5a). Given the charge inversion at position 86 (D86R) and presumed subsequent salt bridge disruption (Fig. 3e), we predicted Solanaceae Rubiscos would not phase separate with CsLinker. Surprisingly, Rubisco from the model Solanaceous plant *Nicotiana benthamiana* (NbRubisco; Tobacco) was readily phase-separated *in vitro* (Fig. 5c). As this behavior was specific to CsLinker (and not EPYC1), we hypothesized the interaction between CsLinker and NbRbcL is based on a similar binding interface as CsRbcL. Analysis of the Rubisco structure from *Nicotiana tabacum* (PDB: 1EJ7<sup>40</sup>) indicated that Lys94 of the RbcL (green in Fig. 5a) is favorably positioned to form a salt bridge with Glu169 of the helix of CsLinker (Extended Data Fig. 22e), that could compensate for the loss of the D86 salt bridge. Given that

all key nightshade plants have conserved CsLinker-interacting interfaces (Extended Data Fig. 22b), we were interested whether CsLinker could phase separate Rubisco from other widely consumed members of this family. In line with the results of *Nicotiana*, Tomato Rubisco (*Solanum lycopersicum*) readily demixed with CsLinker *in vitro* (Fig. 5c and Extended Data Fig. 21). Finally, to explore if the *in vitro* observation of Solanaceae Rubisco condensation extended *in planta*, we transiently co-expressed TurboGFP-tagged CsLinker and mCherry-tagged NbRbcS in *Nicotiana benthamiana*. Strikingly, micron-scale condensates were observed in each chloroplast, which contained both Rubisco and CsLinker, representing the first phase separation of native plant Rubisco *in planta* (Fig. 5d and Extended Data Fig. 23). The condensation of native plant Rubisco represents a significant step forward in the goal of engineering pyrenoids in plants and provides prospect that ongoing engineering of CsLinker could allow phase separation of native non-Solanaceae crop Rubiscos.



**Fig. 5 | CsLinker condenses native plant Rubisco *in vitro* and *in planta*.** **a**, Alignment of the CsLinker-binding interface sequences from algal and plant RbcLs. Interacting residues are shown in black and stylized by interaction type according to Fig. 3f. Substitutions of interacting residues are shown in red. **b**, Confocal fluorescence microscopy images of droplets formed with different Rubiscos and either CsLinker or EPYC1. Scale bars = 5  $\mu$ m. **c**, Images of droplets formed with Solanaceae Rubisco and either CsLinker or EPYC1. **d**, Confocal fluorescence microscopy images of CsLinker-tGFP and NbRbcS-mCherry transiently expressed in *Nicotiana benthamiana* chloroplasts, condensed into Rubisco puncta *in planta*.



# Discussion

Our in-depth characterization of a pyrenoid linker protein from *Chlorella* has yielded significant insight into LLPS driven organelle assembly and provided exciting frontiers for future plant pyrenoid engineering approaches to enhance photosynthesis.

Identification that CsLinker binds to the RbcL and structural characterization of the RbcL binding site has allowed us to make three general observations. Firstly, green lineage pyrenoid linker proteins likely evolved separately and convergently, as previously proposed<sup>15,32,41</sup>, and their binding region was not constrained to the small subunit of Rubisco. Their different sequence composition, binding affinity and binding region suggests the general physical properties of linkers also likely converged at linker-specific optima, which is supported by ongoing modelling efforts<sup>42,43</sup>. Secondly, given that *Chlorella* and *Chlamydomonas* diverged approximately 800 Mya, the capacity to functionally exchange EPYC1 for CsLinker in the *Chlamydomonas* pyrenoid provides a compelling example that the conservation of functionally critical physicochemical properties in intrinsically disordered proteins can be more important than conservation of a specific primary sequence. This example showcases the algal pyrenoid as a tractable model to explore the evolution of biomolecular condensation and LLPS evolution *in vitro* and *in vivo* with clear biological fitness readouts – often a limitation in LLPS systems<sup>31</sup>. We anticipate that our work will enable future studies to systematically test how the physical properties of phase-separating proteins (e.g. sticker number, sticker binding affinity, spacer length and spacer flexibility<sup>44</sup>) impact biological fitness. Thirdly, the high conservation of the CsLinker binding site in the green lineages RbcLs, enabled cross-reactivity of CsLinker with plant Rubiscos and allowed us to demonstrate phase separation of unmodified plant Rubisco for the first time. This finding overcomes a major future barrier for pyrenoid engineering in plants, which is currently dependent on genetic replacement of the multiple host plant RbcS proteins with those of *Chlamydomonas*<sup>23,24,45</sup>.

Whilst our results are encouraging, we acknowledge several areas that will require future work to address. Firstly, we presume that our demonstration of EPYC1 replacement in *Chlamydomonas* is dependent on retention of the native RbcS which is responsible for organizing other essential pyrenoid features through interaction with Rubisco-binding motif (RBM)-containing proteins<sup>21,46</sup>. Future approaches to engineer functional pyrenoids in plants without Rubisco engineering will be dependent on the development of approaches that circumvent RbcS interaction entirely, either by replacing RBMs in *Chlamydomonas* proteins or by identifying and characterizing analogous RbcL-binding parts, possibly from *Chlorella*. Secondly, although *in vivo* expression of CsLinker in both *Chlamydomonas* and *Nicotiana* resulted in Rubisco condensation, the partitioning of Rubisco in the condensate was lower than WT and multiple condensates were observed *in planta*. Whilst the level of CsLinker expression likely plays an important role, these observations also suggest a degree of tunability in the phase separation of Rubisco that is dependent on specific features of the interaction proteins when the interface is not fully conserved. Future studies could likely exploit this tunability to expand the cross-functionality of CsLinker with other non-Solanaceae Rubiscos, making use of the molecular details of the interaction outlined here.

Moving forward, as additional pyrenoids and their corresponding assembly proteins from diverse algae are characterized we expect our understanding of their evolution, the underlying principles of pyrenoid assembly and our ability to model pyrenoid systems to rapidly advance. We envision that this knowledge will provide an expanded parts list for pyrenoids and give us new tools to predict and modulate pyrenoid properties and thereby accelerate future efforts to engineer pyrenoids in plants.

## Methods

### Strains and culture conditions

*Chlorella sorokiniana* UTEX1230 (SAG 211-8k), *Chlamydomonas reinhardtii* WT (CC-4533), ΔEPYC1 (CC-5360) and resulting strains were maintained on 1.5% agar Tris-acetate-phosphate (TAP) medium with revised trace elements<sup>47</sup> plates in low light (~10 μmol photons m<sup>-2</sup> s<sup>-1</sup>). For Rubisco extraction, growth, immunoelectron microscopy and confocal microscopy experiments, strains were grown in Erlenmeyer flasks in Tris-phosphate (TP) media in medium light (~50 μmol photons m<sup>-2</sup> s<sup>-1</sup>) under ambient CO<sub>2</sub> to a density of ~0.5-1x10<sup>7</sup> cells mL<sup>-1</sup>.

### FLIPper and bioinformatic analysis

FLIPper was originally built using IUPred2A<sup>48</sup> as the disorder prediction and filtering software, which was used in the initial identification of CsLinker. For licensing reasons, the pipeline was subsequently rebuilt using metapredict V2<sup>49</sup> instead and the outputs were largely unchanged. FLIPper is available ([github.com/james-r-barrett/FLIPper](https://github.com/james-r-barrett/FLIPper)) and was used with default settings thresholded based on EPYC1 sequence analysis. Briefly, input sequences are physicochemically filtered, repeats detected with XSTREAM<sup>50</sup> and filtered to contain interacting residues before disorder prediction and filtering with metapredict. Output sequence structures were manually examined for repeated helical regions by AlphaFold 2<sup>51</sup> prediction in ColabFold v1.5<sup>52</sup>. Differential gene expression analysis was completed according to ref.<sup>53</sup>, using publicly available data (PRJNA343632).

### Co-Immunoprecipitation Mass Spectrometry

Polyclonal rabbit antibodies were raised to Rubisco (EVWKEIKFEFETIDTL-cooh) and CsLinker (PTPVSNSGVRSAMSSG-amide) peptides (YenZym Antibodies LLC, USA). The control rabbit antibody was raised to *Chlamydomonas* BST2 (PDLDSINAAAPNGNGSHNGN-amide). 1x10<sup>9</sup> *Chlorella* cells grown in TP sparged with 0.01% CO<sub>2</sub> were lysed by ultrasonication in 10 mL IP buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl, 0.1 mM EDTA, 12.5% Glycerol (w/v), 5 mM DTT, 1x cOmplete protease inhibitor tablet / 50 mL). 1 mL of clarified lysate (30 minutes at 50,000 g) was applied to 200 μL of protein A Dynabeads™ (Invitrogen) loaded with 32 μg of respective antibodies that had been blocked with BSA (2 mg mL<sup>-1</sup>) for 1 hour and washed 3 times with IP buffer. The lysate was incubated for 3 hours before washing. Bound protein was trypsin-digested from the beads, post reduced with DTE, and alkylated with iodoacetamide prior to UPLC separation by 60SPD EvoSep One (EvoSep, Denmark) method and acquisition by PASEF-DIA using a timsTOF HT (Bruker, USA). Data were searched using DIA-NN<sup>54</sup> and filtered to 1% FDR with a minimum of two unique peptides.

### Immunoelectron Microscopy

Cells were fixed in 1% glutaraldehyde, 2% formaldehyde in 0.1 M sodium cacodylate (pH 7.4) for 1 hour before resuspension to 2% (w/v) agar. Agar blocks were dehydrated using a 10-50% ethanol gradient at 4°C, followed by 70-100% at -20°C, then infiltrated with LR White resin containing 0.5% benzoin methyl ether (London Resin Company, UK) and polymerised in gelatin capsules for 24 hours at -20°C and 24 hours at -10°C under UV light. 70 nm sections were cut using a Leica UCT7 ultramicrotome with a Diatome knife and mounted on nickel grids. All imaging was completed using an FEI Tecnai T12 BioTWIN TEM operating at 120kV with Ceta CCD camera.

For immunolabelling, grids were blocked with 3% BSA (w/v) in PBS for 30 minutes prior to primary incubation with either purified antibody (0.09 μg mL<sup>-1</sup>) or pre-immune serum (1:5,000 dilution) for 1 hour at 30°C in a humidity chamber. Secondary incubation was completed with a 1:40 dilution of Goat anti-Rabbit IgG 10 nm gold conjugate for a further hour (Merck).

### Absolute quantification mass spectrometry

6x10<sup>6</sup> *Chlorella* cells grown in TP sparged with 0.01% CO<sub>2</sub> were boiled in 50 μL of Laemmli buffer in triplicate for the biological samples. The standards were separately boiled in duplicate in the same volume. Both the samples and standards were trypsin-digested from SDS-PAGE gels, post reduced with DTE, and alkylated with iodoacetamide prior to separation by UPLC with a 25 cm PepMap column (ThermoFisher Scientific) and 1 hour acquisition by DDA using an Orbitrap Fusion Lumos Tribrid (ThermoFisher Scientific). Technical replicate injections were used for the biological samples, after which the values were averaged. LC-MS chromatograms were aligned using Progenesis Q1 and the MS2 spectrum was searched using Mascot with a 1% FDR. Matches were mapped to MS1 intensity using Progenesis Q1 and summed at protein level. The external calibration curve was used to calculate protein abundance in the biological samples (Extended Data Fig. 7).

### Cloning, plasmids, and strains

For *E. coli*, the mature CsLinker sequence was predicted by TargetP 2.0<sup>55</sup>, codon-optimised in Geneious Prime using the *E. coli* K-12 codon usage table and synthesised (TWIST Bioscience). The sequence was ligation-independent cloned (LIC) into pET His6 GFP TEV LIC cloning vector (addgene #29663) to produce the His-mEGFP-TEV-CsLinker plasmid. The mEGFP-CsLinker fusion was subsequently PCR-amplified and LIC cloned into pET MBP His6 LIC cloning vector (addgene #37237) to produce His-mEGFP-TEV-CsLinker-TEV-MBP-His plasmid. α3 (residues 142-207) and α3-α4 (residues 142-272) fragments were PCR-amplified from this plasmid and Gibson assembled back into the PCR-amplified backbone to produce His-mEGFP-TEV-α3-TEV-MBP-His and His-mEGFP-TEV-α3-α4-TEV-MBP-His respectively. SDM constructs were constructed by introducing nucleotide exchanges in the primers used to amplify the α3-α4 fragment from His-mEGFP-TEV-α3-α4-TEV-MBP-His. The SDM

sequences were Gibson-assembled back into the backbone to create His-mEGFP-TEV- $\alpha$ 3(F173DL174D)- $\alpha$ 4(F239DM240D)-TEV-MBP-His and His-mEGFP-TEV- $\alpha$ 3(E169AR176A)- $\alpha$ 4(E235AR242A)-TEV-MBP-His constructs.

For *Chlamydomonas* chloroplast expression, the mature CsLinker was codon optimised in Geneious Prime using the *Chlamydomonas* chloroplast codon usage table (Kazusa) and synthesised (TWIST Bioscience). The mVenus sequence was codon optimised using Codon Usage Optimizer v0.92 (<https://github.com/khai-/CUO>). Sequences were PCR-amplified and golden gate assembled into pME\_Cp\_2\_098 (Gift from René Inckemann and Tobias Erb). The  $\Delta$ RbcL plasmid was created by Gibson assembly of the *aadA* gene amplified from CC-5168 into plasmid P-67 cpDNA EcoRI 14 (Chlamydomonas Resource Centre) to replace the RbcL CDS in frame. The RbcL\_D86H plasmid was created by KLD site-directed mutagenesis (New England Biolabs) of the P-67 cpDNA EcoRI 14 plasmid (Chlamydomonas Resource Centre).

For *Nicotiana benthamiana* transient expression, 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 and HSP+NOS double terminator according to<sup>56</sup>. The NbRbcS was synthesised with AtRbcS1A transit peptide by IDT and assembled into pICH47751 with AtRbcS1A promoter, C-terminal mCherry tag and OCS terminator.

### Rubisco Extraction

Rubisco was purified from *Chlamydomonas* and *Chlorella* according to ref.<sup>57</sup> with the addition of a 16.5 hour, 37,000 rpm 10-30% sucrose gradient ultra-centrifugation step performed in an SW41-Ti rotor prior to anion exchange using a HiTrap 5 mL Q XL column (Cytiva). The same method was used for *Ulva*, Spinach, *Adiantum*, *Nicotiana* and Tomato, except lysis was completed by manual agitation in a blender. Rubisco was labelled using an Atto594 protein labelling kit (Sigma-Aldrich), according to manufacturer's instructions.

### E. coli protein purification

All constructs were purified from *E. coli* BL21 (DE3) strains harbouring respective plasmids. Cells were grown to OD<sub>600</sub> of 0.5-0.8 in Luria Broth or <sup>15</sup>NH<sub>4</sub>Cl minimal media for the NMR samples, before induction with 1 mM isopropyl  $\beta$ -D-thiogalactopyranoside (IPTG) for 3 h at 37 °C. Pellets were snap-frozen before ultrasonic lysis in high salt buffer (50 mM Tris-HCl, 500 mM NaCl, 25 mM imidazole, pH 8.0) with 2 mM phenylmethanesulfonylfluoride (PMSF). Soluble protein was applied to an IMAC column (HisTrap™ FF Crude 5 mL - Cytiva), washed with high salt buffer and eluted with a linear gradient to 500 mM imidazole in high salt buffer. For untagged CsLinker,  $\alpha$ 3,  $\alpha$ 3- $\alpha$ 4,  $\alpha$ 3(F173DL174D)- $\alpha$ 4(F239DM240D) and  $\alpha$ 3(E169AR176A)- $\alpha$ 4(E235AR242A) constructs, the N-terminal mEGFP and C-terminal MBP were cleaved overnight with TEV protease produced according to refs.<sup>58</sup>. The cleaved solution was passed over an IMAC column equilibrated with high salt buffer and the untagged flow-through was collected. Size-exclusion chromatography (SEC) was completed on the flow-through using a HiLoad® 16/600 Superdex® 75 pg (Cytiva) equilibrated with 50 mM Tris-HCl, 500 mM NaCl, pH 8.0 buffer. For the mEGFP-CsLinker, no TEV cleavage or second IMAC was completed, but protein was exposed to SEC.

### Droplet sedimentation assay

All assays were completed in 5  $\mu$ L reaction volumes in which Rubisco and linker were mixed and incubated at room temperature for 15 minutes. Sedimentation was performed at 10,000 g for 10 minutes. Band intensity was quantified in Fiji.

### In vitro confocal microscopy and fluorescence recovery after photobleaching (FRAP)

All *in vitro* confocal and FRAP experiments were completed on either a Zeiss LSM880 or Zeiss LSM980 confocal microscope with a 63x 1.4 numerical aperture (NA) Plan-Apo oil-immersion lens (Carl Zeiss, Germany) in 5  $\mu$ L reaction volumes formed in  $\mu$ -Slide 15 Well 3D coverslips (ibidi, Germany). For FRAP experiments, the sample volumes were overlaid with 30  $\mu$ L of ibidi Anti-Evaporation oil.

FRAP experiments were completed on droplets with diameter  $\sim$ 1  $\mu$ M, where half of the droplet was bleached in half FRAP experiments. 15 pre-bleach images were taken before bleaching (100% 488 nm intensity, 1 cycle for mEGFP, 100% 488 nm + 561 nm intensity, 5 cycles for Atto594). Bleach depth was consistently 60-75%. FRAP images were processed in Fiji. Briefly, fluorescence images were translationally stabilised using the Image Stabilizer plugin (4 pyramid levels, 0.99 template update coefficient) output of the brightfield images and the mean gray value in the bleached, unbleached and background ROIs was measured. Background values were subtracted from bleached and unbleached values before photobleach normalisation using the unbleached references was completed. Full-scale normalisation was completed using the average pre-bleach intensity. An exponential model was fitted to the post-bleach data ( $y(t) = A * (1 - e^{-kt})$ ), where A is the plateau, k is fitted and t is post-bleach time).

### Native PAGE gel-shift assay

Rubisco and fragments were mixed in 5  $\mu$ L reaction volumes and incubated at room temperature for 30 minutes. 1.6  $\mu$ L of loading buffer (80 mM Tris-HCl pH 8.0, 200 mM NaCl, 40% glycerol) was added prior to loading onto 4-20% Mini-PROTEAN® TGX™ gels. Electrophoresis was completed for 4 hours at 100 V at 4 °C.

### Surface plasmon resonance (SPR)

All SPR experiments were completed in triplicate on a Biacore™ T100 fitted with a T200 upgrade kit with a sensor temperature of 25 °C. Immobilisation of Rubisco was completed according to ref.<sup>32</sup> and experiments were completed with modifications. The analyte was injected at 15  $\mu\text{L min}^{-1}$  for 30 seconds followed by 360 seconds of dissociation. After each replicate set the chip was washed with 1M NaCl in running buffer, after which the chip was washed for 360 seconds with running buffer. Binding to the reference chip was negligible. Fitting of the reference-subtracted curves was completed using hill equation ( $y(x) = B_{\text{max}} \cdot x / (K_D + x)$ ), where  $B_{\text{max}}$  is the maximum specific binding,  $K_D$  is the apparent dissociation constant, and  $x$  is the analyte concentration.

### Single-particle cryogenic electron microscopy data collection and image processing

Chlorella Rubisco and the  $\alpha 3$ - $\alpha 4$  fragment were mixed in a buffer compatible with cryo-EM experiments (200 mM sorbitol, 50 mM HEPES, 50 mM KOAc, 2 mM  $\text{Mg}(\text{OAc})_2 \cdot 4\text{H}_2\text{O}$  and 1 mM  $\text{CaCl}_2$  at pH 6.8) at final concentrations of 0.5  $\mu\text{M}$  and 16  $\mu\text{M}$  respectively and incubated at room temperature for 10 minutes.  $\alpha 3$ - $\alpha 4$  had the same apparent  $K_{0.5}$  for CsRubisco in this buffer as the Tris buffer used for native PAGE experiments (Extended Data Fig. 12f). 2.5  $\mu\text{L}$  of solution was applied to R1.2/1.3 Cu 400-mesh grids (Quantifoil, Germany) that had been glow-discharged for 60 seconds with a current of 15 mA in a PELCO easiGlow system. Using an FEI Mark IV Vitrobot (ThermoFisher Scientific) with chamber at 4 °C and 95% relative humidity, the grids were blotted for 8 seconds with a blot force of -5 before rapid plunge-freezing in liquid ethane.

Data was collected on a 200 kV Glacios cryo-electron microscope equipped with a Falcon IV direct electron detector at the University of York. Automated data collection was performed using EPU in AFIS mode (ThermoFisher Scientific). A nominal magnification of 240,000x and electron fluence of 50 ( $\text{e}^-/\text{\AA}^2$ ) with a calibrated pixel size of 0.574  $\text{\AA}^2$  was used during collection in which each exposure was 6.52 seconds. The 100  $\mu\text{m}$  objective aperture was inserted and the C2 aperture was 50  $\mu\text{m}$ . A range of defocus value were used (-1.8, -1.6, -1.4, -1.2, -1.0, -0.8  $\mu\text{m}$ ).

Relion3 was used for processing and 3D reconstruction<sup>59,60</sup>. Of the 1568 EER frames, 32 were grouped to give a fluence/frame of 1.02  $\text{e}^-/\text{\AA}^2$ . Relion's implementation of MotionCor2 was used for motion correction prior to CTF estimation with CTFFIND4, assuming a spherical aberration of 2.7 mm<sup>61</sup>. Initially 465 particles were manually picked and reference-free 2D classification was completed. 237,035 particles were autopicked using selected 2D class averages. Particles were extracted with 2x binning and 2D classification was re-completed. 2D classes presenting clear Rubisco structure were selected and a C1 symmetry 3D classification was completed. A single 3D class with clear secondary structures was used for auto-refinement with D4 symmetry (73,962 particles). After CTF refinement and Bayesian polishing in Relion<sup>62</sup> a 20  $\text{\AA}$  low-pass filtered mask of the 3D refined map, expanded by 10 pixels with a soft edge of 6 pixels was used for solvent masking and resolution estimation. This map has D4 symmetry and represented the CsRubisco holoenzyme (EMD-18049).

In the 3D class used for the D4 map, additional low-resolution density was observed on the equator of Rubisco, suggesting a sub-stoichiometrically bound partner (Extended Data Fig. 14d). The predicted helix of  $\alpha 3$  (Fig. 1f) was built into one region of additional density in UCSF Chimera (Extended Data Fig. 15a)<sup>63</sup>. The soft, featureless mask was created from these coordinates by lowpass filtering to 20  $\text{\AA}$ , extension by 5 pixels and softening by 6 pixels. This mask was used for two rounds of C1 3D classification with a D4 symmetry expanded particle dataset created from the 73,962 polished particles used for the D4 map (591,696 effective particles). 133,171 symmetry expanded particles were used for the 3D reconstruction of the C1 CsRubisco- $\alpha 3$ - $\alpha 4$  map, which was processed as for the D4 map (EMD-18050).

### Single-particle cryogenic electron microscopy model building, fitting and refinement

A holoenzyme model of Chlorella Rubisco was built in UCSF Chimera using AlphaFold 2 structural prediction of the CsRbcL and CsRbcS sequences from *Chlorella sorokiniana* UTEX1230. This model was rigid-body fitted into the CsRubisco holoenzyme map using UCSF Chimera. Flexible fitting was performed in COOT 0.9<sup>64</sup>, using one of each CsRbcL and CsRbcS chain. Real-space refinement was completed in Phenix<sup>65</sup>. The coordinates of this refinement were applied to the other 7 CsRbcL/CsRbcS chains. For  $\alpha 3$ - $\alpha 4$  model building, the AlphaFold 2 predicted helix of  $\alpha 3$  was manually built into the additional density present in the C1 CsRubisco- $\alpha 3$ - $\alpha 4$  map in COOT. The side chains of residue E169, R171, E172, E175 and R179 were removed prior to refinement in Phenix, due to a lack of supporting density. Both the structures derived from the C1 and D4 maps were validated using MolProbity<sup>66</sup>. Figures were created in UCSF Chimera and UCSF ChimeraX<sup>67</sup>.

### NMR Spectroscopy

Spectra were recorded at 10 °C on a Bruker Advance Neo 700Mhz spectrometer equipped with a TCI Prodigy CryoProbe (Bruker). Samples were analysed in buffer containing 15 mM sodium phosphate pH 8.0, 150 mM NaCl and 5%  $\text{D}_2\text{O}$ . Protein concentrations were 920  $\mu\text{M}$  for  $\alpha 3$  and 500  $\mu\text{M}$  for  $\alpha 3$ - $\alpha 4$ . Spectra were processed using TopSpin (Bruker) and analysed using CCPN Analysis<sup>68</sup>.

### Chlamydomonas Transformation

Chloroplast transformations were completed by particle bombardment with a Biolistic PDS-1000/He Particle Delivery System (Bio-Rad). Per bombardment, 0.5 mg of 550 nm gold nanoparticles (Seashell Technologies Inc.) were incubated with 1  $\mu\text{g}$  of plasmid DNA and prepared according to manufacturer's instructions.  $1 \times 10^7$  cells were plated in a 4 cm diameter on TAP plates and placed ~9 cm below a 1100 psi rupture disk. After firing under vacuum conditions, cells were recovered for ~24 hours before re-plating to selection conditions. Transformants into the  $\Delta\text{EPYC1}$  strain for RbcL knockout and CsLinker reintroduction

were re-plated to TAP plates containing 100  $\mu\text{g mL}^{-1}$  spectinomycin under low light ( $\sim 10 \mu\text{mol photons m}^{-2} \text{ s}^{-1}$ ). Transformants from the  $\Delta\text{EPYC1}\Delta\text{RbcL}$  complementation with D86H mutated CrRbcL were plated on TP plates and recovered in 3%  $\text{CO}_2$ -air mix at  $\sim 50 \mu\text{mol photons m}^{-2} \text{ s}^{-1}$ . 16 transformants from each transformation were propagated 4 times on selection plates before checking for homoplasmic integration of the genetic material.

Transformation of mCherry-tagged *Chlamydomonas* RbcS was completed with pLM035<sup>15</sup>, using a NEPA21 electroporator according to ref.<sup>69</sup>.

#### ***Chlamydomonas* growth assays**

Spot test assays were completed according to ref.<sup>15</sup>.

#### ***Chlamydomonas* confocal microscopy**

All images were captured on a Zeiss LSM880 confocal microscope in Airyscan mode with a 63x 1.4 numerical aperture (NA) Plan-Apo oil-immersion lens (Carl Zeiss, Germany).  $\mu$ -Slide 18 Well chambered coverslips (ibidi, Germany) with 10  $\mu\text{L}$  of cell suspension and 30  $\mu\text{L}$  of 1% TP-low-melting point agarose were used for imaging.

#### **Transient expression *in planta***

Overnight cultures of electrocompetent GV3101 *Agrobacterium* harbouring the relevant plasmids were grown overnight in LB. Cultures were resuspended in 10 mM  $\text{MgCl}_2$  to  $\text{OD}_{600}$  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  $\mu\text{M photons m}^{-2} \text{ s}^{-1}$ .

### **Data availability**

Proteomics data were deposited in MassIVE, with ProteomeXchange identifier PXD044179. Electron density maps were deposited in EMDB with accession codes EMD-18049 (D4) and EMD-18050 (C1), and their corresponding coordinates in the PDB with accession codes 8Q04 and 8Q05 respectively. Code for the Fast Linker Identification Pipeline for Pyrenoids (FLIPPer) is available in a public GitHub repository (<https://github.com/james-r-barrett/FLIPPer>). All other associated data will be made available in a Zenodo repository.



## References

1. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* **281**, 237–240 (1998).
2. Bar-Even, A. *et al.* The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* **50**, 4402–4410 (2011).
3. Shih, P. M. *et al.* Biochemical characterization of predicted Precambrian RuBisCO. *Nat. Commun.* **7**, 10382 (2016).
4. Bouvier, J. W., Emms, D. M. & Kelly, S. Rubisco is evolving for improved catalytic efficiency and CO<sub>2</sub> assimilation in plants. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2321050121 (2024).
5. Tcherkez, G. G. B., Farquhar, G. D. & Andrews, T. J. Despite slow catalysis and confused substrate specificity, all ribulose biphosphate carboxylases may be nearly perfectly optimized. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7246–7251 (2006).
6. Maheshwari, C. *et al.* Targeted knockdown of ribulose-1, 5-bisphosphate carboxylase-oxygenase in rice mesophyll cells. *J. Plant Physiol.* **260**, 153395 (2021).
7. Stitt, M. *et al.* Decreased ribulose-1,5-bisphosphate carboxylase-oxygenase in transgenic tobacco transformed with “antisense” rbcS : II. Flux-control coefficients for photosynthesis in varying light, CO<sub>2</sub>, and air humidity. *Planta* **183**, 555–566 (1991).
8. Fichtner, K. *et al.* Decreased ribulose-1,5-bisphosphate carboxylase-oxygenase in transgenic tobacco transformed with “antisense” rbcS. *Planta* **190**, (1993).
9. Dietz, K.-J. & Heber, U. Rate-limiting factors in leaf photosynthesis. I. Carbon fluxes in the calvin cycle. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **767**, 432–443 (1984).
10. Quick, W. P. *et al.* Decreased ribulose-1,5-bisphosphate carboxylase-oxygenase in transgenic tobacco transformed with “antisense” rbcS : I. Impact on photosynthesis in ambient growth conditions. *Planta* **183**, 542–554 (1991).
11. Farquhar, G. D., von Caemmerer, S. & Berry, J. A. A biochemical model of photosynthetic CO<sub>2</sub> assimilation in leaves of C<sub>3</sub> species. *Planta* **149**, 78–90 (1980).
12. Raven, J. A. Rubisco: still the most abundant protein of Earth? *The New phytologist* vol. 198 1–3 (2013).
13. Carmo-Silva, E., Scales, J. C., Madgwick, P. J. & Parry, M. A. J. Optimizing Rubisco and its regulation for greater resource use efficiency. *Plant Cell Environ.* **38**, 1817–1832 (2015).
14. Bar-On, Y. M. & Milo, R. The global mass and average rate of rubisco. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 4738–4743 (2019).
15. Mackinder, L. C. M. *et al.* A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 5958–5963 (2016).
16. Barrett, J., Girr, P. & Mackinder, L. C. M. Pyrenoids: CO<sub>2</sub>-fixing phase separated liquid organelles. *Biochim. Biophys. Acta Mol. Cell Res.* **1868**, 118949 (2021).
17. Oh, Z. G. *et al.* A linker protein from a red-type pyrenoid phase separates with Rubisco via oligomerizing sticker motifs. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2304833120 (2023).
18. Adler, L. *et al.* New horizons for building pyrenoid-based CO<sub>2</sub>-concentrating mechanisms in plants to improve yields. *Plant Physiol.* (2022) doi:10.1093/plphys/kiac373.
19. Fei, C., Wilson, A. T., Mangan, N. M., Wingreen, N. S. & Jonikas, M. C. Modelling the pyrenoid-based CO<sub>2</sub>-concentrating mechanism provides insights into its operating principles and a roadmap for its engineering into crops. *Nat Plants* **8**, 583–595 (2022).
20. Mukherjee, A. *et al.* Thylakoid localized bestrophin-like proteins are essential for the CO<sub>2</sub> concentrating mechanism of *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 16915–16920 (2019).
21. Itakura, A. K. *et al.* A Rubisco-binding protein is required for normal pyrenoid number and starch sheath morphology in *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 18445–18454 (2019).

22. Atkinson, N. *et al.* The pyrenoidal linker protein EPYC1 phase separates with hybrid *Arabidopsis-Chlamydomonas* Rubisco through interactions with the algal Rubisco small subunit. *J. Exp. Bot.* (2019) doi:10.1093/jxb/erz275.
23. Atkinson, N., Mao, Y., Chan, K. X. & McCormick, A. J. Condensation of Rubisco into a proto-pyrenoid in higher plant chloroplasts. *Nat. Commun.* **11**, 6303 (2020).
24. Atkinson, N., Stringer, R., Mitchell, S. R., Seung, D. & McCormick, A. J. SAGA1 and SAGA2 promote starch formation around proto-pyrenoids in *Arabidopsis* chloroplasts. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2311013121 (2024).
25. Hofmann, H. *et al.* Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 16155–16160 (2012).
26. Freeman Rosenzweig, E. S. *et al.* The Eukaryotic CO<sub>2</sub>-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization. *Cell* **171**, 148–162.e19 (2017).
27. Hovde, B. T. *et al.* Genomic characterization reveals significant divergence within *Chlorella sorokiniana* (Chlorellales, Trebouxiophyceae). *Algal Research* **35**, 449–461 (2018).
28. Cortona, A. D. *et al.* Neoproterozoic origin and multiple transitions to macroscopic growth in green seaweeds. *Proceedings of the National Academy of Sciences* **117**, 2551–2559 (2020).
29. Guo, W. *et al.* 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*. *Chem. Eng. J.* **430**, 133037 (2022).
30. Wunder, T., Cheng, S. L. H., Lai, S.-K., Li, H.-Y. & Mueller-Cajar, O. The phase separation underlying the pyrenoid-based microalgal Rubisco supercharger. *Nat. Commun.* **9**, 5076 (2018).
31. McSwiggen, D. T., Mir, M., Darzacq, X. & Tjian, R. Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes Dev.* **33**, 1619–1634 (2019).
32. He, S. *et al.* The structural basis of Rubisco phase separation in the pyrenoid. *Nature Plants* **1–11** (2020).
33. Oltrogge, L. M. *et al.* Multivalent interactions between CsoS2 and Rubisco mediate  $\alpha$ -carboxysome formation. *Nat. Struct. Mol. Biol.* **27**, 281–287 (2020).
34. Wang, H. *et al.* Rubisco condensate formation by CcmM in  $\beta$ -carboxysome biogenesis. *Nature* **566**, 131–135 (2019).
35. Ilca, S. L. *et al.* Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nat. Commun.* **6**, 8843 (2015).
36. Kellogg, E. & Juliano, N. The structure and function of RuBisCO and their implications for systematic studies. *Am. J. Bot.* **84**, 413 (1997).
37. Goudet, M. M. M. *et al.* Rubisco and carbon concentrating mechanism (CCM) co-evolution across Chlorophyte and Streptophyte green algae. *New Phytol.* (2020) doi:10.1111/nph.16577.
38. Meyer, M. T. *et al.* Rubisco small-subunit  $\alpha$ -helices control pyrenoid formation in *Chlamydomonas*. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19474–19479 (2012).
39. Stotz, M. *et al.* Structure of green-type Rubisco activase from tobacco. *Nat. Struct. Mol. Biol.* **18**, 1366–1370 (2011).
40. Duff, A. P., Andrews, T. J. & Curmi, P. M. The transition between the open and closed states of rubisco is triggered by the inter-phosphate distance of the bound bisphosphate. *J. Mol. Biol.* **298**, 903–916 (2000).
41. Long, B. M., Matsuda, Y. & Moroney, J. V. Algal chloroplast pyrenoids: Evidence for convergent evolution. *Proceedings of the National Academy of Sciences of the United States of America* vol. 121 e2402546121 (2024).
42. GrandPre, T. *et al.* Impact of Linker Length on Biomolecular Condensate Formation. *PRX Life* **1**, 023013 (2023).
43. Payne-Dwyer, A. *et al.* Predicting Rubisco:Linker Condensation from Titration in the Dilute Phase. *arXiv [cond-mat.soft]* (2023).
44. Choi, J.-M. & Pappu, R. V. The Stickers and Spacers Framework for Describing Phase Behavior of Multivalent Intrinsically Disordered Proteins. *Biophys. J.* **118**, 492a (2020).

45. Atkinson, N. *et al.* Rubisco small subunits from the unicellular green alga *Chlamydomonas* complement Rubisco-deficient mutants of *Arabidopsis*. *New Phytol.* **214**, 655–667 (2017).
46. Meyer, M. T. *et al.* Assembly of the algal CO<sub>2</sub>-fixing organelle, the pyrenoid, is guided by a Rubisco-binding motif. *Sci Adv* **6**, (2020).
47. Kropat, J. *et al.* A revised mineral nutrient supplement increases biomass and growth rate in *Chlamydomonas reinhardtii*. *Plant J.* **66**, 770–780 (2011).
48. Mészáros, B., Erdos, G. & Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337 (2018).
49. Emenecker, R. J., Griffith, D. & Holehouse, A. S. Metapredict V2: An update to metapredict, a fast, accurate, and easy-to-use predictor of consensus disorder and structure. *bioRxiv* 2022.06.06.494887 (2022) doi:10.1101/2022.06.06.494887.
50. Newman, A. M. & Cooper, J. B. XSTREAM: a practical algorithm for identification and architecture modeling of tandem repeats in protein sequences. *BMC Bioinformatics* **8**, 382 (2007).
51. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
52. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
53. Barratt, L. J., Reynolds, I. J., Franco Ortega, S. & Harper, A. L. Transcriptomic and co-expression network analyses on diverse wheat landraces identifies candidate master regulators of the response to early drought. *Front. Plant Sci.* **14**, 1212559 (2023).
54. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44 (2020).
55. Almagro Armenteros, J. J. *et al.* Detecting sequence signals in targeting peptides using deep learning. *Life Sci Alliance* **2**, (2019).
56. Engler, C. *et al.* A golden gate modular cloning toolbox for plants. *ACS Synth. Biol.* **3**, 839–843 (2014).
57. Shivhare, D. & Mueller-Cajar, O. *In Vitro* Characterization of Thermostable CAM Rubisco Activase Reveals a Rubisco Interacting Surface Loop. *Plant Physiol.* **174**, 1505–1516 (2017).
58. Kapust, R. B. *et al.* Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* **14**, 993–1000 (2001).
59. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
60. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, (2018).
61. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
62. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCr* **6**, 5–17 (2019).
63. Pettersen, E. F. *et al.* UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
64. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
65. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
66. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
67. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
68. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).

69. Yamano, T. & Fukuzawa, H. Transformation of the Model Microalga *Chlamydomonas reinhardtii* Without Cell-Wall Removal. in *Electroporation Protocols: Microorganism, Mammalian System, and Nanodevice* (eds. Li, S., Chang, L. & Teissie, J.) 155–161 (Springer US, New York, NY, 2020).
70. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

**Acknowledgments:** J.B. and S.M. were supported by BBSRC studentships (BB/M0111511a and BB/T007222/1, respectively). J.B., L.C.M.M. and A.J.M. were supported by a CTRF grant (AP23-1\_023). M.I.S.N. was supported by an NWO Rubicon 2022-1 grant (Project ID: 019.221EN.010). Y.M. was funded by a postgraduate research scholarship from the Darwin Trust of Edinburgh. C.D., M.C.L, M.J.P. and L.C.M.M. were supported by EPSRC funding (EP/W024063/1) as part of the York Physics of Pyrenoids Project (YP3) who are also thanked for fruitful discussions. A.J.M. and L.C.M.M. were supported by BBSRC-NSF/BIO funding (BB/S015337/11 and BB/Y000323/1) and Bill and Melinda Gates Foundation/Foreign Commonwealth and Development Office funding (INV-054558). J.N.B was supported by a UKRI Future Leader Fellowship (MR/T040742/1), as was L.C.M.M. (MR/TO20679/1). In addition, we wish to thank University of York Bioscience Technology Facility staff Dr. Adam Dowle, Dr. Grant Calder, Dr. Andrew Leech and Dr. Clare Steele-King for their assistance in mass spectrometry, FRAP, SPR and immunoelectron experiments respectively. Philipp Girr is thanked for his mentorship and for supplying the TEM image of the *Chlamydomonas* pyrenoid. Tobias Wunder and Oliver Mueller-Cajar are thanked for their assistance in the conception of the project. René Inckemann and Tobias Erb are thanked for the gift of the pME\_Cp\_2\_098 plasmid. Luke Oltrogge and David Savage are thanked for allowing the use of Rubisco graphics. Katy Davis is thanked for the supply of Solanaceae plant material. We would also like to thank the BBSRC, the Wellcome Trust (grant number 206161/Z/17/Z), Tony Wild, and the Wolfson Foundation for funding the Eleanor and Guy Dodson Building and associated Cryo-EM facilities, in addition to Dr. Johan Turkenburg and Sam Hart for coordinating data collection.

**Author contributions:** J.B. and L.C.M.M. conceived the study and wrote the manuscript with contributions from M.J.P., A.J.M. and M.C.L.. J.B. performed biochemical characterisations, in vivo *Chlamydomonas* experiments. M.I.S.N. performed co-immunoprecipitation, absolute quantification and immunoelectron experiments. J.B conceived and constructed FLIPper with assistance from S.M.. C.D. performed NMR experiments. A.S. performed sequence validation. J.B. and J.N.B. performed cryo-EM experiments and analysis. Y.M. and A.J.M. designed and performed in planta experiments.

**Competing interests:** None