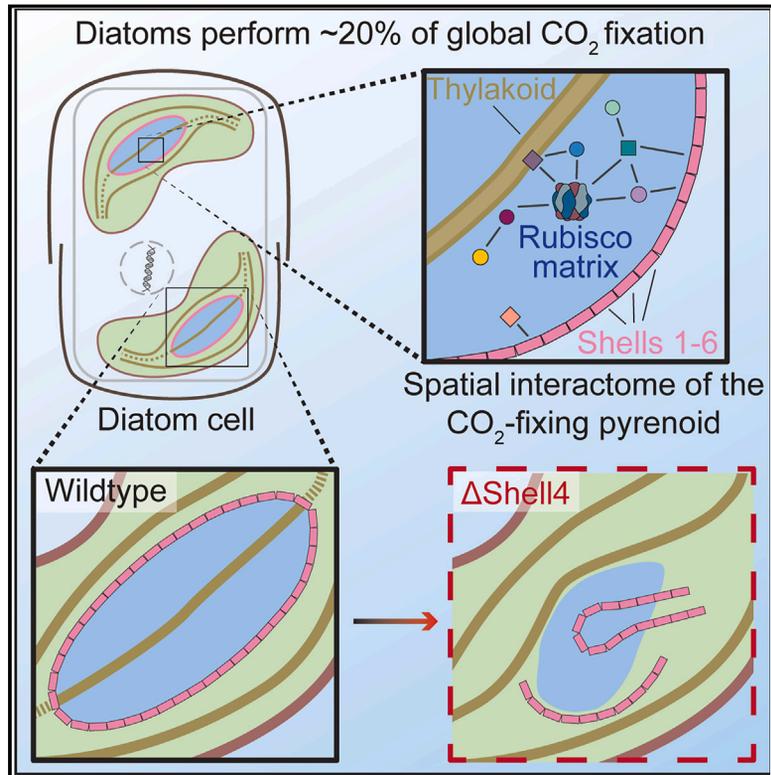


A protein blueprint of the diatom CO₂-fixing organelle

Graphical abstract



Authors

Onyou Nam, Sabina Musial, Manon Demulder, ..., James N. Blaza, Benjamin D. Engel, Luke C.M. Mackinder

Correspondence

ben.engel@unibas.ch (B.D.E.),
luke.mackinder@york.ac.uk (L.C.M.M.)

In brief

A conserved family in the protein interaction network of the CO₂-fixing pyrenoid of diatoms forms a pyrenoid-encapsulating layer that is essential for organelle architecture and efficient CO₂ fixation.

Highlights

- A spatial interaction network reveals 10 previously unknown diatom pyrenoid proteins
- Six Shell proteins encapsulate the pyrenoids of diatoms
- Shell1/2 and Shell4 are essential for efficient CO₂ fixation and pyrenoid architecture
- Shell proteins are widely conserved in major phytoplankton primary producers



Article

A protein blueprint of the diatom CO₂-fixing organelle

Onyou Nam,^{1,2} Sabina Musiał,^{1,2} Manon Demulder,³ Caroline McKenzie,^{1,2} Adam Dowle,¹ Matthew Dowson,^{1,2} James Barrett,^{1,2} James N. Blaza,⁴ Benjamin D. Engel,^{3,*} and Luke C.M. Mackinder^{1,2,5,*}

¹Department of Biology, University of York, York YO10 5DD, UK

²Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5DD, UK

³Biozentrum, University of Basel, Spitalstrasse 41, 4056 Basel, Switzerland

⁴York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, UK

⁵Lead contact

*Correspondence: ben.engel@unibas.ch (B.D.E.), luke.mackinder@york.ac.uk (L.C.M.M.)

<https://doi.org/10.1016/j.cell.2024.09.025>

SUMMARY

Diatoms are central to the global carbon cycle. At the heart of diatom carbon fixation is an overlooked organelle called the pyrenoid, where concentrated CO₂ is delivered to densely packed Rubisco. Diatom pyrenoids fix approximately one-fifth of global CO₂, but the protein composition of this organelle is largely unknown. Using fluorescence protein tagging and affinity purification-mass spectrometry, we generate a high-confidence spatially defined protein-protein interaction network for the diatom pyrenoid. Within our pyrenoid interaction network are 10 proteins with previously unknown functions. We show that six of these form a shell that encapsulates the Rubisco matrix and is critical for pyrenoid structural integrity, shape, and function. Although not conserved at a sequence or structural level, the diatom pyrenoid shares some architectural similarities to prokaryotic carboxysomes. Collectively, our results support the convergent evolution of pyrenoids across the two main plastid lineages and uncover a major structural and functional component of global CO₂ fixation.

INTRODUCTION

Approximately one-third of global carbon fixation takes place in pyrenoids.¹ Pyrenoids are biomolecular condensates of the principal CO₂-fixing enzyme Rubisco found in the chloroplasts of algae.² There are two major chloroplast lineages, the green and red plastids, with algae within these lineages proposed to have convergently evolved pyrenoids.^{3,4} Nearly all of our knowledge of pyrenoid function and composition comes from the model terrestrial green plastid containing alga *Chlamydomonas reinhardtii*.^{2,5} *C. reinhardtii* is a powerful model organism,⁶ but global carbon fixation is primarily driven by oceanic red plastid containing algae, such as diatoms, where our knowledge is still in its infancy.^{7–9} Diatoms are responsible for up to 20% of global net primary production, are estimated to fix ~70 gigatons CO₂ per year, and are fundamental for the long-term storage of carbon by driving the flux of organic material from the ocean surface to sediments.^{10,11}

Pyrenoids are found at the heart of algal CO₂ concentrating mechanisms (CCMs).^{2,8} CCMs overcome the slow diffusion of CO₂ in water and the catalytic limitations of Rubisco by actively pumping inorganic carbon from the external environment into the cell and releasing it as CO₂ in the pyrenoid, where it can be fixed by densely clustered Rubisco. In *C. reinhardtii*, a disor-

dered linker protein, essential pyrenoid component 1 (EPYC1), condenses Rubisco to form the liquid-liquid phase-separated matrix of the pyrenoid.^{1,12,13} A shared Rubisco binding motif found in EPYC1 and numerous other pyrenoid components enables targeting to and structural organization of the pyrenoid. Proteins containing this motif are proposed to link the matrix to the surrounding starch sheath and to specialized matrix-traversing thylakoids called pyrenoid tubules.¹⁴ Inorganic carbon in the form of HCO₃[–] is shuttled into the pyrenoid tubules by bestrophin-like proteins¹⁵ where a carbonic anhydrase converts it to membrane-permeable CO₂ that is then fixed by Rubisco in the pyrenoid matrix.^{5,15,16} The surrounding starch acts as a diffusion barrier to minimize CO₂ leakage out of the pyrenoid.^{17,18} Fluorescent protein tagging,^{14,19} affinity purification followed by mass spectrometry (APMS),¹⁹ and proximity labeling²⁰ have enabled a high-confidence pyrenoid proteome to be determined, with multiple components now functionally characterized. This has enabled *C. reinhardtii* proto-pyrenoid engineering in plants²¹ and a parts list of components that should theoretically enable the engineering of a functional pyrenoid-based CCM to enhance plant photosynthesis.^{17,22}

Diatom pyrenoids share certain features with the *C. reinhardtii* pyrenoid, including condensed Rubisco and traversing thylakoids. They also show some limited conservation of inorganic



carbon delivery proteins.⁹ However, the level of conservation for structural proteins remains unclear.^{8,9} The only structural component identified so far for the diatom pyrenoid is pyrenoid component 1 (PYCO1), a Rubisco linker protein found in the pyrenoid of the pennate diatom *Phaeodactylum tricornutum* and suggested to be responsible for phase separating Rubisco to form the pyrenoid matrix.⁷ However, PYCO1 is not widely conserved, being absent in centric diatoms, and its functional importance is yet to be determined. Outside of PYCO1, most previous diatom CCM research has focused on inorganic carbon uptake. In *P. tricornutum*, several candidates belonging to the solute carrier 4 (SLC4) family of transporters have been proposed for HCO₃⁻ uptake at the plasma and chloroplast membranes.^{23,24} Bestrophin-like proteins are also implicated in the shuttling of HCO₃⁻ into the thylakoid lumen, where a θ -type carbonic anhydrase has been identified that is thought to function by releasing CO₂ from the thylakoid membranes that traverse the pyrenoid matrix.^{25,26} In the centric diatom *Thalassiosira pseudonana*, SLC4 candidates have been implicated in inorganic carbon uptake across both the plasma and thylakoid membranes.^{9,27} Recently, two bestrophin-like proteins were localized to the *T. pseudonana* pyrenoid, likely in the pyrenoid penetrating thylakoid (PPT)^{28,29} and a θ -type carbonic anhydrase 2 (θ CA2) confirmed to be located in the PPT.⁹ However, supporting functional data are missing for these proteins in *T. pseudonana*. In contrast to *C. reinhardtii*, diatoms lack a starch sheath encapsulating their pyrenoids. In *C. reinhardtii* flux balance modeling of pyrenoid function¹⁷ and analysis of starch mutants¹⁸ indicate that a diffusion barrier is essential for efficient pyrenoid function to minimize CO₂ leakage. How the diatom pyrenoid minimizes CO₂ leakage is a substantial, outstanding question.

Here, we advance our knowledge of the diatom pyrenoid by developing an iterative approach of fluorescent protein tagging followed by APMS in *T. pseudonana*, which belongs to a biogeochemically important genus. These data enabled us to build a high-confidence diatom pyrenoid interaction network, identifying multiple pyrenoid proteins, many with no previously known functional domains. A family of these proteins form a static shell that encapsulates the Rubisco matrix and is critical for pyrenoid shape, structural integrity, and CCM function. Our findings provide new insight into a globally important organelle and provide additional molecular parts for engineering a CCM into crop plants to improve productivity.

RESULTS AND DISCUSSION

Rubisco co-immunoprecipitation mass spectrometry to identify diatom pyrenoid components

Although diatoms play a central role in global biogeochemical cycles, very little is known about the diatom pyrenoid. In *T. pseudonana*, cells have two chloroplasts, each surrounded by four membranes, with each chloroplast containing a single centrally positioned lenticular-shaped pyrenoid (Figure 1A). As a starting point to understand *T. pseudonana* pyrenoid composition, we performed co-immunoprecipitation coupled with mass spectrometry (colPMS), using the main pyrenoid component, Rubisco, as a bait protein (Figure 1B). To immunoprecipitate Ru-

bisco, we used an antibody raised to a conserved 12 amino acid surface-exposed peptide on the Rubisco large subunit (rbcL) (Figures S1A and S1B). By comparing two independent colPMS experiments, consisting of two and three technical replicates, respectively, against non-antibody control experiments, we identified 36 putative pyrenoid components out of a total of 1,167 detected proteins identified with 2 or more spectral counts (Table S1). For these exploratory experiments, we applied a relaxed cut-off based on the fold-change enrichment of bestrophin-like protein 2 (BST2), which we had previously localized to the pyrenoid²⁸ and would expect to only have a weak enrichment due to it being predicted to be a membrane protein (Figure 1B). Top hits from our rbcL colPMS experiments were then fed into an iterative fluorescent protein tagging, localization, and APMS framework that we used to build a spatially defined pyrenoid proteome (Figure 1C).

Development of a high-throughput tagging pipeline in *T. pseudonana* identifies multiple pyrenoid components

To enable rapid cycling through our iterative pipeline, we set out to establish high-throughput fluorescent protein tagging and screening in diatoms. We initially adapted our Golden Gate modular cloning-based episomal assembly framework²⁸ to be 96-well compatible and combined it with multi-well diatom transformation via bacterial conjugation. We coupled this with 48-well plate strain maintenance and 96-well plate flow cytometry screening for clonal fluorophore-fusion expressing lines (Figure 2A). As colPMS data on a small scale are inherently noisy with both false positives and false negatives,³⁰ we applied our tagging pipeline to 22 Rubisco colPMS hits to validate if they were bona fide pyrenoid proteins. Nourseothricin-positive transformants were picked and screened for positive fluorescence using flow cytometry. Due to typical mosaic colony presence,³¹ either multiple rounds of screening or screening of several independent colonies was required to identify stable monomeric enhanced green fluorescent protein (mEGFP) expressing lines. We found that screening 8–12 colonies would typically yield a stable cell population with >90% of cells mEGFP positive (Figure S1C). Positive lines were subsequently imaged by confocal microscopy (Figures 2B and S2). From the initially identified Rubisco-interacting proteins, we localized 13 proteins to distinct sub-regions of the pyrenoid (Figures 2B and S2A; see below for further discussion). However, 9 candidates were either localized to chloroplast sub-regions adjacent to the pyrenoid (4) or to distinct non-chloroplast regions (5) (Figure S2B), indicating that our colPMS data contains false positives.

To accurately determine sub-pyrenoid localization of components, we implemented a dual-tagging approach using two spectrally compatible fluorophores. We first developed a pyrenoid matrix marker for co-localization. In green algae, nuclear-encoded rbcS-fluorescent protein fusions have been powerful for understanding sub-pyrenoid spatial organization¹⁹ and for determining the liquid-like properties of the pyrenoid.¹² As the rbcS of diatoms is chloroplast encoded, and no *T. pseudonana* chloroplast transformation protocol is available, we wondered if we could target an episomal-expressed rbcS-mEGFP to the pyrenoid. Using the N-terminal signal and transit peptide sequences from the nuclear-encoded chloroplast localized BST2 protein,³²

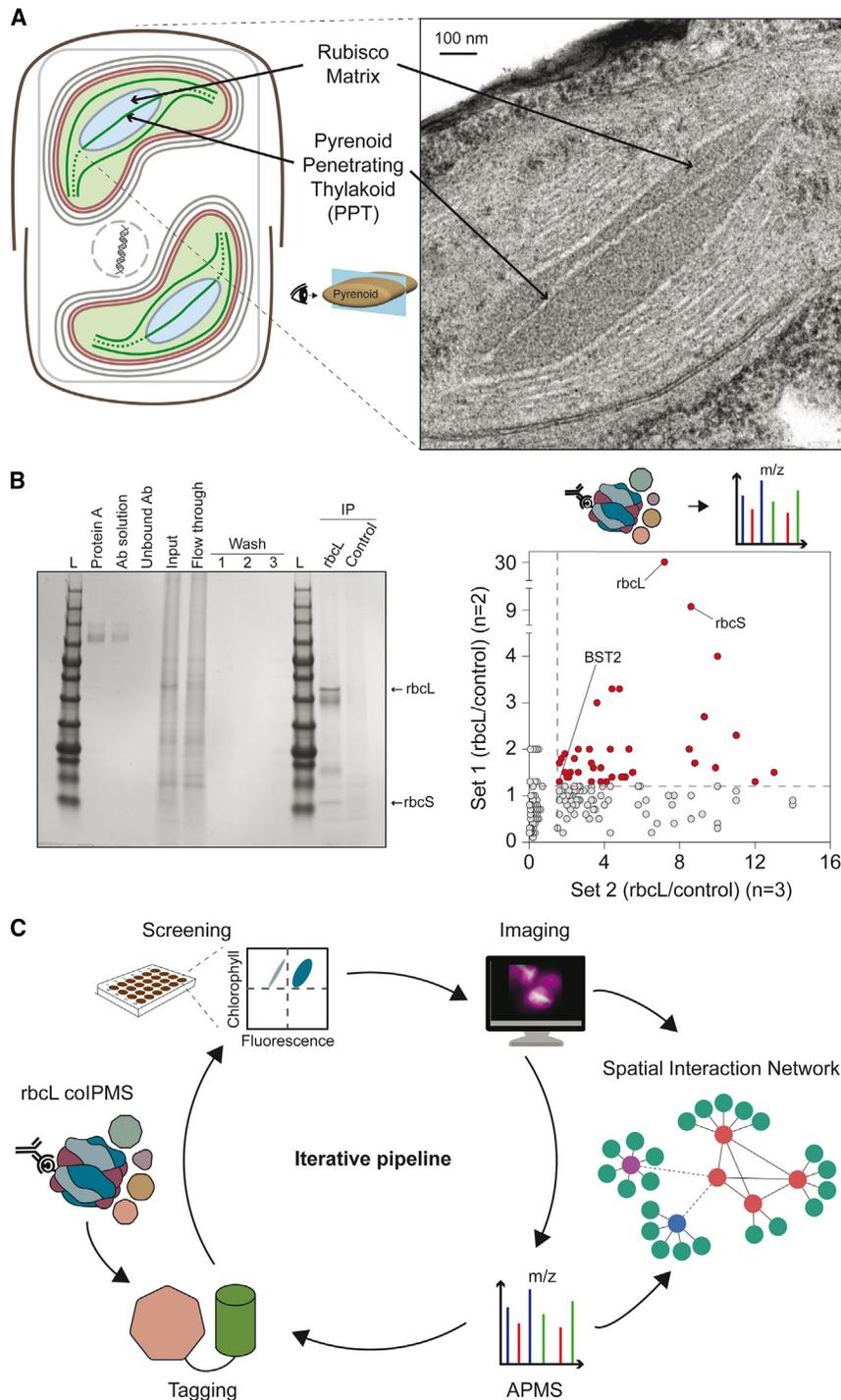


Figure 1. Identification of candidate diatom pyrenoid components

(A) The diatom *T. pseudonana* has two chloroplasts, each containing a single pyrenoid traversed by a specialized membrane called the pyrenoid penetrating thylakoid (PPT). The connection between the PPT and the wider thylakoid network is unresolved. (B) Rubisco co-immunoprecipitation followed by mass spectrometry (colPMS) to identify candidate pyrenoid components. Left: example Rubisco large subunit (rbcL) colIP with different steps resolved by SDS-PAGE. Right: two experiments consisting of either two or three technical colIPs were performed and the antibody-bound fractions submitted for mass spectrometry. Cutoffs (dashed lines) were determined based on bestrophin-like protein 2 (BST2) that was previously shown to be pyrenoid localized. Ab, antibody; L, protein ladder.

(C) An iterative pipeline to determine the pyrenoid spatial interaction network. RbcL colPMS data fed into a high-throughput fluorescence protein tagging pipeline. Pyrenoid confirmed proteins were then used for affinity purification followed by mass spectrometry (APMS). Interactors were tagged to confirm pyrenoid localization and used in subsequent APMS rounds.

See also [Figure S1](#).

localization of BST2, with the rbcS signal extending outside that of BST2, supporting a PPT localization of BST2 ([Figure 2D](#)).

Establishing large-scale APMS to build a pyrenoid interaction network

To expand the pyrenoid proteome, we developed and optimized an APMS pipeline in *T. pseudonana* using our GFP-tagged pyrenoid proteins as baits. Lines expressing GFP-tagged proteins were typically grown at ambient CO₂, where the CCM is fully active.³³ In triplicate, GFP-trap nanobodies were used to enrich for target proteins from cell lysate, and their interactors were determined via liquid chromatography-tandem mass spectrometry (LC-MS/MS) ([Figure 3A](#); [Table S2](#)). Protein-protein interactions were stringently defined by comparing both CompPASS (comparative proteomic analysis software suite)³⁴ and SAINT (significance analysis of interactome)³⁵

we successfully targeted rbcS to the pyrenoid, allowing us to clearly delineate the pyrenoid matrix ([Figure 2C](#)). Second, we tested assembling two target genes with different fluorophores on the same episome. We decided to initially validate the BST2 localization by making an episome with BST2-mEGFP and rbcS-mScarlet-I, a fluorophore we had previously validated using our system.²⁸ Using this approach, we observed a clear pyrenoid

scores ([Figure 3A](#); [Table S3](#)), which use different weighting criteria to identify true interactors from non-specific background using label-free proteomic quantitation data. We set interaction confidence thresholds based on the known interaction of rbcS with rbcL, which resulted in proteins in the top 2.2% for CompPASS and top 1% for SAINT being designated as high-confidence interactors.

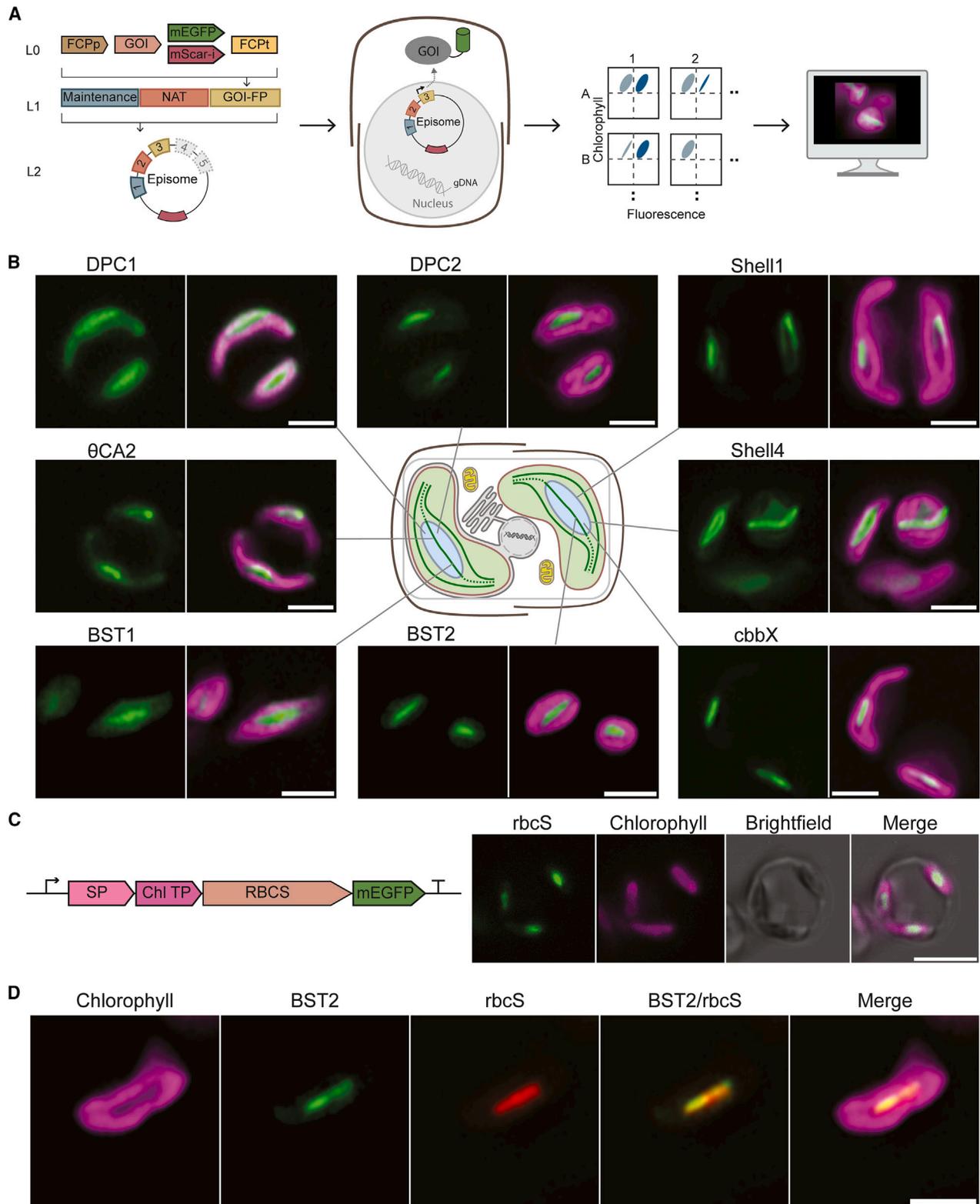


Figure 2. High-throughput fluorescent protein tagging identifies multiple diatom pyrenoid components

(A) Candidate genes were cloned using Golden Gate modular cloning, transformed into *T. pseudonana* via bacterial conjugation, and screened by flow cytometry for fluorescence prior to imaging. Gene of interest (GOI) expression was driven by the fucoxanthin chlorophyll a/c-binding protein promoter (FCPp) and terminated using the FCP terminator (FCPt).

(legend continued on next page)

Our initial 13 pyrenoid-localized proteins included the Rubisco small subunit *rbcS*; BST1 and BST2, bestrophin-like proteins proposed to be involved in HCO_3^- uptake into the PPTs²⁸; θCA2 , most likely involved in CO_2 release from HCO_3^- within the PPTs^{25,26}; *cbbX*, a nuclear-encoded red-type Rubisco activase³⁶ that until now has not been localized in algae with red plastids³⁷; and 8 uncharacterized proteins with no clear function. We named these previously uncharacterized pyrenoid proteins diatom pyrenoid components 1 and 2 (DPC1 and DPC2) and Shells 1–6. While DPC1 and DPC2 were predominantly in the pyrenoid matrix, the initial tagging of Shell1 and Shell4 showed that they may encapsulate the pyrenoid (Figure 2B; see below). A subset of these pyrenoid-localized components was utilized for APMS using our iterative pipeline (Figures 1C and 3B; Table S4). Subsequently, two additional components, DPC3 and DPC4, which had strong interactions with Shell4 and *rbcS* but again with no sequence predictable function, were localized and fed into our APMS pipeline. Whereas DPC4 was found throughout the pyrenoid with a matrix-type localization, DPC3 appeared to encapsulate the pyrenoid similar to the Shell proteins (Figure 3C). Combining the data and using our stringent interaction scoring approach enabled us to construct a high-confidence pyrenoid interaction network for *T. pseudonana* built from 11 baits and containing 46 additional protein nodes linked by 57 interaction edges (Figure 3B). In the network, interaction confidence can be further interpreted by CompPASS and SAINT score magnitude (line thickness and color, respectively, in Figure 3B) as well as the number of connecting edges with baits.

rbcS and DPC2 appear to be key hub proteins, each linking four nodes that further link to confirmed pyrenoid components. How Rubisco is packaged into the pyrenoid is unknown in *T. pseudonana*. With the absence of an EPYC1 or PYCO1 homolog to phase separate Rubisco, it was hypothesized that an alternative repeat protein could be fulfilling this role. Unexpectedly, none of the pyrenoid proteins identified in our study contain a repeated sequence with the expected frequency of ~60 amino acids,^{1,38} potentially indicating that pyrenoid assembly in *T. pseudonana* could be based on biophysical principles that are different than in both green algal and pennate diatom pyrenoids. The localization of DPC2 solely to the pyrenoid (Figure 2B), its Rubisco interaction (Figure 3B), and its relatively high abundance from whole-cell mass spectrometry (Figure 3D; Table S5) make DPC2 a strong candidate for future studies.

***T. pseudonana* has six Shell homologs, and Shell proteins are found across algal lineages containing secondary red plastids**

In our initial tagging, we were intrigued to see that two proteins with no annotated functional domains appeared to encapsulate the pyrenoid matrix (Figure 2B: Shell1, Shell4). In the well-char-

acterized green algal pyrenoid, chloroplast synthesized starch forms a sheath that encapsulates the Rubisco matrix and acts as a CO_2 leakage barrier to enhance CCM efficiency.^{17,18} In diatoms, the main carbohydrate storage molecule is chrysolaminaran, which is stored in cytosolic vacuoles³⁹ with no clear carbohydrate barrier surrounding the diatom pyrenoid. Pyrenoid modeling of inorganic carbon fluxes has shown that a CO_2 diffusion barrier is critical for an efficient CCM,¹⁷ thus opening up the question of how diatoms and other algal lineages that lack starch encapsulated pyrenoids operate efficient CCMs.² Although protein encapsulation of pyrenoids has not previously been identified, cyanobacterial Rubisco-containing carboxysomes have a well-characterized protein shell that is proposed to act as a CO_2 barrier to minimize leakage^{40–43} and is essential for carboxysome biogenesis and shape.^{42,43} Instead of starch, we hypothesized a protein shell could have an analogous function in diatoms. We set out to further understand the importance of the Shell proteins for pyrenoid function.

BLAST analysis of Shell1 and Shell4 identified four additional homologs in the *T. pseudonana* genome that all contain two predicted β sheet domains (Figures 4A and S3) and are predicted to be structurally similar to each other but distinct from DPC3 (Figures S3C and S3D). We explored the distribution of these proteins across different evolutionary lineages. BLAST analysis against the NCBI database identified homologs in stramenopiles (including other diatoms), pelagophytes, and haptophytes, all of which are photosynthetic algae that contain red plastids originated from secondary endosymbiotic events (Figure 4B; Table S6). However, Shell proteins are absent in the rhodophytes, the primary endosymbiotic red plastid lineage that was engulfed by a heterotrophic host to form the red plastid secondary endosymbiotic lineages. This suggests that Shell proteins either were absent in the engulfed red alga but originated in the heterotrophic host prior to endosymbiosis, were present in the engulfed red alga but have since been lost in rhodophytes, or evolved after engulfment. Given that we found no Shell homologs in extant non-photosynthetic organisms or in pyrenoid-containing rhodophytes, it seems plausible that these proteins evolved following the secondary endosymbiotic event. Further supporting a role in pyrenoid function of Shell proteins, transmission electron microscopy (TEM) data from available literature indicates that algae found within the Shell protein-containing clades all possess pyrenoids.²

Shell proteins encapsulate the pyrenoid matrix and localize to distinct sub-regions

To more precisely localize the six Shell proteins, we co-expressed *rbcS*-mScarlet-I with mEGFP-tagged Shell proteins (Figures 5A and S4). Analysis of fluorescence intensity of pyrenoid transects (Figures 5B and S4) and max intensity z stack projections (Figures 5C and S4) indicate that all six proteins

(B) Multiple proteins localized to the canonical pyrenoid region within the chloroplast. Identified pyrenoid components with no sequence-predicted function are termed diatom pyrenoid components (DPCs) or Shells. Green: mEGFP fusion protein; magenta: chlorophyll autofluorescence. Scale bars: 2 μm .

(C) Development of a pyrenoid matrix marker. By using the BST2 signal and transit peptide sequences, we could express chloroplast encoded *rbcS* from nuclear episomes and target it to the chloroplast. Scale bar: 5 μm .

(D) Co-localization of BST2-mEGFP (green) and *rbcS*-mScarlet-I (red) expressed from the same episome. Scale bar: 2 μm .

See also Figures S1 and S2.

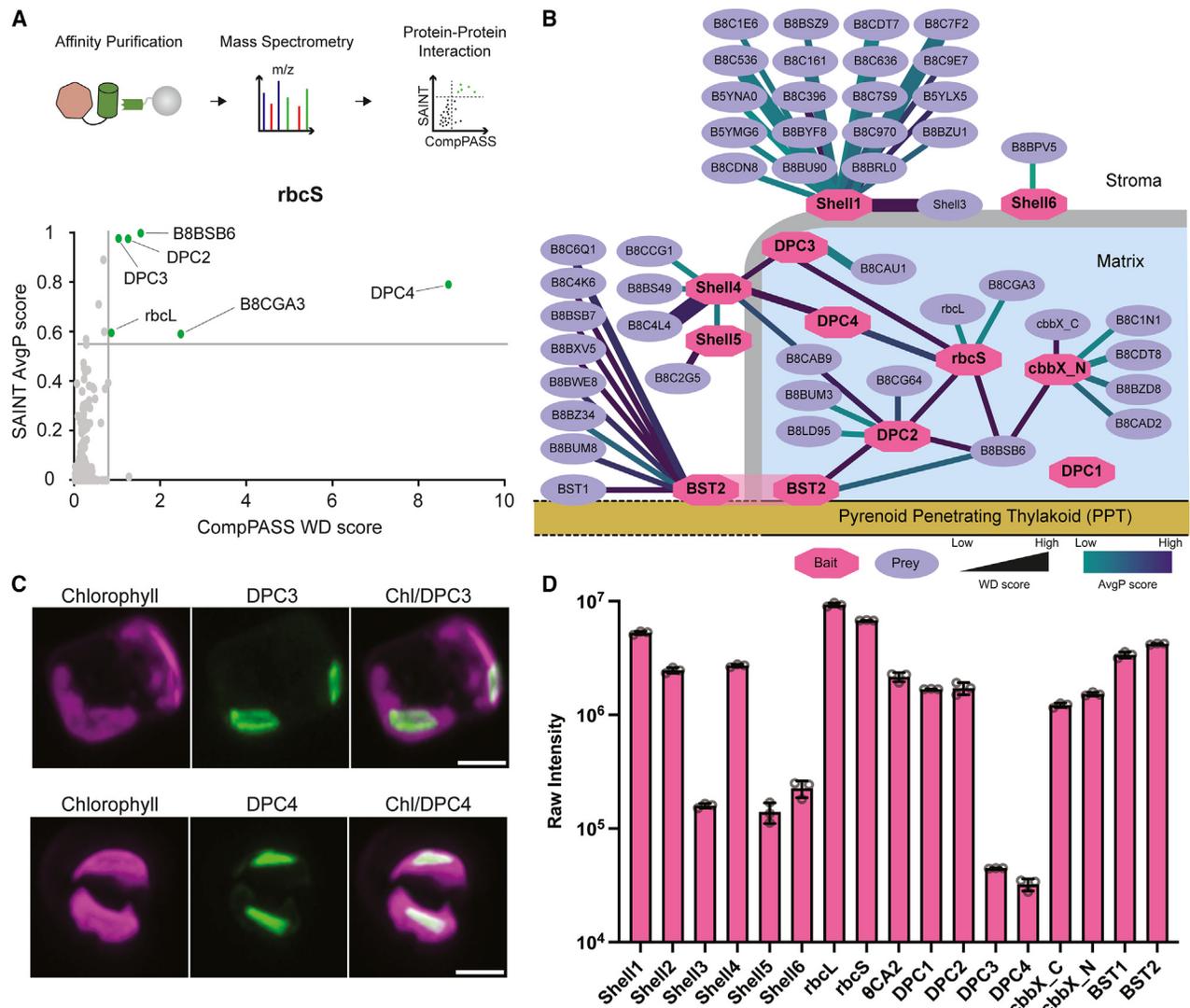


Figure 3. A spatial interaction network of the diatom pyrenoid

(A) Top: mEGFP-tagged proteins were affinity purified using GFP nanobodies, interactors identified using mass spectrometry, and high-confidence protein-protein interactions determined using SAINT and CompPASS analysis. Bottom: analysis example for *rbcS* APMS. *rbcS* is not present in the plot due to bait spectral counts being set to zero prior to SAINT and CompPASS analysis.

(B) Spatially defined protein-protein interaction network of the pyrenoid. Where available, positioning is based on localization determined by confocal microscopy. Due to 96% similarity of the mature proteins of Shell1 and Shell2, all peptides were assigned to Shell1 in the APMS study.

(C) Confocal localization of DPC3 and DPC4 confirming pyrenoid localization (max intensity z stack projections). Green: mEGFP fusion protein; magenta: chlorophyll. Scale bars: 2 μm.

(D) Whole-cell mass spectrometry raw intensity values of pyrenoid confirmed proteins. Data are from WT cells ran in triplicate. Error bars: SD of the mean. See also Figure S2.

encapsulate the Rubisco matrix of the pyrenoid. To prioritize Shell proteins for further investigation, we looked at their relative abundance in our whole-cell mass spectrometry data performed on wild-type (WT) cells (Figure 3D; Table S5). The most abundant homologs were Shell1 and Shell2, which share 92% identity, and Shell4, which shares 34% identity with Shell1 (Figure S3B). As the diatom pyrenoid shape can be generalized as an elliptical cylinder with curved ends (Figure 1), the apparent localization of the Shell proteins is highly dependent on the orientation of the chloroplast and pyrenoid during fluorescence microscopy imaging

(Figures 5A–5C and S4), making the assignment to sub-pyrenoid surface regions challenging. To explore if Shell1 and Shell4 occupy the same regions of the shell, we co-expressed Shell1 and Shell4 with different fluorescent tags. Imaging shows that they localize to distinct regions of the pyrenoid periphery, suggesting that they may have different roles in the formation or function of the shell (Figures 5D–5F).

Although Shell proteins were identified in our *rbcL* colPMS and *rbcS* APMS data, they did not fall above the stringent thresholds we defined for high-confidence interactors with either *rbcS* or

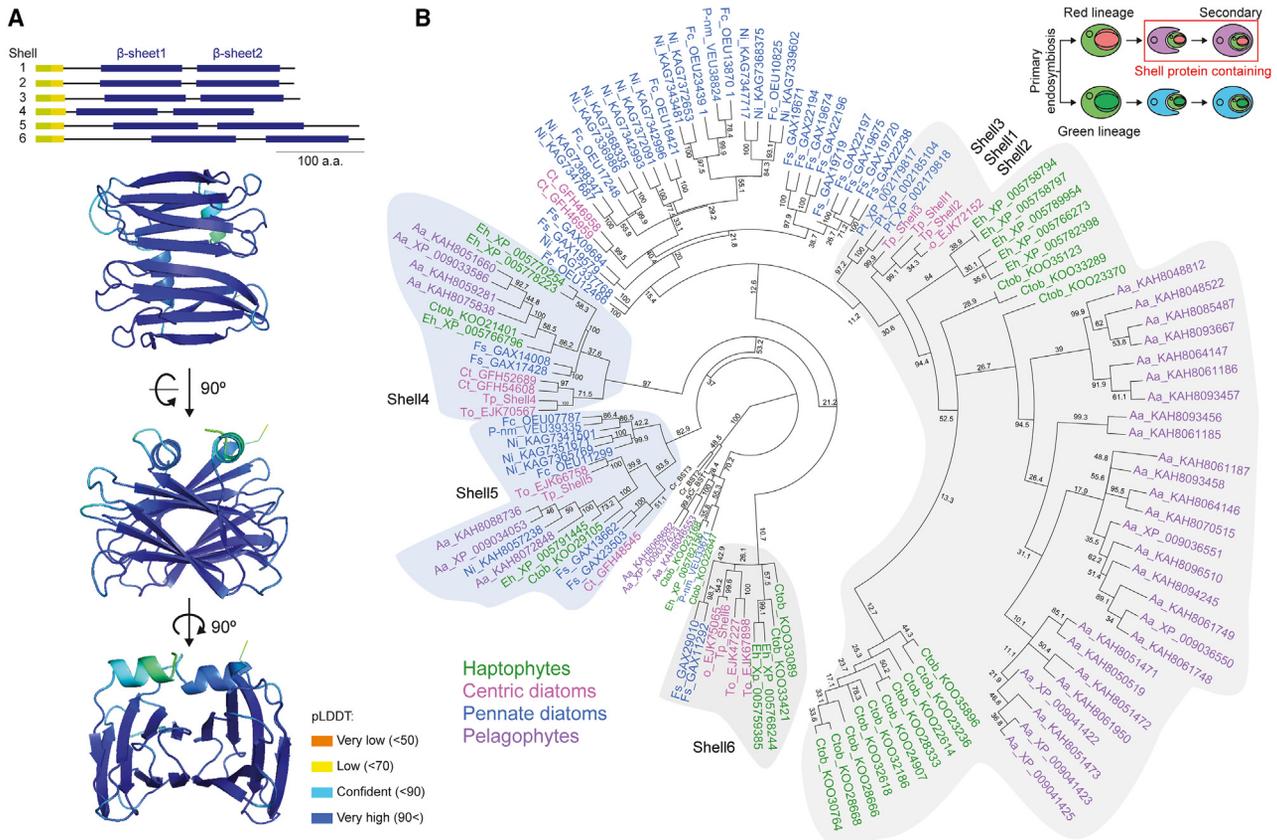


Figure 4. *T. pseudonana* has six Shell homologs, and Shell proteins are found across red plastid secondary endosymbionts

(A) *T. pseudonana* has six Shell protein homologs that all contain two β sheet regions. Top: domain alignment of Shell homologs, green is signal peptide, and yellow is transit peptide. Bottom: AlphaFold2 structure prediction of Shell1. pLDDT: predicted local distance difference test. The N-terminal and C-terminal disordered regions are not shown due to low-confidence structural prediction.

(B) Phylogenetic analysis indicates that Shell proteins are found in red plastid secondary endosymbionts, including stramenopiles, pelagophytes, and haptophytes. Bootstrap values are shown. Cartoon inset represents the two main plastid lineages and where Shell proteins are present. Aa, *Aureococcus anophagefferens*; Ct, *Chaetoceros tenuissimus*; Ctob, *Chrysochromulina tobinii*; Eh, *Emiliania huxleyi*; Fs, *Fistulifera solaris*; Fc, *Fragilariopsis cylindrus*; Ni, *Nitzschia inconspicua*; Pt, *Phaeodactylum tricornutum*; P-nm, *Pseudo-nitzschia multistriata*; To, *Thalassiosira oceanica*; Tp, *Thalassiosira pseudonana*. *Chlamydomonas reinhardtii* bestrophins (Cr_BST1–3) have been used as an outgroup. See also Figure S3.

rbcL. Their pyrenoid-peripheral localization and low-confidence Rubisco interaction suggest that Shell proteins could interact indirectly with the Rubisco matrix. The shell-like pattern displayed by DPC3 (Figure 3C) raises the possibility that this protein could act as an intermediary Rubisco matrix-shell adaptor or potentially an additional shell component. The co-expression of DPC3 with Shell4 shows that they co-localize (Figures 5G–5I). The relatively low protein abundance of DPC3 in comparison to Shell1 and Shell4 (Figure 3D) indicates that DPC3 is most likely a minor protein constituent of the pyrenoid surface.

The pyrenoid matrix and Shell proteins have minimal mobility in the pyrenoid

To understand the dynamics of pyrenoid components, we leveraged our tagged lines. If the shell proteins form a lattice around the Rubisco matrix, similar to carboxysome shell proteins in assembled carboxysomes,^{44–46} then we would expect them to form a static layer. We investigated the mobility of Shell1 and

Shell4 by fluorescence recovery after photobleaching (FRAP). Both Shell1 and Shell4 showed minimal rearrangement after photobleaching (Figures 6A and 6B). Although this does not provide direct evidence that they form a CO₂ diffusion barrier, the static nature of the pyrenoid shell is consistent with our current understanding of both cyanobacterial carboxysome shell and green algal starch sheath CO₂ diffusion barriers.

In *C. reinhardtii*, the pyrenoid matrix has liquid-like properties. Rubisco, EPYC1, and Rubisco activase (RCA1) all show rapid rearrangement in FRAP experiments on the timescale of ~30 s after photobleaching.¹² In *T. pseudonana*, Rubisco is densely packed in the pyrenoid but does not form ordered arrays.⁴⁷ Surprisingly, FRAP experiments on both *T. pseudonana* rbcS and cbbX showed minimal rearrangement over minute timescales (Figure S5A). This observation aligns with recent data from the pennate diatom *P. tricornutum*, where *in vitro* phase-separated Rubisco demonstrates minimal rearrangement over similar timescales, and the linker protein PYCO1 appears

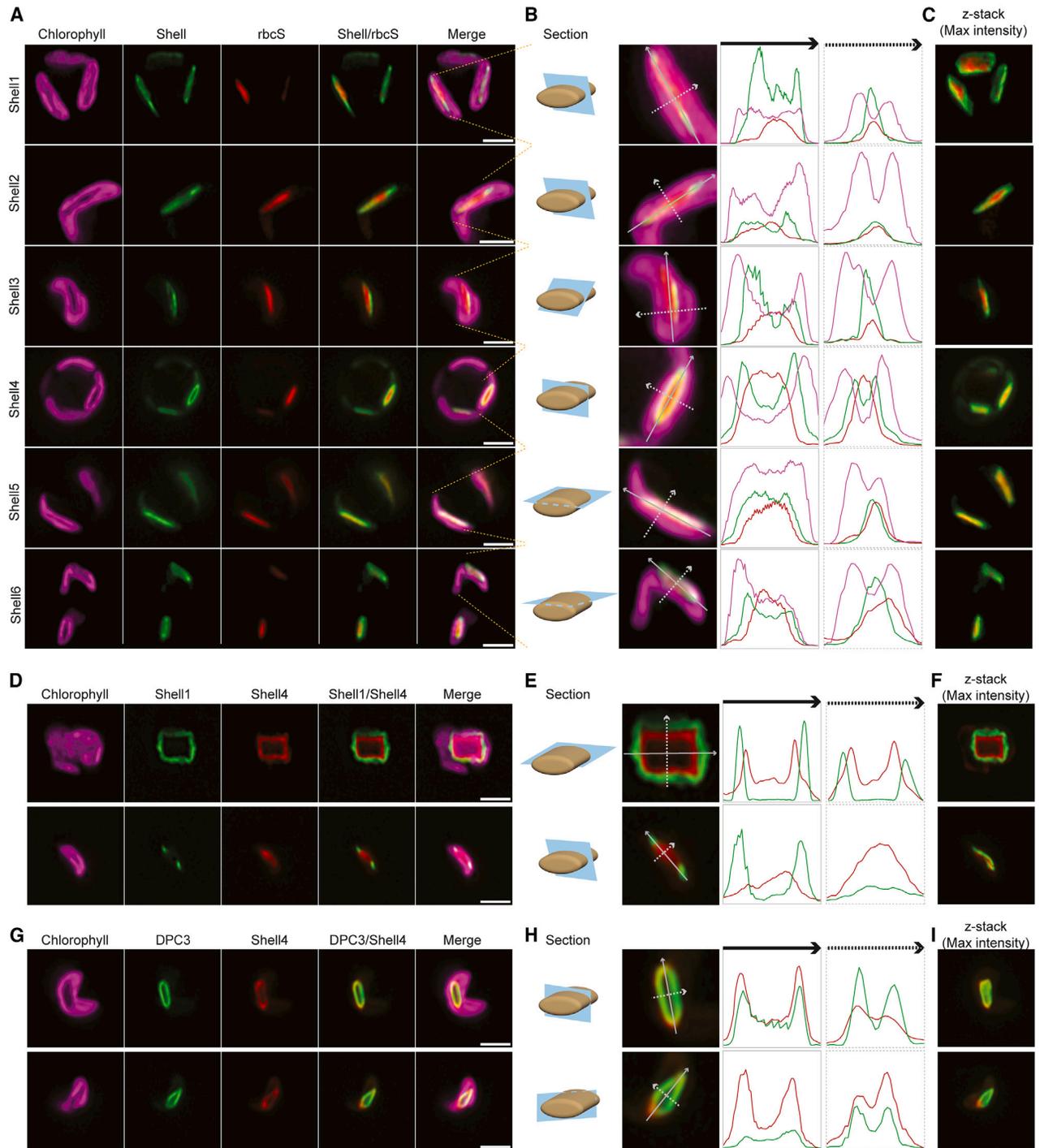


Figure 5. Shell proteins encapsulate the pyrenoid matrix, and Shell1 and Shell4 occupy distinct regions of the pyrenoid surface

(A) Co-localization of Shell proteins with Rubisco.
 (B) Fluorescence intensity cross-sections of chlorophyll, Shell, and rbcS across the pyrenoid.
 (C) z stack max intensity projections of (B).
 (D) Co-localization of Shell1 and Shell4 in two cells with different pyrenoid orientations.
 (E) Fluorescence intensity cross-sections of Shell1 and Shell4 across the pyrenoid.
 (F) z stack max intensity projections of (E).
 (G) Co-localization of DPC3 and Shell4 in two cells with different pyrenoid orientations.

(legend continued on next page)

immobile *in vivo*.⁷ This lack of dynamic mixing opens a considerable question about how *cbbX* can sufficiently access inhibited Rubisco to reactivate it. Collectively, this indicates that the diatom pyrenoid has different mesoscale properties compared to the *C. reinhardtii* pyrenoid and is relatively static once Rubisco is assembled into the pyrenoid. This observation aligns with some cyanobacterial carboxysomes, where Rubisco forms ordered arrays upon packing.^{48,49}

Shell1 can self-assemble *in vitro* to form tubes and sheets

To form an encapsulating protein layer around the Rubisco matrix, we would expect that Shell proteins can undergo higher-order self-assembly. To test this, we purified recombinantly expressed Shell1 (Figure S5B) from *Escherichia coli* and performed cryogenic electron microscopy (cryo-EM). We observed the formation of tubes and sheets (Figures 6C, S5C, and S5D) in the micrographs, which agrees with the structural studies by Shimakawa et al.⁴⁷ This observation is reminiscent of the self-assembly of bacterial microcompartment hexamers, the main building block of carboxysome shells, which have been shown to self-assemble into both sheets⁵⁰ and tubes *in vitro*.⁵¹ The ability to self-assemble into stable higher-order structures likely explains the limited mobility of Shell1 observed *in vivo* and supports the hypothesis that Shell proteins could provide a continuous, encapsulating layer at the pyrenoid surface to function as a CO₂ barrier, control metabolite flux, and/or provide structural integrity to the pyrenoid. We also attempted cryo-EM of recombinantly expressed Shell4 but saw no higher-order *in vitro* structures under the conditions tested. This failure of Shell4 to assemble into tubes or sheets could be due to technical reasons or could indicate a biologically relevant difference between Shell1 and Shell4. Sequence alignment and AlphaFold3 structure predictions show that Shell4 lacks the C-terminal extension of Shell1 (Figures S3A and S3E), which forms an inter-subunit contact and may stabilize the formation of higher-order lattices.⁴⁷

Shell1/2 and Shell4 are essential for CCM function and pyrenoid architecture

The *in vitro* self-assembly of Shell1 into sheets and tubes and the immobile nature of Shell1 and Shell4 *in vivo* suggest they may be required for pyrenoid structural integrity. To test this, we used our MoClo Golden Gate system to simultaneously tag *rbcS* with mEGFP and disrupt either *Shell1/2* (*Shell1* and *Shell2*) or *Shell4* expression by CRISPR. Due to *Shell1* and *Shell2* having 93% DNA sequence similarity (Figure S3B), single guide RNAs were designed to simultaneously target both genes (Figure S6A). Edited lines were grown under high CO₂ conditions and selected by mEGFP fluorescence. Biallelic gene editing was confirmed by Sanger sequencing (Figures S6A and S6B). Imaging of *rbcS*-mEGFP in the *shell1/2* mutant revealed that cells lacking both *Shell1* and *Shell2* failed to form a lenticular pyrenoid and instead

typically possessed a single spherical pyrenoid per chloroplast, although the presence of multiple pyrenoids was also observed (Figures 6D and S6D). This suggests that Shell1/2 are required for the lenticular shape of the pyrenoid, and in their absence, the pyrenoid assembles into a sphere. Similarly, the *shell4* mutant failed to form a lenticular pyrenoid, forming more oval-shaped pyrenoids that were subtly different in shape from the more spherical pyrenoids in the *shell1/2* mutant (Figures 6D and S6D). This hints that Shell4 may have a distinct structural role to Shell1/2, in line with their different localizations at the pyrenoid periphery. The more spherical appearance of Rubisco in the absence of the shell is consistent with surface tension effects observed in liquid-liquid phase-separated condensates, suggesting dynamic Rubisco condensation may also play a role in pyrenoid matrix assembly *in vivo* in *T. pseudonana*, as shown *in vitro* for *P. tricornutum*⁷ and as seen for carboxysome assembly.^{46,52}

Consistent with their disrupted pyrenoid morphology, the *shell1/2* and *shell4* mutants had severely reduced growth at atmospheric CO₂, which was fully rescued by supplying elevated CO₂ (Figure 6E), demonstrating that both Shell1/2 and Shell4 are required for a functional CCM. Interestingly, a *shell3* mutant did not show abnormal pyrenoid morphology by TEM (Figures S6C and S6E) and had no growth defect at ambient CO₂ (Figure 6E). Relative to other Shell homologs, Shell3 has a low abundance (Figure 3D) and thus could potentially play a minor role or be compensated by Shell1/2, which have 73% amino acid sequence similarity.

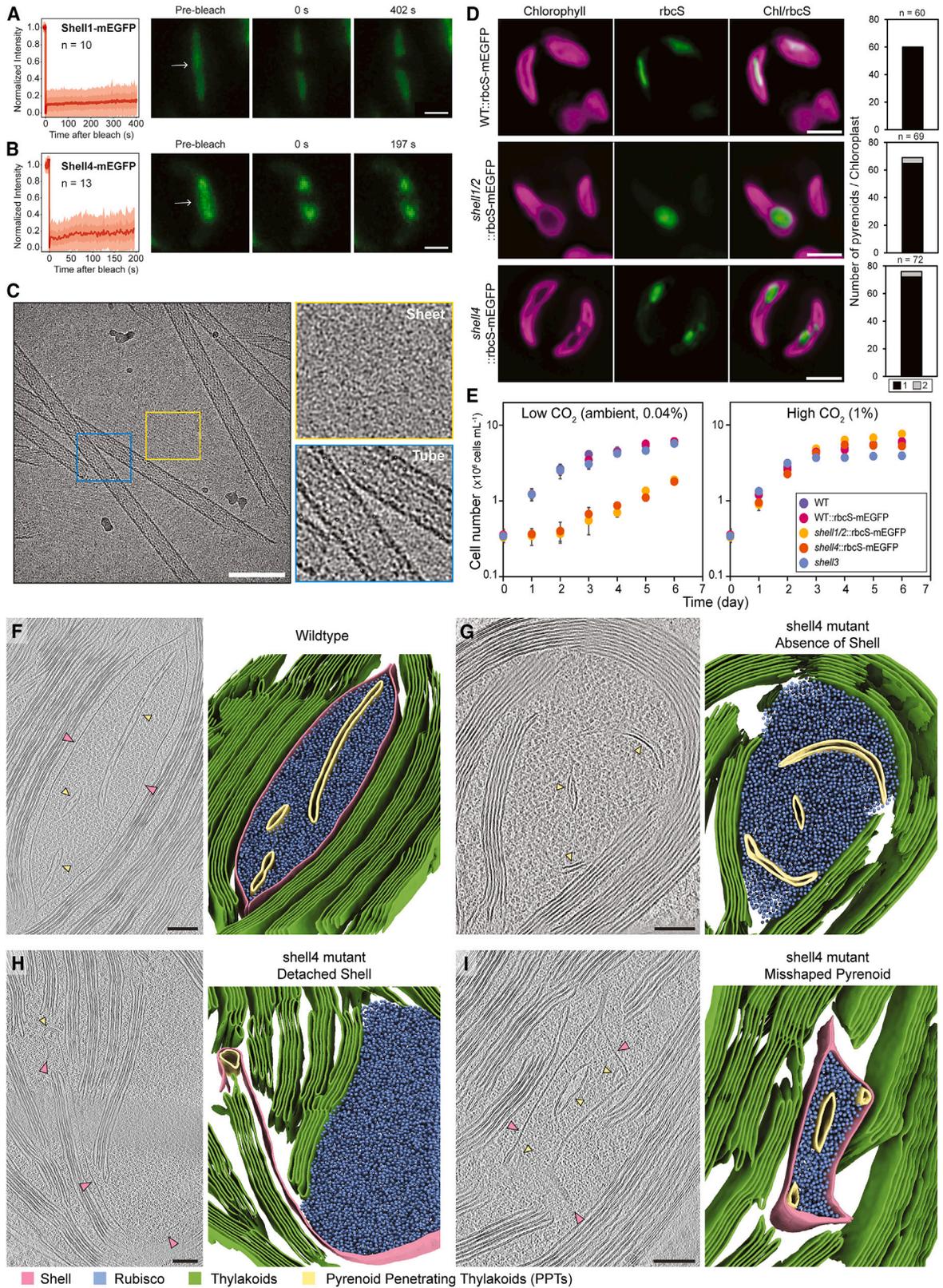
In a parallel study, it was shown that *shell1/2* knockout mutants completely lack a protein shell around the pyrenoid.⁴⁷ To see if our *shell4* mutant has a similar architectural defect, we performed cryo-electron tomography (cryo-ET).⁵³ WT *T. pseudonana* cells contain lenticular-shaped pyrenoids, where the Rubisco matrix is encapsulated in a protein shell and bisected end-to-end by one or two PPTs that contain densities in their lumen (Figure 6F).⁴⁷ In contrast, the *shell4* mutant has misshaped pyrenoids with a diverse range of morphologies (Figures 6G–6I and S7). These include (1) spherical pyrenoids that lack a visible shell and have abnormal PPTs that do not bisect the matrix (similar to the *shell1/2* mutant⁴⁷), (2) pyrenoids with a shell that is detached from the Rubisco matrix, and (3) pyrenoids that are encapsulated in a shell but have an atypical distorted shape and abnormal PPTs. Whole-cell proteomics on the *shell4* mutant showed the almost complete absence of Shell4 but largely unaltered levels of Shell1/2, suggesting that the Shell we observe in the *shell4* mutant is composed of Shell1/2 proteins (Figure S6F; Table S7). Shells in the *shell4* mutant show apparent decreased affinity for the Rubisco matrix (detached shells) as well as apparent increased interaction with PPT membranes (attachment to PPTs located both within the matrix and displaced in the stroma). Interestingly, these mutant shells were also often observed to self-interact, forming double- or triple-layered sheets (Figure S7). We therefore hypothesize that Shell4

(H) Fluorescence intensity cross-sections of DPC3 and Shell4 across the pyrenoid.

(I) z stack max intensity projections of (H).

Scale bars: 2 μ m. For (B), (E), and (H), the level of zoom changes between images.

See also Figure S4.



(legend on next page)

promotes interaction with Rubisco, whereas Shell1/2 have high affinity for PPT membranes and Shell1/2 proteins in adjacent sheets. Taken together, the complete absence of a shell in the *shell1/2* mutant seen by cryo-ET⁴⁷ and the presence of a shell but severely misshaped pyrenoids in the *shell4* mutant suggest that Shell1/2 and Shell4 have distinct roles: Shell1/2 form the main structural component of the shell, and Shell4 is critical for pyrenoid organization, ensuring that the Rubisco matrix, PPT, and shell are correctly assembled together. This is further supported by (1) the whole-cell proteomics, where Shell1/2 are more abundant than Shell4, (2) the pyrenoid interaction network, where Shell4 directly and indirectly is connected to several pyrenoid matrix components but Shell1 has no such connections, and (3) the propensity of Shell1 but not Shell4 to form high order assemblies *in vitro*.

The diverse pyrenoid structural defects in the *shell4* mutant make it difficult to assign the high CO₂ requiring phenotype to a specific pyrenoid structural defect. The disrupted PPTs proposed to be involved in inorganic carbon delivery to Rubisco could reduce inorganic carbon fluxes to the Rubisco matrix. If the shell has a role in restricting CO₂ diffusion, its disruption could make the pyrenoid leaky, thereby reducing the effective CO₂ concentration around Rubisco. The shell could also be critical for efficient delivery and partitioning of Calvin cycle intermediates between the pyrenoid and the surrounding stroma. Finally, pyrenoid shape could be important to maximize membrane contact area and to minimize diffusion distances.

Collectively, these data support that the diatom Shell proteins 1, 2, and 4 are critical for pyrenoid architecture and CCM function and may act as a CO₂ diffusion barrier, although further experimental proof is required for the latter.

Perspective

The development of a high-throughput tagging and APMS pipeline in a model diatom has enabled us to generate a spatial interaction network of the diatom pyrenoid, providing molecular insight into how diatoms help drive the global carbon cycle. We have identified and confirmed via localization 13 previously unknown pyrenoid components, of which 10 have no conserved functional domains. Six of these components constitute a protein shell that encapsulates the pyrenoid and is found across diverse species with red plastids derived from secondary endosymbiotic events. Knockout of the most abundant shell components, Shell1/2 and Shell4, resulted in large pyrenoid structural

changes and poor growth at atmospheric levels of CO₂. Cryo-ET on the *shell4* mutant indicates that Shell4 is critical for correct organization of the PPT, shell, and matrix, while our data along with data in a parallel study⁴⁷ indicate that Shell1/2 is most likely the major structural component of the shell.

Four additional pyrenoid components, DPC1–4, have no clear function that can be predicted from their sequence. DPC3 showed a shell-like localization pattern, co-localizing with Shell4 and interacting with both Shell4 and *rbcS*, suggesting a potential role in mediating shell-Rubisco matrix interactions. The molecular mechanism of Rubisco condensation to form the pyrenoid matrix is currently unknown in *T. pseudonana*, with no homologs or functional analogs of EPYC1 or PYCO1 (Rubisco likers in *C. reinhardtii* and *P. tricornutum*) identified in our study. This suggests that pyrenoid matrix formation may be different in *T. pseudonana*. The matrix localization and interaction partners of DPC2 and DPC4 suggest that they may have a central role in pyrenoid matrix assembly/function. DPC2, DPC3, and DPC4 are prime targets for future characterization.

Close to 50% of global carbon fixation is performed by biomolecular condensates of Rubisco.² This includes prokaryotic cyanobacterial carboxysomes and eukaryotic algal pyrenoids. Nearly all present knowledge of pyrenoid structure and function comes from the green plastid lineage alga, *C. reinhardtii*. However, pyrenoids are proposed to have convergently evolved both between plastid lineages and within plastid lineages.^{2,4} Insights from our data suggest that diatom pyrenoids have similarities to both green plastid pyrenoids and prokaryotic carboxysomes. Similarities between the *T. pseudonana* and *C. reinhardtii* pyrenoid include dense Rubisco packaging around specialized thylakoid membranes (PPTs) for CO₂ delivery, CO₂ delivery to the PPTs via bestrophin family channels,^{28,29} and CO₂ release within the acidic lumen driven by constrained localization of a carbonic anhydrase (*θCA2*; Kikutani et al.²⁵ and Shimakawa et al.²⁶ and this study) within the PPTs. However, the encapsulation by a protein shell layer composed of homologs, some with different subshell localizations, is analogous to carboxysome shell proteins.⁴⁰ Additionally, the static nature of Rubisco, *cbbX*, and shell proteins contrasts with the dynamic properties of the *C. reinhardtii* pyrenoid and aligns more with carboxysomes. Another major outstanding question is the connection of the PPT with the broader thylakoid network; in both our TEM and cryo-ET, we are yet to visualize an unambiguous connection as seen in nearly all pyrenoid-containing algae,² including *C. reinhardtii*⁵⁴ and

Figure 6. Shell proteins are static on the pyrenoid surface, can self-assemble into sheets and tubes, and are essential for CCM function and pyrenoid architecture

(A and B) Fluorescence recovery after photobleaching (FRAP) experiments for Shell1 (A) and Shell4 (B). Arrows indicate photobleached regions. Scale bars: 1 μ m. Plots show SEM and SD of the mean.

(C) Shell1 can form tubes and sheets as seen by cryo-EM. Scale bar: 100 nm.

(D) Tagged *rbcS* in WT (WT::*rbcS*-mEGFP, top), a *shell1/2* mutant (*shell1/2*::*rbcS*-mEGFP, middle), and a *shell4* mutant line (*shell4*::*rbcS*-mEGFP, bottom). Scale bars: 2 μ m. Bar charts indicate pyrenoid number per chloroplast in corresponding line.

(E) Growth rate of WT, WT::*rbcS*-mEGFP, *shell1/2*::*rbcS*-mEGFP, *shell4*::*rbcS*-mEGFP, and a *shell3* mutant. Error bars: SD of the mean.

(F–I) Cryo-ET of WT *T. pseudonana* (F) and the *shell4*::*rbcS*-mEGFP line (G–I). In each panel, the left image shows a 2D slice through the tomogram, and the right image shows the corresponding 3D segmentation of the Shell structure (pink), Rubisco matrix (blue), thylakoids (green), and PPTs (yellow). Scale bars: 100 nm. In total, 25 tomograms of the *shell4* mutant were obtained. Of these, 6/25 displayed Rubisco matrices with a total absence of shell proteins, 12/25 had shell density that only partially enclosed the Rubisco matrix, and 7/25 showed pyrenoids where either the shell was folded multiple times upon itself, the PPTs did not show their canonical organization, or Rubisco packing looked atypically sparse. Additional tomograms of the *shell4* mutant are shown in Figure S7.

See also Figures S5–S7.

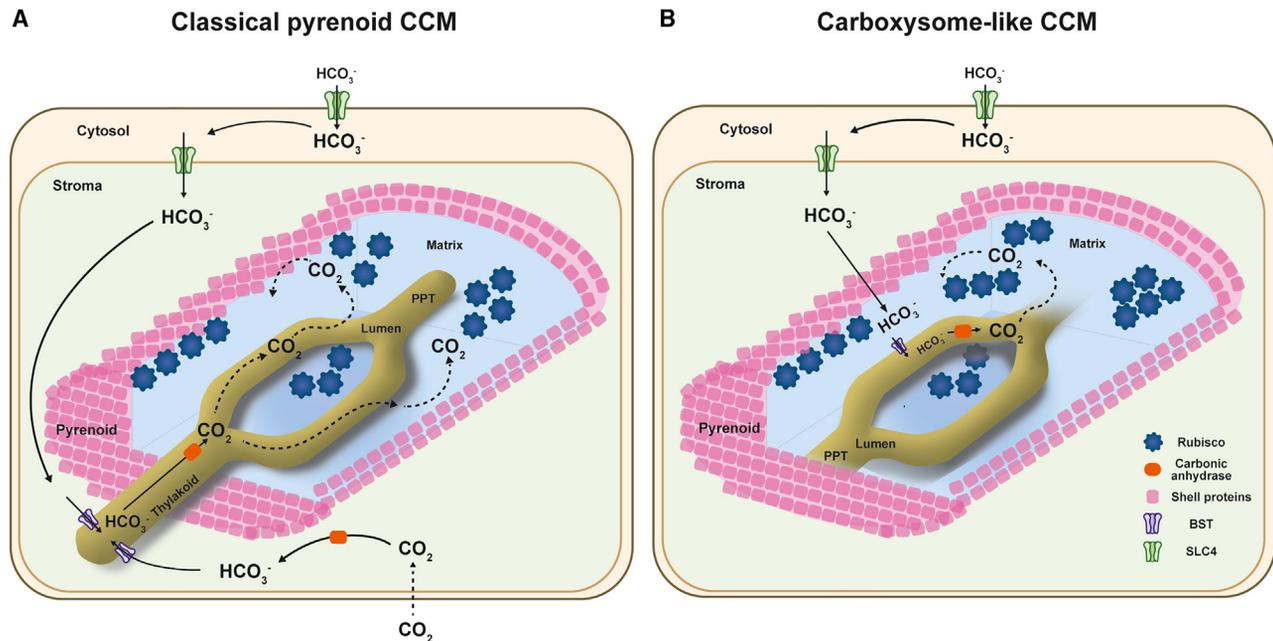


Figure 7. Structural and functional models of the *T. pseudonana* pyrenoid-based CO_2 -concentrating mechanism

(A) Proposed model based on a classical pyrenoid CCM where there is a connection between the PPT and the broader thylakoid network.

(B) Proposed model based on a carboxysome-like CCM, where the shell completely encapsulates the pyrenoid and the PPT is separated from the broader thylakoid network.

Models integrate data from this study with the available literature. For simplicity, the multiple membranes encapsulating the chloroplast and components with unknown or unclear function have been omitted. See text for further discussion.

P. tricornutum.⁴⁷ If there is no connection, this would align the *T. pseudonana* pyrenoid even closer to carboxysomes and require a new functional model for the *T. pseudonana* CCM. It is tempting to postulate that the expansion and diversification of the Shell protein family in *T. pseudonana* vs. *P. tricornutum* (6 homologs vs. 2 homologs; Figure 4B) may be critical for the complete encapsulation of the Rubisco matrix and PPT, with Shell4 a strong candidate that could enable this.

In Figure 7, we propose two models for the *T. pseudonana* CCM: a more classical pyrenoid-based CCM where a PPT-thylakoid connection is present (Figure 7A) and a completely Shell-encapsulated pyrenoid that would align closer to a carboxysome system (Figure 7B). In both models, SLC4 family proteins contribute to sodium dependent HCO_3^- transport at the plasma membrane.^{8,23,24} The mechanism of Ci transport across the four chloroplast membranes is still unknown, although SLC4 transporters are also proposed to play a role here²⁷ along with the carbonic anhydrase LCIP63 (low CO_2 inducible protein of 63 kDa)⁵⁵ and the vacuolar-type ATPase.⁵⁶ In the classical model, Ci delivery to the pyrenoid is potentially analogous to *C. reinhardtii*, relying on channeling of HCO_3^- into the thylakoid lumen via BST1 and BST2.²⁸ HCO_3^- could then diffuse into the PPT, where it is dehydrated to CO_2 via θCA2 ^{25,26} and diffuses to Rubisco packaged within the pyrenoid. In a more carboxysome-like system, HCO_3^- accumulated in the stroma would have to transit across the shell and subsequently be dehydrated to CO_2 . CO_2 release could potentially still occur via channeling into the PPT by bestrophins, and its release accelerated via θCA2 . In the clas-

sical pyrenoid model, like in *C. reinhardtii*, the low luminal pH of the PPTs established by the light reactions of photosynthesis would provide protons for HCO_3^- dehydration and the energetic driving force of the CCM.⁵⁷ In a more carboxysome-like system, protons for HCO_3^- dehydration could come from Rubisco, which has been modeled to produce two protons for every carboxylation reaction.⁵⁸ In both models, CO_2 leakage out of the pyrenoid would be minimized by the proteinaceous shell that encapsulates the pyrenoid. The shell is also critical for maintaining the correct architecture of the pyrenoid, including its lenticular shape that should minimize diffusion distances of CO_2 from the PPTs to Rubisco.

Other than Rubisco, there appear to be no sequence or structural similarities between carboxysomes and *T. pseudonana* pyrenoid proteins. This further supports the convergent evolution of pyrenoids and that a broad range of biophysical, structural, and functional properties, some previously associated with carboxysomes, can be expected as more pyrenoids are characterized across diverse algae.

A core structural component of pyrenoids is a CO_2 leakage barrier, with the starch sheath in *C. reinhardtii* shown both experimentally and theoretically to be required for efficient CCM function.^{17,18} As engineering of a pyrenoid into plants progresses, a major future challenge will be CO_2 diffusion barrier engineering.^{17,22} This is thought to require multiple starch-synthesis-related steps correctly localized to the pyrenoid periphery. The diatom Shell proteins could potentially provide an alternative biotechnology solution to this challenge.

Limitations of the study

Well-documented technical limitations of protein tagging could result in protein mislocalization and inaccurate protein-protein interactions. C-terminal tagging could result in masking interactions that could alter protein targeting and modulate native interactions with other proteins. The use of a constitutive promoter and expression of proteins in trans-to-native copies could also modify protein localization and interaction partners. Reported protein-protein interactions were not validated by a parallel method and may not mean a direct interaction but could be mediated through an additional component. Although DPC3 and Shell4 were shown to interact and co-localize, co-localization does not confirm a direct interaction.

Where possible, we validated two independently tagged lines for localization; however, in some cases, only a single stable line was generated. For *shell4* mutants, four independent lines were sequence verified for biallelic editing. Both pyrenoid shape disruption by *rbcS*-mEGFP imaging and growth defects at ambient CO₂ were seen across all edited lines, and a single representative line was chosen for further detailed studies. For *shell1/2* mutants, multiple lines were generated that resulted in pyrenoid shape disruption by *rbcS*-mEGFP imaging; however, only one was confirmed for biallelic editing of both *Shell1* and *Shell2*; this line was prioritized for further studies.

While we propose that the shell may act as a CO₂ barrier analogous to the carboxysome shell, we provide no direct evidence for this. We also do not understand how the Rubisco substrate, ribulose 1,5-bisphosphate (RuBP), and product, 3-phosphoglycerate (PGA), cross the shell. Although we show that *Shell1* and *Shell4* do not co-localize, we do not understand how the other *Shell* components are orientated relatively to each other, whether *Shell* proteins other than *Shell1* can homo-oligomerize, and whether *Shell* proteins can hetero-oligomerize. All our images are snapshots of living cells. Proteins could change localization and interactions depending on cell-cycle state and during pyrenoid division. Further, the dynamics of *Shell1*, *Shell4*, Rubisco, and *cbbX* may change at different states of the cell cycle to enable pyrenoid division and pyrenoid growth. Finally, how the pyrenoid matrix is assembled is still an open question.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Luke C.M. Mackinder (luke.mackinder@york.ac.uk).

Materials availability

All unique/stable reagents generated in this study are available from the [lead contact](#) without restriction.

Data and code availability

- Mass spectrometry and proteomic identification data are referenced in ProteomeXchange (ProteomeXchange: PXD052522) and can be accessed via MassIVE (MassIVE: MSV000094846). Cryo-ET reconstructed tomograms are in the Electron Microscopy Data Bank (EMDB: EMD-51423, EMD-51424, EMD-51425, EMD-51427), and raw data are in the Electron Microscopy Public Image Archive (EMPIAR: EMPIAR-12250). All data are publicly available as of the date of publica-

tion. In addition, this paper analyzes existing, publicly available data, accessible at EMBL/Genbank/DBJ databases under the accession: GCA_000149405.1.

- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

We thank Oliver Mueller-Cajar for kindly sharing the anti-rbcL antibody, the Technology Facility in the York Department of Biology for the access/support for microscopy and flow cytometry, and members of the Mackinder lab for fruitful discussions. Thanks to Glenn Harper for optimizing the TEM. We thank Alexander Neuhaus (cryo-EM facility, University of Münster), Carlos Fernandez Rodriguez and Mohamed Chami (BioEM facility, University of Basel), and Daniel Böhringer, Miroslav Peterek, and Bilal M. Qureshi (ETH Zürich ScopeM facility) for access to FIB and TEM instruments and assistance with cryo-ET data acquisition as well as Wojciech Wietrzynski for cryo-ET data acquisition. Cryo-ET analysis was performed at the sciCORE (<http://scicore.unibas.ch/>) scientific computing center at the University of Basel. The York Glacios was funded by the Wellcome Trust (206161/Z/17/Z), in space supported by the Wolfson Foundation and Dr. Tony Wild. We thank Johan Turkenburg and Sam Hart for assistance with facility operations. The York Centre of Excellence in Mass Spectrometry was created thanks to a major capital investment through Science City York, supported by Yorkshire Forward with funds from the Northern Way Initiative, and subsequent support from EPSRC (EP/K039660/1 and EP/M028127/1). B.D.E. acknowledges funding from ERC consolidator grant “cryocean” (fulfilled by the Swiss State Secretariat for Education, Research, and Innovation, M822.00045). L.C.M.M. was supported by a UKRI Future Leader Fellowship (MR/T020679/1), EPSRC funding (EP/W024063/1) as part of the York Physics of Pyrenoids Project (YP3), and BBSRC funding (BB/S015337/1 and BB/X004953/1).

AUTHOR CONTRIBUTIONS

O.N. and L.C.M.M. designed and supervised the study. O.N. carried out the experiments unless otherwise mentioned. M.D. and B.D.E. performed cryo-ET of the *shell4* mutant. S.M. purified the *Shell1* protein and carried out cryo-EM with support from J.N.B. C.M. and O.N. performed the phylogenetic analysis of *Shell* proteins. A.D., M.D., and J.B. provided bioinformatics and data analysis support. A.D. oversaw the mass spectrometry and peptide mapping. O.N., C.M., and L.C.M.M. analyzed and interpreted the data. O.N. and M.D. created the figures. O.N. and L.C.M.M. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

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

- [KEY RESOURCES TABLE](#)
- [EXPERIMENTAL MODELS AND STUDY PARTICIPANT DETAILS](#)
 - Strains and Culturing
- [METHOD DETAILS](#)
 - Episome Assemblies using Golden Gate Cloning
 - Genome editing
 - *T. pseudonana* Transformation via Bacterial Conjugation
 - Flow Cytometry
 - Confocal Fluorescent Microscopy
 - Co-Immunoprecipitation and Affinity Purification
 - Mass Spectrometry
 - Protein-Protein Interaction Network analysis
 - *In vivo* Fluorescence Recovery After Photobleaching

- Data visualization
- Structure Prediction
- Transmission Electron Microscopy
- Blast search
- Phylogeny
- Recombinant Shell1 protein purification
- Cryo-Electron Microscopy of recombinantly expressed Shell1
- Cellular Cryo-Electron Tomography

● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2024.09.025>.

Received: October 17, 2023

Revised: June 18, 2024

Accepted: September 13, 2024

Published: October 4, 2024

REFERENCES

1. Mackinder, L.C.M., Meyer, M.T., Mettler-Altmann, T., Chen, V.K., Mitchell, M.C., Caspari, O., Freeman Rosenzweig, E.S., Pallesen, L., Reeves, G., Itakura, A., et al. (2016). A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. *Proc. Natl. Acad. Sci. USA* *113*, 5958–5963. <https://doi.org/10.1073/pnas.1522866113>.
2. Barrett, J., Girr, P., and Mackinder, L.C.M. (2021). Pyrenoids: CO₂-fixing phase separated liquid organelles. *Biochim. Biophys. Acta Mol. Cell Res.* *1868*, 118949. <https://doi.org/10.1016/j.bbamcr.2021.118949>.
3. Raven, J.A., Cockell, C.S., and De La Rocha, C.L. (2008). The evolution of inorganic carbon concentrating mechanisms in photosynthesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* *363*, 2641–2650. <https://doi.org/10.1098/rstb.2008.0020>.
4. Meyer, M.T., Goudet, M.M.M., and Griffiths, H. (2020). The Algal Pyrenoid. In *Photosynthesis in Algae: Biochemical and Physiological Mechanisms*, A.W.D. Larkum, A.R. Grossmann, and J.A. Raven, eds. (Springer International Publishing), pp. 179–203. https://doi.org/10.1007/978-3-030-33397-3_9.
5. He, S., Crans, V.L., and Jonikas, M.C. (2023). The pyrenoid: the eukaryotic CO₂-concentrating organelle. *Plant Cell* *35*, 3236–3259. <https://doi.org/10.1093/plcell/koad157>.
6. Findinier, J., and Grossman, A.R. (2023). *Chlamydomonas*: Fast tracking from genomics. *J. Phycol.* *59*, 644–652. <https://doi.org/10.1111/jpy.13356>.
7. Oh, Z.G., Ang, W.S.L., Poh, C.W., Lai, S.-K., Sze, S.K., Li, H.-Y., Bhushan, S., Wunder, T., and Mueller-Cajar, O. (2023). A linker protein from a red-type pyrenoid phase separates with Rubisco via oligomerizing sticker motifs. *Proc. Natl. Acad. Sci. USA* *120*, e2304833120. <https://doi.org/10.1073/pnas.2304833120>.
8. Tsuji, Y., Nakajima, K., and Matsuda, Y. (2017). Molecular aspects of the biophysical CO₂-concentrating mechanism and its regulation in marine diatoms. *J. Exp. Bot.* *68*, 3763–3772. <https://doi.org/10.1093/jxb/erx173>.
9. Nawaly, H., Tanaka, A., Toyoshima, Y., Tsuji, Y., and Matsuda, Y. (2023). Localization and characterization of carbonic anhydrases in *Thalassiosira pseudonana*. *Photosynth. Res.* *156*, 217–229. <https://doi.org/10.1007/s11120-023-01007-z>.
10. Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* *281*, 237–240. <https://doi.org/10.1126/science.281.5374.237>.
11. Benoiston, A.-S., Ibarbalz, F.M., Bittner, L., Guidi, L., Jahn, O., Dutkiewicz, S., and Bowler, C. (2017). The evolution of diatoms and their biogeochemical functions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* *372*, 20160397. <https://doi.org/10.1098/rstb.2016.0397>.
12. 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., et al. (2017). The Eukaryotic CO₂-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization. *Cell* *171*, 148–162.e19. <https://doi.org/10.1016/j.cell.2017.08.008>.
13. Wunder, T., Cheng, S.L.H., Lai, S.-K., Li, H.-Y., and Mueller-Cajar, O. (2018). The phase separation underlying the pyrenoid-based microalgal Rubisco supercharger. *Nat. Commun.* *9*, 5076. <https://doi.org/10.1038/s41467-018-07624-w>.
14. Meyer, M.T., Itakura, A.K., Patena, W., Wang, L., He, S., Emrich-Mills, T., Lau, C.S., Yates, G., Mackinder, L.C.M., and Jonikas, M.C. (2020). Assembly of the algal CO₂-fixing organelle, the pyrenoid, is guided by a RuBisCO-binding motif. *Sci. Adv.* *6*, eabd2408. <https://doi.org/10.1126/sciadv.abd2408>.
15. 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., et al. (2019). Thylakoid localized bestrophin-like proteins are essential for the CO₂ concentrating mechanism of *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. USA* *116*, 16915–16920. <https://doi.org/10.1073/pnas.1909706116>.
16. Sinetova, M.A., Kupriyanova, E.V., Markelova, A.G., Allakhverdiev, S.I., and Pronina, N.A. (2012). Identification and functional role of the carbonic anhydrase Cah3 in thylakoid membranes of pyrenoid of *Chlamydomonas reinhardtii*. *Biochim. Biophys. Acta* *1817*, 1248–1255. <https://doi.org/10.1016/j.bbabi.2012.02.014>.
17. Fei, C., Wilson, A.T., Mangan, N.M., Wingreen, N.S., and 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. *Nat. Plants* *8*, 583–595. <https://doi.org/10.1038/s41477-022-01153-7>.
18. Toyokawa, C., Yamano, T., and Fukuzawa, H. (2020). Pyrenoid starch sheath is required for LCIB localization and the CO₂-concentrating mechanism in green algae. *Plant Physiol.* *182*, 1883–1893. <https://doi.org/10.1104/pp.19.01587>.
19. Mackinder, L.C.M., Chen, C., Leib, R.D., Patena, W., Blum, S.R., Rodman, M., Ramundo, S., Adams, C.M., and Jonikas, M.C. (2017). A spatial interactome reveals the protein organization of the algal CO₂-concentrating mechanism. *Cell* *171*, 133–147.e14. <https://doi.org/10.1016/j.cell.2017.08.044>.
20. Lau, C.S., Dowle, A., Thomas, G.H., Girr, P., and Mackinder, L.C.M. (2023). A phase-separated CO₂-fixing pyrenoid proteome determined by TurboID in *Chlamydomonas reinhardtii*. *Plant Cell* *35*, 3260–3279. <https://doi.org/10.1093/plcell/koad131>.
21. Atkinson, N., Mao, Y., Chan, K.X., and McCormick, A.J. (2020). Condensation of Rubisco into a proto-pyrenoid in higher plant chloroplasts. *Nat. Commun.* *11*, 6303. <https://doi.org/10.1038/s41467-020-20132-0>.
22. Adler, L., Díaz-Ramos, A., Mao, Y., Pukacz, K.R., Fei, C., and McCormick, A.J. (2022). New horizons for building pyrenoid-based CO₂-concentrating mechanisms in plants to improve yields. *Plant Physiol.* *190*, 1609–1627. <https://doi.org/10.1093/plphys/kiac373>.
23. Nakajima, K., Tanaka, A., and Matsuda, Y. (2013). SLC4 family transporters in a marine diatom directly pump bicarbonate from seawater. *Proc. Natl. Acad. Sci. USA* *110*, 1767–1772. <https://doi.org/10.1073/pnas.1216234110>.
24. Nawaly, H., Matsui, H., Tsuji, Y., Iwayama, K., Ohashi, H., Nakajima, K., and Matsuda, Y. (2023). Multiple plasma membrane SLC4s contribute to external HCO₃⁻ acquisition during CO₂ starvation in the marine diatom *Phaeodactylum tricatum*. *J. Exp. Bot.* *74*, 296–307. <https://doi.org/10.1093/jxb/erac380>.
25. Kikutani, S., Nakajima, K., Nagasato, C., Tsuji, Y., Miyatake, A., and Matsuda, Y. (2016). Thylakoid luminal θ -carbonic anhydrase critical for growth and photosynthesis in the marine diatom *Phaeodactylum tricatum*. *Proc. Natl. Acad. Sci. USA* *113*, 9828–9833. <https://doi.org/10.1073/pnas.1603112113>.

26. Shimakawa, G., Okuyama, A., Harada, H., Nakagaito, S., Toyoshima, Y., Nagata, K., and Matsuda, Y. (2023). Pyrenoid-core CO₂-evolving machinery is essential for diatom photosynthesis in elevated CO₂. *Plant Physiol.* 193, 2298–2305. <https://doi.org/10.1093/plphys/kiad475>.
27. Matsuda, Y., Hopkinson, B.M., Nakajima, K., Dupont, C.L., and Tsuji, Y. (2017). Mechanisms of carbon dioxide acquisition and CO₂ sensing in marine diatoms: a gateway to carbon metabolism. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372, 20160403. <https://doi.org/10.1098/rstb.2016.0403>.
28. Nam, O., Grouneva, I., and Mackinder, L.C.M. (2022). Endogenous GFP tagging in the diatom *Thalassiosira pseudonana*. Preprint at bioRxiv. <https://doi.org/10.1101/2022.09.30.510313>.
29. Nigishi, M., Shimakawa, G., Yamagishi, K., Amano, R., Ito, S., Tsuji, Y., Nagasato, C., and Matsuda, Y. (2024). Low-CO₂-inducible bestrophins outside the pyrenoid sustain high photosynthetic efficacy in diatoms. *Plant Physiol.* 195, 1432–1445. <https://doi.org/10.1093/plphys/kiad137>.
30. Wang, S., Wu, R., Lu, J., Jiang, Y., Huang, T., and Cai, Y.-D. (2022). Protein-protein interaction networks as miners of biological discovery. *Proteomics* 22, e2100190. <https://doi.org/10.1002/pmic.202100190>.
31. Huang, W., and Daboussi, F. (2017). Genetic and metabolic engineering in diatoms. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372, 20160411. <https://doi.org/10.1098/rstb.2016.0411>.
32. Huesgen, P.F., Alami, M., Lange, P.F., Foster, L.J., Schröder, W.P., Overall, C.M., and Green, B.R. (2013). Proteomic amino-terminal profiling reveals targeting information for protein import into complex plastids. *PLoS One* 8, e74483. <https://doi.org/10.1371/journal.pone.0074483>.
33. Clement, R., Dimnet, L., Maberly, S.C., and Gontero, B. (2016). The nature of the CO₂-concentrating mechanisms in a marine diatom, *Thalassiosira pseudonana*. *New Phytol.* 209, 1417–1427. <https://doi.org/10.1111/nph.13728>.
34. Sowa, M.E., Bennett, E.J., Gygi, S.P., and Harper, J.W. (2009). Defining the human deubiquitinating enzyme interaction landscape. *Cell* 138, 389–403. <https://doi.org/10.1016/j.cell.2009.04.042>.
35. Choi, H., Larsen, B., Lin, Z.-Y., Breitkreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z.S., Tyers, M., Gingras, A.-C., and Nesvizhskii, A.I. (2011). SAINT: probabilistic scoring of affinity purification–mass spectrometry data. *Nat. Methods* 8, 70–73. <https://doi.org/10.1038/nmeth.1541>.
36. Mueller-Cajar, O., Stotz, M., Wendler, P., Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Structure and function of the AAA+ protein CbbX, a red-type RuBisCO activase. *Nature* 479, 194–199. <https://doi.org/10.1038/nature10568>.
37. Meyer, M.T., Whittaker, C., and Griffiths, H. (2017). The algal pyrenoid: key unanswered questions. *J. Exp. Bot.* 68, 3739–3749. <https://doi.org/10.1093/jxb/erx178>.
38. Barrett, J., Naduthodi, M.I.S., Mao, Y., Dégut, C., Musial, S., Salter, A., Leake, M.C., Plevin, M.J., McCormick, A.J., Blaza, J.N., et al. (2024). A promiscuous mechanism to phase separate eukaryotic carbon fixation in the green lineage. Preprint at bioRxiv. <https://doi.org/10.1101/2024.04.09.588658>.
39. Kroth, P.G., Chiovitti, A., Gruber, A., Martin-Jezequel, V., Mock, T., Parker, M.S., Stanley, M.S., Kaplan, A., Caron, L., Weber, T., et al. (2008). A model for carbohydrate metabolism in the diatom *Phaeodactylum tricoratum* deduced from comparative whole genome analysis. *PLoS One* 3, e1426. <https://doi.org/10.1371/journal.pone.0001426>.
40. Rae, B.D., Long, B.M., Badger, M.R., and Price, G.D. (2013). Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO₂ fixation in cyanobacteria and some proteobacteria. *Microbiol. Mol. Biol. Rev.* 77, 357–379. <https://doi.org/10.1128/MMBR.00061-12>.
41. Kerfeld, C.A., and Melnicki, M.R. (2016). Assembly, function and evolution of cyanobacterial carboxysomes. *Curr. Opin. Plant Biol.* 31, 66–75. <https://doi.org/10.1016/j.pbi.2016.03.009>.
42. Cameron, J.C., Wilson, S.C., Bernstein, S.L., and Kerfeld, C.A. (2013). Biogenesis of a bacterial organelle: the carboxysome assembly pathway. *Cell* 155, 1131–1140. <https://doi.org/10.1016/j.cell.2013.10.044>.
43. Long, B.M., Rae, B.D., Rolland, V., Förster, B., and Price, G.D. (2016). Cyanobacterial CO₂-concentrating mechanism components: function and prospects for plant metabolic engineering. *Curr. Opin. Plant Biol.* 31, 1–8. <https://doi.org/10.1016/j.pbi.2016.03.002>.
44. Kong, W.-W., Zhu, Y., Zhao, H.-R., Du, K., Zhou, R.-Q., Li, B., Yang, F., Hou, P., Huang, X.-H., Chen, Y., et al. (2024). Cryo-electron tomography reveals the packaging pattern of RuBisCOs in *Synechococcus* β-carboxysome. *Structure* 32, 1110–1120.e4. <https://doi.org/10.1016/j.str.2024.05.007>.
45. Evans, S.L., Al-Hazeem, M.M.J., Mann, D., Smetacek, N., Beavil, A.J., Sun, Y., Chen, T., Dykes, G.F., Liu, L.-N., and Bergeron, J.R.C. (2023). Single-particle cryo-EM analysis of the shell architecture and internal organization of an intact α-carboxysome. *Structure* 31, 677–688.e4. <https://doi.org/10.1016/j.str.2023.03.008>.
46. Turnšek, J.B., Oltrogge, L.M., and Savage, D.F. (2024). Conserved and repetitive motifs in an intrinsically disordered protein drive α-carboxysome assembly. *J. Biol. Chem.* 300, 107532. <https://doi.org/10.1016/j.jbc.2024.107532>.
47. Shimakawa, G., Demulder, M., Flori, S., Kawamoto, A., Tsuji, Y., Nawaly, H., Tanaka, A., Tohda, R., Ota, T., Matsui, H., et al. (2024). Diatom pyrenoids are encased in a protein shell that enables efficient CO₂ fixation. *Cell* 187, 5919–5934. <https://doi.org/10.1016/j.cell.2024.09.013>.
48. Metskas, L.A., Ortega, D., Oltrogge, L.M., Blikstad, C., Lovejoy, D.R., Laughlin, T.G., Savage, D.F., and Jensen, G.J. (2022). RuBisCO forms a lattice inside alpha-carboxysomes. *Nat. Commun.* 13, 4863. <https://doi.org/10.1038/s41467-022-32584-7>.
49. Ni, T., Sun, Y., Burn, W., Al-Hazeem, M.M.J., Zhu, Y., Yu, X., Liu, L.-N., and Zhang, P. (2022). Structure and assembly of cargo RuBisCO in two native α-carboxysomes. *Nat. Commun.* 13, 4299. <https://doi.org/10.1038/s41467-022-32004-w>.
50. Sutter, M., Faulkner, M., Aussignargues, C., Paasch, B.C., Barrett, S., Kerfeld, C.A., and Liu, L.-N. (2016). Visualization of bacterial microcompartment facet assembly using high-speed atomic force microscopy. *Nano Lett.* 16, 1590–1595. <https://doi.org/10.1021/acs.nanolett.5b04259>.
51. Noël, C.R., Cai, F., and Kerfeld, C.A. (2016). Purification and characterization of protein nanotubes assembled from a single bacterial microcompartment shell subunit. *Adv. Mater. Interfaces* 3, 1500295. <https://doi.org/10.1002/admi.201500295>.
52. Wang, H., Yan, X., Aigner, H., Bracher, A., Nguyen, N.D., Hee, W.Y., Long, B.M., Price, G.D., Hartl, F.U., and Hayer-Hartl, M. (2019). RuBisCO condensate formation by CcmM in β-carboxysome biogenesis. *Nature* 566, 131–135. <https://doi.org/10.1038/s41586-019-0880-5>.
53. Young, L.N., and Villa, E. (2023). Bringing Structure to Cell Biology with Cryo-Electron Tomography. *Annu. Rev. Biophys.* 52, 573–595. <https://doi.org/10.1146/annurev-biophys-111622-091327>.
54. Engel, B.D., Schaffer, M., Kuhn Cuellar, L., Villa, E., Plietzko, J.M., and Baumeister, W. (2015). Native architecture of the *Chlamydomonas* chloroplast revealed by in situ cryo-electron tomography. *eLife* 4, e04889. <https://doi.org/10.7554/eLife.04889>.
55. Jensen, E.L., Clement, R., Kosta, A., Maberly, S.C., and Gontero, B. (2019). A new widespread subclass of carbonic anhydrase in marine phytoplankton. *ISME J.* 13, 2094–2106. <https://doi.org/10.1038/s41396-019-0426-8>.
56. Yee, D.P., Samo, T.J., Abbriano, R.M., Shimasaki, B., Vernet, M., Mayali, X., Weber, P.K., Mitchell, B.G., Hildebrand, M., Decelle, J., et al. (2023). The V-type ATPase enhances photosynthesis in marine phytoplankton and further links phagocytosis to symbiogenesis. *Curr. Biol.* 33, 2541–2547.e5. <https://doi.org/10.1016/j.cub.2023.05.020>.
57. Burlacot, A., Dao, O., Auroy, P., Cuié, S., Li-Beisson, Y., and Peltier, G. (2022). Alternative photosynthesis pathways drive the algal

- CO₂-concentrating mechanism. *Nature* 605, 366–371. <https://doi.org/10.1038/s41586-022-04662-9>.
58. Long, B.M., Förster, B., Pulsford, S.B., Price, G.D., and Badger, M.R. (2021). Rubisco proton production can drive the elevation of CO₂ within condensates and carboxysomes. *Proc. Natl. Acad. Sci. USA* 118, e2014406118. <https://doi.org/10.1073/pnas.2014406118>.
 59. Strand, T.A., Lale, R., Degnes, K.F., Lando, M., and Valla, S. (2014). A new and improved host-independent plasmid system for RK2-based conjugal transfer. *PLoS One* 9, e90372. <https://doi.org/10.1371/journal.pone.0090372>.
 60. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. <https://doi.org/10.1038/nmeth.2019>.
 61. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nat. Methods* 19, 679–682. <https://doi.org/10.1038/s41592-022-01488-1>.
 62. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493–500. <https://doi.org/10.1038/s41586-024-07487-w>.
 63. Zheng, S.Q., Palovcak, E., Armache, J.-P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* 14, 331–332. <https://doi.org/10.1038/nmeth.4193>.
 64. Lamm, L., Zufferey, S., Righetto, R.D., Wietrzynski, W., Yamauchi, K.A., Burt, A., Liu, Y., Zhang, H., Martinez-Sanchez, A., Ziegler, S., et al. (2024). MemBrain v2: an end-to-end tool for the analysis of membranes in cryo-electron tomography. Preprint at bioRxiv. <https://doi.org/10.1101/2024.01.05.574336>.
 65. Mastronarde, D.N., and Held, S.R. (2017). Automated tilt series alignment and tomographic reconstruction in IMOD. *J. Struct. Biol.* 197, 102–113. <https://doi.org/10.1016/j.jsb.2016.07.011>.
 66. Buchholz, T.-O., Krull, A., Shahidi, R., Pigino, G., Jékely, G., and Jug, F. (2019). Content-aware image restoration for electron microscopy. *Methods Cell Biol.* 152, 277–289. <https://doi.org/10.1016/bs.mcb.2019.05.001>.
 67. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* 30, 70–82. <https://doi.org/10.1002/pro.3943>.
 68. Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* 152, 36–51. <https://doi.org/10.1016/j.jsb.2005.07.007>.
 69. Sievers, F., and Higgins, D.G. (2014). Clustal omega. *Curr. Protoc. Bioinformatics* 48, 3.13.1–3.13.16. <https://doi.org/10.1002/0471250953.bi0313s48>.
 70. Guillard, R.R., and Rytner, J.H. (1962). Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt, and *Detonula confervacea* (Cleve) Gran. *Can. J. Microbiol.* 8, 229–239. <https://doi.org/10.1139/m62-029>.
 71. Guillard, R.R.L. (1975). Culture of Phytoplankton for Feeding Marine Invertebrates. In *Culture of Marine Invertebrate Animals*, W.L. Smith and M.H. Chanley, eds. (Springer), pp. 29–60.
 72. Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PLoS One* 6, e16765. <https://doi.org/10.1371/journal.pone.0016765>.
 73. Karas, B.J., Diner, R.E., Lefebvre, S.C., McQuaid, J., Phillips, A.P.R., Noddings, C.M., Brunson, J.K., Valas, R.E., Deerinck, T.J., Jablanovic, J., et al. (2015). Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* 6, 6925.
 74. Concordet, J.-P., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* 46, W242–W245. <https://doi.org/10.1093/nar/gky354>.
 75. Park, J., Bae, S., and Kim, J.-S. (2015). Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. *Bioinformatics* 31, 4014–4016. <https://doi.org/10.1093/bioinformatics/btv537>.
 76. Tan, S. (2001). A modular polycistronic expression system for overexpressing protein complexes in *Escherichia coli*. *Protein Expr. Purif.* 21, 224–234. <https://doi.org/10.1006/prep.2000.1363>.
 77. Schaffer, M., Mahamid, J., Engel, B.D., Laugks, T., Baumeister, W., and Plitzko, J.M. (2017). Optimized cryo-focused ion beam sample preparation aimed at in situ structural studies of membrane proteins. *J. Struct. Biol.* 197, 73–82. <https://doi.org/10.1016/j.jsb.2016.07.010>.
 78. Hagen, W.J.H., Wan, W., and Briggs, J.A.G. (2017). Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* 197, 191–198. <https://doi.org/10.1016/j.jsb.2016.06.007>.
 79. Khavnekar, S., Erdmann, P.S., and Wan, W. (2024). TOMOMAN: a software package for large scale cryo-electron tomography data preprocessing, community data sharing, and collaborative computing. Preprint at bioRxiv. <https://doi.org/10.1101/2024.05.02.589639>.
 80. Grant, T., and Grigorieff, N. (2015). Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* 4, e06980. <https://doi.org/10.7554/eLife.06980>.
 81. Hrabe, T., Chen, Y., Pfeffer, S., Cuellar, L.K., Mangold, A.-V., and Förster, F. (2012). PyTom: a python-based toolbox for localization of macromolecules in cryo-electron tomograms and subtomogram analysis. *J. Struct. Biol.* 178, 177–188. <https://doi.org/10.1016/j.jsb.2011.12.003>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-rbcL	Oliver Mueller-Cajar; Oh et al. ⁷	N/A
anti-GFP nanobody-trap magnetic agarose beads	ChromoTek	Cat#gtma-20; RRID:AB_2631358
Bacterial and virus strains		
<i>E. coli</i> : strain DH5 α	Centre for Novel Agricultural Products, York	DH5 α
<i>E. coli</i> : strain C41 (DE3)	Centre for Novel Agricultural Products, York	C41 (DE3)
Chemicals, peptides, and recombinant proteins		
Protein A Dynabeads	Invitrogen	Cat#10001D
cOmplete, EDTA-free Protease Inhibitor Cocktail	Roche	Cat#11836170001
Digitonin	Sigma-Aldrich	Cat#D141
Critical commercial assays		
Phusion High-Fidelity DNA polymerase	Thermo Fisher Scientific	Cat#F530L
QIAprep Spin Miniprep Kit	Qiagen	Cat#27106
Deposited data		
Raw mass spectrometry data	This paper	ProteomeXchange: PXD052522; MassIVE: MSV000094846
<i>Thalassiosira pseudonana</i> genome assembly (ASM14940v2)	Joint Genome Institute	https://protists.ensembl.org/Thalassiosira_pseudonana/Info/Annotation/
Cellular tomograms	This paper	EMDB: EMD-51423, EMD-51424, EMD-51425, EMD-51427
Raw cryo-ET data	This paper	EMPIAR: EMPIAR-12250
<i>Thalassiosira antarctica</i> Rubisco model	Protein Data Bank	PDB: 5MZZ
NCBI protein database	NCBI	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Experimental models: Organisms/strains		
<i>T. pseudonana</i> : wild-type	Scottish Culture Collection of Algae and Protozoa	CCAP1085/12 (equivalent to CCMP1335)
<i>T. pseudonana</i> : rbcS-mEGFP	This paper	N/A
<i>T. pseudonana</i> : <i>shell1/2</i> KO::rbcS-mEGFP	This paper	N/A
<i>T. pseudonana</i> : <i>shell4</i> KO::rbcS-mEGFP	This paper	N/A
Oligonucleotides		
<i>shell1/2</i> KO sgRNA1	This study	N/A
<i>shell1/2</i> KO sgRNA2	This study	N/A
<i>shell4</i> KO sgRNA1	This study	N/A
<i>shell4</i> KO sgRNA2	This study	N/A
Recombinant DNA		
pTA-MOB	Rahmi Lale; Strand et al. ⁵⁹	N/A
pLM1015 and pLM1016-rbcS	This study	N/A
pLM1019 and pLM1039-Shell1	This study	N/A
pLM1020-Shell2	This study	N/A
pLM1056-Shell4	This study	N/A
pLM1369-pOPT-His6-MBP-Shell1-GST	This study	N/A
Diatom MoClo parts are available upon request	Nam et al. ²⁸	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Fiji	Schindelin et al. ⁶⁰	https://imagej.net/software/fiji/downloads
Cytoscape	Cytoscape	https://www.cytoscape.org/
AlphaFold2	Mirdita et al. ⁶¹	https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb
AlphaFold3	Abramson et al. ⁶²	https://alphafoldserver.com/
Geneious Prime	Geneious	https://www.geneious.com/updates/geneious-prime-2023-2
CytExpert	Beckman Coulter	Version 2.4
MotionCorr v2.1	Zheng et al. ⁶³	https://emcore.ucsf.edu/ucsf-software
MemBrain v2	Lamm et al. ⁶⁴	https://github.com/teamtomo/membrain-seg
Amira v 2021.2	FEI, Thermo Fisher Scientific	https://www.thermofisher.com/us/en/home/industrial/electron-microscopy/electron-microscopy-instruments-workflow-solutions/3d-visualization-analysis-software/amira-life-sciences-biomedical.html
IMOD v4.11	Mastronarde and Held ⁶⁵	https://bio3d.colorado.edu/imod/index.html
Cryo-CARE v0.2.1	Buchholz et al. ⁶⁶	https://github.com/juglab/cryoCARE_T2T
UCSF ChimeraX	Pettersen et al. ⁶⁷	https://www.rbvi.ucsf.edu/chimerax/
Serial EM software	Mastronarde ⁶⁸	https://bio3d.colorado.edu/SerialEM/
Tomography 5.11 software	Thermo Fisher Scientific	https://www.thermofisher.com/ch/en/home/electron-microscopy/products/software-em-3d-vis/tomography-software.html
Clustal Omega	Sievers and Higgins ⁶⁹	https://www.ebi.ac.uk/jdispatcher/msa/clustalo
GraphPad Prism version 10	GraphPad	https://www.graphpad.com/
Microsoft Office	Microsoft	https://www.office.com/
Adobe Illustrator	Adobe	https://www.adobe.com/uk/creativecloud.html
SAINT	Choi et al. ³⁵	http://sourceforge.net/projects/saint-apms/files/
cRomppass	Sowa et al. ³⁴	https://github.com/dnusinow/cRomppass/blob/master/R/cRomppass.R
Other		
Poly-L-lysine-coated 8 well m-slides	Ibidi	Cat#80824
Digital haemocytometer Countess II FL	Thermo Fisher Scientific	Cat#AMQAF1000 (Discontinued)
CytoFLEX	Beckman Coulter	LX355 or 375
Confocal Microscope	Zeiss	LSM880 or 980
R 2/1 holey carbon-foil 200-mesh copper EM grids	Quantifoil Micro Tools	Cat#N1-c15ncu20-01
Autogrid supports	Thermo Fisher Scientific	N/A
Aquilos 2	Thermo Fisher Scientific	N/A
Titan Krios 300kV	Thermo Fisher Scientific	N/A
Vitrobot plunger	Thermo Fisher Scientific	N/A

EXPERIMENTAL MODELS AND STUDY PARTICIPANT DETAILS

Strains and Culturing

The background *Thalassiosira pseudonana* strain for all experiments was wildtype (WT) CCAP1085/12 (Scottish Culture Collection of Algae and Protozoa, equivalent to CCMP1335). WT cells were axenically maintained in artificial seawater (ASW) (32 g L⁻¹, Instant Ocean SS15-10) supplemented with half-strength (F/2) Guillard F solution^{70,71} at 20°C under continuous illumination of ~50 μmol photons m⁻² s⁻¹. All strains were grown at ambient CO₂ except *shell1/2* deletion, *shell4* deletion and *Shell1*-mEGFP lines, which were maintained under 1% CO₂. Exponentially growing cultures were used for growth assays. Once the pre-cultures reached 2–4 × 10⁶ cell mL⁻¹, cells were harvested by centrifugation at 3000 xg for 10 min. Then the cells were resuspended at a concentration of 3 × 10⁵ cell mL⁻¹ in 10 mL of fresh media. To ensure optimal gas exchange, cultures were grown in 6-well plates swirling twice a day and incubated in growth chambers with water-saturated 0.04% CO₂ (ambient air, LC) or 1% CO₂ (HC) in air. Cell density was

monitored daily by counting cells using a digital haemocytometer (Countess II FL, ThermoFisher Scientific). All experiments were performed in biological triplicates.

E. coli strain DH5 α was used for standard cloning procedures and *E. coli* strain C41 (DE3) was used for Shell1 and Shell4 recombinant protein expression. All *E. coli* growth was performed in Luria Broth containing relevant antibiotics.

METHOD DETAILS

Episome Assemblies using Golden Gate Cloning

Level 0, 1, and 2 (L0, L1, and L2) plasmids were assembled by Golden Gate (GG) cloning⁷² using the custom parts from the diatom MoClo framework.²⁸ Using genome version ASM14940v2, target genes without stop codons were synthesized by Twist Bioscience (Table S4). The regulatory elements, fucoxanthin chlorophyll *a/c*-binding protein (FCP) promoter and terminator, and the fluorescent protein (FP) tags (mEGFP and mScarlet-I) harboring L0 plasmids were used to build L1 plasmids together with the gene-of-interest (GOI) for FP tagging. Subsequently, L1 plasmids were assembled into L2 plasmids (episomes) in the following order: episomal maintenance elements (position 1, P1),⁷³ nourseothricin (NAT) resistance cassette (P2), FP tagged GOI (P3), and for dual-FP tagging (P4). For the CRISPR-Cas9 knock-out episomes, Cas9 occupied P3 and two sgRNAs for each target both under the U6 promoter occupied P4 and P5. P6 was used for simultaneous knock-out and *rbcS*-mEGFP tagging. Each GG assembly was performed in 20 μ L containing 40 fmols of each component with 10x ligase buffer (NEB), 10 units T4 DNA ligase (NEB), and 10 units restriction enzyme (BsaI for L1 assembly or Bpil for L0/L2 assembly, Thermo Fisher Scientific). The reaction was incubated in a thermocycler by switching between 37°C and 16°C for 5 min intervals in a total of 20–30 cycles, followed by 37°C for 5 min and terminated after incubating at 65°C for 20 min. 3 μ L of the reaction were transformed into 50 μ L chemically competent DH5 α *E. coli* cells.

Genome editing

To generate the CRISPR-Cas9 knock-out mutant lines, two single guide RNA (sgRNA) sequences were designed to target each gene using CRISPOR⁷⁴ (<http://crispor.gi.ucsc.edu/crispor.py>) and Cas-Designer⁷⁵ (<http://www.rgenome.net/cas-designer/>). The nucleotides selected to generate the *shell1/2* mutant were 5'-CCCTGCCATTGAATCTGTGC-3' (267–286 *shell1*, 264–287 *shell2*) and 5'-TGGCAACAGAGTCAACGGTG-3' (594–613 *shell1*, 591–610 *shell2*); for the *shell3* mutant were 5'-GTCCGAGACCGAGC GAACGT-3' (109–128) and 5'-GCCGAAGCCTTGATGAAATC-3' (197–216); and for the *shell4* mutant were 5'-GAGGGACAATCACGT GAGCAC-3' (148–168) and 5'-GCCCGACTGGACTCCCATGA-3' (261–280).

T. pseudonana Transformation via Bacterial Conjugation

Episomes were delivered to *T. pseudonana* via bacterial conjugation according to Karas et al.⁷³ with minor modifications. Episome plasmids were transformed into *E. coli* (TransforMax EPI300) harboring the pTA_Mob⁵⁹ mobility plasmid (gift from R. Lale) via electroporation (Bio-Rad). Transformed cells were spread onto LB agar plates containing both gentamycin (10 μ g mL⁻¹) and kanamycin (25 μ g mL⁻¹) for selection overnight at 37°C. Colonies were inoculated for subsequent conjugation. Cultures (150 mL) grown at 37°C to OD₆₀₀ of 0.3–0.4 were harvested by centrifugation (3,000 \times g, 5 min) and resuspended in 800 μ L of SOC media. Liquid grown *T. pseudonana* WT culture was harvested by centrifugation (3,000 \times g, 5 min) and resuspended at a concentration of 2 \times 10⁸ cells mL⁻¹ in 1/2ASW-F/2. Equal volume (200 μ L) of *E. coli* and *T. pseudonana* WT cells were gently mixed by pipetting. Next the mixture of cells was plated on 1/2ASW-F/2, 5% LB, 1% agar plates and incubated in the dark for 90 min at 30°C. The plates were transferred to 20°C with continuous illumination (\sim 50 μ mol photons m⁻² s⁻¹) and grown overnight. Next day, 500 μ L of 1/2ASW-F/2 medium was added to the plate for scraping and resuspending the cells. Up to 200 μ L of resuspended cells were spread onto 1% (w/v) 1/2ASW-F/2 agar plates with 100 μ g mL⁻¹ nourseothricin for selection. Colonies appeared after 6–14 days.

Flow Cytometry

mEGFP and mScarlet-I expression was analyzed by flow cytometry using either CytoFLEX LX355 or 375 (Beckman Coulter) analyzers. Forward scattered (FSC) and side scattered photons by the 488 nm laser were used to distinguish diatoms from cell culture debris. FSC-height versus FSC-area signal was used to separate single events from sample aggregates. Chlorophyll autofluorescence excited by a 561 nm laser and emitted photons detected with a 675/25 filter was used to ensure all the diatom cells were fully intact. mEGFP fluorescence excited by a 488 nm laser was detected by an avalanche photodiode detector with a 525/40 bandpass filter. All the data analysis was done using CytExpert software (Beckman Coulter).

Confocal Fluorescent Microscopy

Fluorescence imaging was performed using a Zeiss LSM880 confocal microscope with a 63x objective, 1.4 numerical aperture (NA) Plan-Apo oil-immersion lens (Carl Zeiss). All imaging was done in Airyscan mode except for the *rbcS*-mEGFP line, which was imaged in confocal mode. 20 μ L of cell suspension were pipetted on 8 well μ -Slide chambered coverslips (ibidi) overlaid with 180 μ L of 1.5% F/2-low-melting point agarose (Invitrogen) for imaging. Excitation lasers and emission filters were as follows: mEGFP excitation 488 nm, emission 481–541 nm; mScarlet-I excitation 561 nm, emission 561–633 nm; and chlorophyll excitation 633 nm, emission 642–712 nm. All the microscopic images were processed using Fiji.⁶⁰

Co-Immunoprecipitation and Affinity Purification

For rbcL coIP, 50 mL of WT *T. pseudonana* cells grown in log phase ($2\text{--}3 \times 10^6$ cells mL⁻¹) were harvested by centrifugation (3,000 *xg*, 10 min). The pellets were resuspended in CoIP buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl, 0.1 mM EDTA, 12.5% glycerol) containing 5 mM DTT and protease inhibitor cocktail tablets (PIs, cOmplete EDTA-free, Roche). Cells were lysed by sonication for 3 min (ON 3 sec, OFF 12 sec). The lysates were centrifuged for 20 min (20,817 *xg*, 4°C) to separate the supernatant (soluble lysate) from the pellet. 200 µL of Protein A (Dynabeads Protein A, Invitrogen) beads were washed twice in coIP buffer containing PIs. 32 µg of anti-rbcL antibody in 500 µL coIP buffer containing PIs was added to the washed beads and incubated at 4°C for 2 hours. After incubation beads were washed twice in coIP buffer containing PIs. For blocking, 500 µL of BSA (2 mg mL⁻¹) was added and incubated at 4°C for 1 hour. After incubation beads were washed twice in coIP buffer containing PIs. Subsequently, the soluble lysates were added to protein A beads primed with antibody and incubated at 4°C for 3 hours. After incubation, beads were washed three times with coIP buffer containing PIs and 0.1% digitonin (SigmaAldrich). For elution, 200 µL of 1x SDS loading dye was added and boiled at 95°C for 5 min. The supernatant was collected without any beads and ran on an SDS-PAGE gel for ~1.5 cm. Gels were sliced for further in-gel digestion for LC-MS/MS (see below).

For mEGFP tagged lines affinity purification (AP), 50 mL of GFP tagged *T. pseudonana* lines grown to exponential phase ($2\text{--}3 \times 10^6$ cells mL⁻¹) were harvested by centrifugation (3,000 *xg*, 10 min). The pellets were resuspended in an AP buffer (200 mM D-sorbitol, 50 mM HEPES, 50 mM KOAc, 2 mM Mg(OAc)₂, 1 mM CaCl₂) containing protease inhibitor cocktail tablets (cOmplete EDTA-free, Roche), 2% digitonin (SigmaAldrich), 1 mM PMSF, 0.5 mM NaF and 0.15 mM Na₃VO₄. Cells were lysed by sonication for 30 sec (On 3 sec, Off 15 sec) twice. The lysates were centrifuged for 20 min (20,817 *xg*, 4°C) and the supernatant was incubated with mEGFP-Trap Agarose beads (ChromoTek) for 1 hour according to the manufacturer's instructions. Subsequently, beads were washed twice with AP buffer containing 0.1% digitonin and a final wash without digitonin. All steps were performed at 4°C.

Mass Spectrometry

For rbcL coIPMS, samples were in-gel digested with 0.2 µg sequencing-grade, modified porcine trypsin (Promega), following reduction with 1.5 mg mL⁻¹ dithioerythritol and alkylation with 9.5 mg mL⁻¹ iodoacetamide. Digests were incubated overnight at 37°C. Peptides were extracted by washing three times with aqueous 50% (v:v) acetonitrile containing 0.1% (v:v) trifluoroacetic acid, before drying in a vacuum concentrator and reconstituting in aqueous 0.1% (v:v) trifluoroacetic acid. Peptides were loaded onto an mClass nanoflow UPLC system (Waters) equipped with a nanoEase M/Z Symmetry 100 Å C₁₈, 5 µm trap column (180 µm x 20 mm, Waters) and a PepMap, 2 µm, 100 Å, C 18 EasyNano nanocapillary column (75 mm x 500 mm, Thermo). The trap wash solvent was aqueous 0.05% (v:v) trifluoroacetic acid and the trapping flow rate was 15 µL min⁻¹. The trap was washed for 5 min before switching flow to the capillary column. Separation used gradient elution of two solvents: solvent A, aqueous 0.1% (v:v) formic acid; solvent B, acetonitrile containing 0.1% (v:v) formic acid. The flow rate for the capillary column was 330 nL min⁻¹ and the column temperature was 40°C. The linear multi-step gradient profile was: 3–10% B over 5 mins, 10–35% B over 85 mins, 35–99% B over 10 mins and then proceeded to wash with 99% solvent B for 5 min. The column was returned to initial conditions and re-equilibrated for 15 min before subsequent injections. The nanoLC system was interfaced with an Orbitrap Fusion Tribrid mass spectrometer (Thermo) with an EasyNano ionisation source (Thermo). Positive ESI-MS and MS² spectra were acquired using Xcalibur software (version 4.0, Thermo). Instrument source settings were: ion spray voltage, 1900–2100 V; sweep gas, 0 Arb; ion transfer tube temperature; 275°C. MS¹ spectra were acquired in the Orbitrap with: 120,000 resolution, scan range: *m/z* 375–1,500; AGC target, 4e⁵; max fill time, 100 ms. Data dependent acquisition was performed in top speed mode using a 1 s cycle, selecting the most intense precursors with charge states >1. Easy-IC was used for internal calibration. Dynamic exclusion was performed for 50 s post precursor selection and a minimum threshold for fragmentation was set at 5e³. MS² spectra were acquired in the linear ion trap with: scan rate, turbo; quadrupole isolation, 1.6 *m/z*; activation type, HCD; activation energy: 32%; AGC target, 5e³; first mass, 110 *m/z*; max fill time, 100 ms. Acquisitions were arranged by Xcalibur to inject ions for all available parallelizable time. Tandem mass spectra peak lists were extracted from Thermo.raw files to .mgf format using MSConvert (ProteoWizard 3.0). Mascot Daemon (version 2.6.0, Matrix Science) was used to submit searches to a locally-running copy of the Mascot program (Matrix Science Ltd., version 2.7.0). Peak lists were searched against the *Thalassiosira pseudonana* subsets of UniProt and NCBI with common proteomic contaminants appended. Search criteria specified: Enzyme, trypsin; Max missed cleavages, 2; Fixed modifications, Carbamidomethyl (C); Variable modifications, Oxidation (M); Peptide tolerance, 3 ppm; MS/MS tolerance, 0.5 Da; Instrument, ESI-TRAP. Peptide identifications were collated and filtered using Scaffold (5.2.0, Proteome Software Inc). Peptide identifications were accepted if they could be established at greater than 51.0% probability to achieve an FDR less than 1.0% by the Percolator posterior error probability calculation. Protein identifications were accepted if they could be established at greater than 5.0% probability to achieve an FDR less than 1.0% and contained at least 2 identified peptides.

For mEGFP tagged lines APMS, samples were on-bead digested using Chromotek's recommended procedure for NanoTraps: protein was digested overnight at 37°C with 25 µL 50 mM Tris-HCl pH 7.5, 2 M urea, 1 mM DTT, 5 µg mL⁻¹ Sequencing Grade Modified Trypsin (Promega). Peptides were eluted with 50 mM Tris-HCl pH 7.5, 2 M urea, 5 mM iodoacetamide before loading onto EvoTip Pure tips for desalting and as a disposable trap column for nanoUPLC using an EvoSep One system. A pre-set EvoSep 60 SPD gradient was used with a 8 cm EvoSep C₁₈ Performance column (8 cm x 150 µm x 1.5 µm). The nanoUPLC system was interfaced to a timsTOF HT mass spectrometer (Bruker) with a CaptiveSpray ionisation source (Source). Positive PASEF-DDA, ESI-MS and MS² spectra were acquired using Compass HyStar software (version 6.2, Thermo). Instrument source settings were: capillary voltage,

1,500 V; dry gas, 3 l min⁻¹; dry temperature; 180°C. Spectra were acquired between *m/z* 100–1,700. The following TIMS settings were applied as: 1/K0 0.6–1.60 V.s cm⁻²; Ramp time, 100 ms; Ramp rate 9.42 Hz. Data dependent acquisition was performed with 10 PASEF ramps and a total cycle time of 1.17 s. An intensity threshold of 2,500 and a target intensity of 20,000 were set with active exclusion applied for 0.4 min post precursor selection. Collision energy was interpolated between 20 eV at 0.5 V.s cm⁻² to 59 eV at 1.6 V.s cm⁻². Pick picking, database searching, significance thresholding and peak area integration was performed using FragPipe (version 19.1). Data were searched against UniProt reference proteome UP000001449, appended with common contaminants and concatenated with reversed sequences for false discovery calculation. Search criteria specified: Enzyme, trypsin; Max missed cleavages, 2; Fixed modifications, Carbamidomethyl (C); Variable modifications, Oxidation (M), Acetylation (Protein N-term); Peptide tolerance, 10 ppm; MS/MS tolerance, 10 ppm; Instrument, IM-MS. Peptide identifications were filtered using Percolator and ProteinProphet to 1% PSM FDR, protein probabilities >99%, best peptide probability >99% and a minimum of two unique peptides. Peak area quantification was extracted using IonQuant with match between run applied. Feature detection tolerances were set to: MS1 mass <10 ppm; RT < 0.4 min; and IM (1/k0) <0.05.

For whole-cell MS samples were ran in triplicate. 50 mL of WT cells were grown in ambient CO₂ conditions and harvested at the log phase (2–3 × 10⁶ cells mL⁻¹) by centrifugation (3,000 *xg*, 10 min). Shell4 mutant lines were grown in 1% CO₂. Cells were lysed by sonication (3 min: ON 3 s, OFF 12 s) in ColP buffer containing 5 mM DTT and PIs followed by centrifugation for 20 min (20,817 *xg*, 4°C). Supernatant was run on an SDS-PAGE gel for ~1.5 cm. In-gel digestion was performed with the addition of 0.2 mg sequencing-grade, modified porcine trypsin (Promega), following reduction with dithioerythritol and alkylation with iodoacetamide. Digests were incubated overnight at 37°C. Peptides were extracted by washing three times with aqueous 50% (v:v) acetonitrile containing 0.1% (v:v) trifluoroacetic acid, before drying in a vacuum concentrator and reconstituting in aqueous 0.1% (v:v) trifluoroacetic acid.

S-TrapTM micro spin column (PROFIT) digestions were performed using the manufacture's protocol. Briefly, protein in 5% SDS (w:w) lysis buffer was reduced with tris(2-carboxyethyl)phosphine and alkylated with methyl methanethiosulfonate before acidification with phosphoric acid and dilution into 100 mM triethylammonium bicarbonate (TEAB) in 90% (v:v) methanol. Protein was washed on s-Trap with the same buffer five times before digestion for 2 h at 47°C with 2 mg Promega Trypsin/Lys-C mixed protease in 50 mM TEAB.

Peptides were loaded onto EvoTip Pure tips for nanoUPLC using an EvoSep One system. A pre-set 30SPD gradient was used with a 15 cm EvoSep C₁₈ Performance column (15 cm × 150 mm × 1.5 mm).

The nanoUPLC system was interfaced to a timsTOF HT mass spectrometer (Bruker) with a CaptiveSpray ionisation source. Positive PASEF-DIA, nanoESI-MS and MS² spectra were acquired using Compass HyStar software (version 6.2, Bruker). Instrument source settings were: capillary voltage, 1,500 V; dry gas, 3 l/min; dry temperature; 180°C. Spectra were acquired between *m/z* 100–1,700. DIA windows were set to 25 Th width between *m/z* 400–1201 and a TIMS range of 1/K0 0.6–1.60 V.s/cm². Collision energy was interpolated between 20 eV at 0.65 V.s/cm² to 59 eV at 1.6 V.s/cm².

LC-MS data, in Bruker.d format, was processed using DIA-NN (1.8.2.27) software and searched against an *in silico* predicted spectral library, derived from the *Thalassiosira pseudonana* subset of UniProt (11934 protein sequences). Search criteria were set to maintain a false discovery rate (FDR) of 1% with heuristic protein inference. High-precision quant-UMS was used for extraction of quantitative values within DIA-NN. Peptide-centric output in.tsv format, was pivoted to protein-centric summaries using KNIME 5.1.2 and data filtered to require protein *q*-values < 0.01 and a minimum of two peptides per accepted protein.

Protein-Protein Interaction Network analysis

Protein abundances quantified using MS2 spectral count measurements of fragment ions from all sample triplicates were run through a CompPASS package in R Studio (<https://github.com/dnusinow/cRomppass/blob/master/R/cRomppass.R>) and a control IP inclusive variation of SAINT analysis in Ubuntu using standard parameters.³⁵ The WD and AvgP scores respectively generated were used as measures of interaction strength between bait and prey proteins. Only interactions which fell in both the top 2.2% WD score and 1% AvgP score were filtered as high confidence interactors. Prior to analysis, bait spectral count data was set to zero to minimize data skewing due to the typically high spectral counts and the inability to distinguish between mEGFP tagged bait and untagged native protein.

In vivo Fluorescence Recovery After Photobleaching

FRAP experiments were performed using a Zeiss LSM980 confocal microscope with a 63x objective 1.4 numerical aperture (NA) Plan-Apo oil-immersion lens (Carl Zeiss). Samples were prepared as in confocal microscopy and overlaid with 200 μL of ibidi anti-evaporation oil. 20 pre-bleach images were taken prior to bleaching (60% 488 nm intensity, 1 cycle). All the images were processed by Fiji.⁶⁰ The Image Stabilizer plugin (4 pyramid levels, 0.99 template update coefficient) output of the brightfield images were used to stabilize the fluorescence images. The mean gray values were measured for the bleached, unbleached and background regions of interest (ROIs). Background values were subtracted from bleached and unbleached values before photobleach normalization using the unbleached reference was completed. The average pre-bleach intensity was used for full-scale normalization.

Data visualization

Network visualization of the interaction network was done using Cytoscape (version 3.10.0) (<https://cytoscape.org/>). The phylogenetic tree was done in Geneious Prime (2023.2.1). *Thalassiosira antarctica* Rubisco model (PDB: 5M22) was used to visualize the

peptide regions utilized for raising the anti-rbcL antibody using ChimeraX.⁶⁷ Adobe Illustrator (2023), Excel (Microsoft) and Prism 10 (GraphPad) were used to generate the figures.

Structure Prediction

AlphaFold structure predictions for Shell 1-6 and DPC3 without signal peptide and chloroplast transit peptides were done using ColabFold v1.5.2-patch: AlphaFold2 using MMseqs2 (<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>).⁶¹ To compare Shell1 and Shell4, the structures of both proteins were predicted using AlphaFold3⁶² and fitted using MatchMaker in ChimeraX.⁶⁷ A sequence alignment of Shell protein sequences was obtained with Clustal Omega.⁶⁹

Transmission Electron Microscopy

50 mL of cell culture grown in log phase were harvested by centrifugation (3,000 xg , 5 min). Cells were fixed in 0.1 M cacodylate buffer (SigmaAldrich), pH 7.4, containing 2.5% glutaraldehyde (SigmaAldrich), 2% formaldehyde (Polysciences) for 1 hour at room temperature. They were then rinsed twice in phosphate-buffered saline (PBS, pH 7.4) for 15 min by removing the liquid at the top of the tube and leaving the cells undisturbed at the bottom of the tube. The cells were secondary fixed with Osmium Tetroxide (1%, in buffer pH 7.2, 0.1M) for 1 hour. After rinsing twice with buffer, again removing the liquid above the cells whilst leaving the cells undisturbed, the cells were dehydrated through an ethanol series of 30, 50, 70, 90 and 100%, with each rinse for 15 min. To ensure thorough dehydration the 100% ethanol step was repeated 3 times. The ethanol was then replaced by Agar low viscosity resin by placing it in increasing concentrations of resin (30% resin: 70 % ethanol. 50:50, 70:30 each change was left for at least 12 hours) until it was in 100% resin. To ensure complete resin infiltration, the 100% resin step was repeated 3 times leaving it overnight between changes. The Eppendorf tubes were then placed in an embedding oven and the resin polymerized at 60°C overnight. The resulting blocks were sectioned with a Leica Ultracut E ultra microtome using a diatome diamond knife. The sections were then stained using a saturated solution of Uranyl acetate (for 15 min) and Reynold's Lead citrate (15 min). The sections were imaged using a JEOL 1400 TEM.

Blast search

Full amino acid sequences of Shell 1-6 were used for Blastp against the NCBI database with default settings: 'Standard databases', 'Non-redundant protein sequences (nr)', 'blastp (protein-protein BLAST)' 'Max target sequences=100' 'Word size =5' 'Max matches in a query range = 0' 'Matrix = BLOSUM62' 'Gap Costs= Existence: 11 Extension: 1' 'Compositional adjustments = Conditional compositional score matrix adjustment' and 'Expect threshold =1'. Hits were sorted by highest to lowest alignment length, and a cut-off length of 100 amino acids was employed.

Phylogeny

The phylogenetic tree was built using Geneious Prime (2023.2.1). 136 proteins were used for MAFFT alignment using default settings (Algorithm = 'Auto', Scoring matrix = 'BLOSUM62', Gap open penalty = '1.53', Offset value = '0.123'). Phylogenetic tree analysis was performed using RAXML with the following settings: Protein Model = 'GAMMA BLOSUM62', Algorithm = 'Rapid Bootstrapping', Number of starting trees or bootstrap replicates = '1000', Parsimony random seed = '1'. A consensus tree was generated by selecting Consensus Tree Builder, Create Consensus Tree, Support Threshold % = 0, Topology Threshold % = 0, Burn-in % = 0, Save tree(s) separately. *C. reinhardtii* BST1-3 sequences (Cre16.g662600, BST1; Cre16.g663400, BST2; Cre16.g663450, BST3) were added to root the tree.

Recombinant Shell1 protein purification

A modified pOPT expression plasmid⁷⁶ containing *shell1* with N-terminal His6-MBP tag and C-terminal GST tag was transformed into *E. coli* expression strain C41 (DE3). Cells were grown in 2 L of liquid Luria Broth medium with 100 $\mu g mL^{-1}$ carbenicillin until OD_{600} reached 0.6 when the culture was induced with 1 mM IPTG for 6 h. Harvested cells were resuspended in a lysis buffer (50 mM Tris-HCl, pH 8.0, 1x cComplete EDTA-free protease inhibitor with 500 mM NaCl and 2 mM PMSF). The cells were lysed by cell disruption (30 kPSI) and the lysate cleared by 10 min of centrifugation at 7,000 xg and 30 min at 20,000 xg before being filtered with a 0.2 μm filter. The supernatant was applied onto a 5 ml HisTrap Fast Flow nickel affinity column (Cytiva) equilibrated with buffer A (50 mM Tris-HCl pH 8.0, 500 mM NaCl) followed by 4CV wash with lysis buffer. The bound protein was then eluted using a linear gradient of imidazole up to 500 mM concentration. The elution was then treated with TEV protease and 3C protease while dialysed overnight at 4°C into 50 mM Tris-HCl, pH 8.0, 500 mM NaCl buffer. The sample was then applied to both a HisTrap Fast Flow nickel column and a GSTrap column to remove the tags. The Shell protein remained in the flowthrough. The protein was then concentrated using an Amicon Ultra concentrator 10,000 MW, applied onto a Superdex 200 pg 16/600 Column to further ensure the purity of the sample, and concentrated again with an Amicon Ultra Concentrator if needed.

Cryo-Electron Microscopy of recombinantly expressed Shell1

QUANTIFOIL® R 1.2/1.3 grids were glow discharged with a PELCO easiGlow system using a pressure of 0.26 mBar, for 60 seconds at a 20 mA current. Shell1 protein sample in 50 mM Tris-HCl (pH 8.0), 50 mM NaCl at 2.3 mg mL^{-1} was applied onto grids, blotted for 8 s using force 10 with a FEI Mark IV Vitrobot (ThermoFisher Scientific) in a chamber at 4°C and 95% relative humidity. Grids were then plunged into liquid ethane.

Data was collected on a 200 kV Glacios cryo-electron microscope with a Falcon IV direct electron detector at the University of York. Automated data collection was performed using EPU in AFIS mode (ThermoFisher Scientific). Micrographs were collected at a nominal magnification of 120,000x and total electron fluence of $50 \text{ e}^-/\text{\AA}^2$ with pixel size of 1.2 \AA^2 . Total exposure time was 8 s. A $100 \text{ }\mu\text{m}$ objective aperture was inserted and the C2 aperture was $50 \text{ }\mu\text{m}$. Defocus values were: -2.0, -1.6, -1.4, -1.2, -1.0.

Cellular Cryo-Electron Tomography

T. pseudonana cultures were grown as described above, with the mutant *shell4* KO::rbcS-mEGFP being supplemented with 5 mM Na_2CO_3 in the F/2 medium. Cells were sedimented at 800 xg for 5 min prior to vitrification, $4 \text{ }\mu\text{L}$ of cell suspension was applied to 200-mesh R2/1 carbon-film covered copper grids (Quantifoil Micro Tools), and grids were blotted and plunge-frozen using a Vitrobot Mark IV (Thermo Fisher Scientific). EM grids were clipped into Autogrid supports (Thermo Fisher Scientific) and loaded into an Aquilos 2 FIB-SEM instrument (Thermo Fisher Scientific), where they were thinned with a Gallium ion beam as previously described.⁷⁷ The resulting EM grids with thin lamellae were transferred to a transmission electron microscope for tomographic imaging. Datasets were acquired on three different microscopes: a 300 kV Titan Krios G4i microscope (Thermo Fisher Scientific), equipped with a Selectris X post-column energy filter (Thermo Fisher Scientific) and a Falcon 4 direct electron detector (Thermo Fisher Scientific) (“M1”); a 300 kV Titan Krios G3i microscope (Thermo Fisher Scientific) equipped with a BioQuantum post-column energy filter (Gatan) and a K3 direct electron detector (Gatan) (“M2”); and Titan G4 equipped with a monochromator, a Selectris X energy filter, and a Falcon 4 camera (“M3”). Tilt-series were obtained using SerialEM 3.8 software⁶⁸ or Tomo5 (Thermo Fisher Scientific). In all cases, tilt-series were acquired using a dose-symmetric tilt scheme,⁷⁸ with 2° steps totaling 60 tilts per series. Each image was recorded in counting mode with ten frames per second. For M1 and M3, data was acquired in EER mode. The target defocus of individual tilt-series ranged from -2 to $-5 \text{ }\mu\text{m}$. Total dose per tilt series was approximately $120 \text{ e}^-/\text{\AA}^2$. Image pixel sizes for microscopes M1, M2, and M3 were 2.42 \AA , $2.143/2.685 \text{ \AA}$, and 2.38 \AA , respectively.

Tomograms were subsequently reconstructed. First, TOMOMAN MATLAB scripts (version 0.6.9 <https://github.com/williamnwan/TOMOMAN/>; <https://doi.org/10.5281/zenodo.4110737>)⁷⁹ were used to preprocess the tomographic tilt series data. Raw frames were then aligned using MotionCor2 (version 1.5.0)⁶³ before dose-weighting the tilt-series,⁸⁰ followed by manual removal of bad tilts. The resulting tilt-series (binned 4 times) were aligned in IMOD (version 4.11)⁶⁵ using patch tracking and were reconstructed by weighted back projection. Cryo-CARE (version 0.2.1)⁶⁶ was applied on reconstructed tomogram pairs from odd and even raw frames to enhance contrast and remove noise. Snapshots of denoised tomograms were captured using the IMOD 3dmod viewer. Denoised tomograms were used as input for automatic segmentation using MemBrain.⁶⁴ The resulting segmentations were manually curated in Amira (version 2021.2; Thermo Fisher Scientific). Rubisco enzymes were template matched using PyTOM (version 0.0.6)⁸¹ on bin4 tomograms. The resulting segmentations and particle coordinates were visualized in ChimeraX.⁶⁷

QUANTIFICATION AND STATISTICAL ANALYSIS

For FRAP experiments, n represents the number of individual replicate measurements on distinct pyrenoids. The mean, S.E.M. and S.D. of each timepoint were calculated from the indicated number of replicates (n) in Excel (Microsoft). The mean S.D. for the growth experiments were conducted using Excel (Microsoft). The whole cell mass spectrometry plots with the mean S.D., were performed using GraphPad PRISM version 10 software (GraphPad).

Supplemental figures

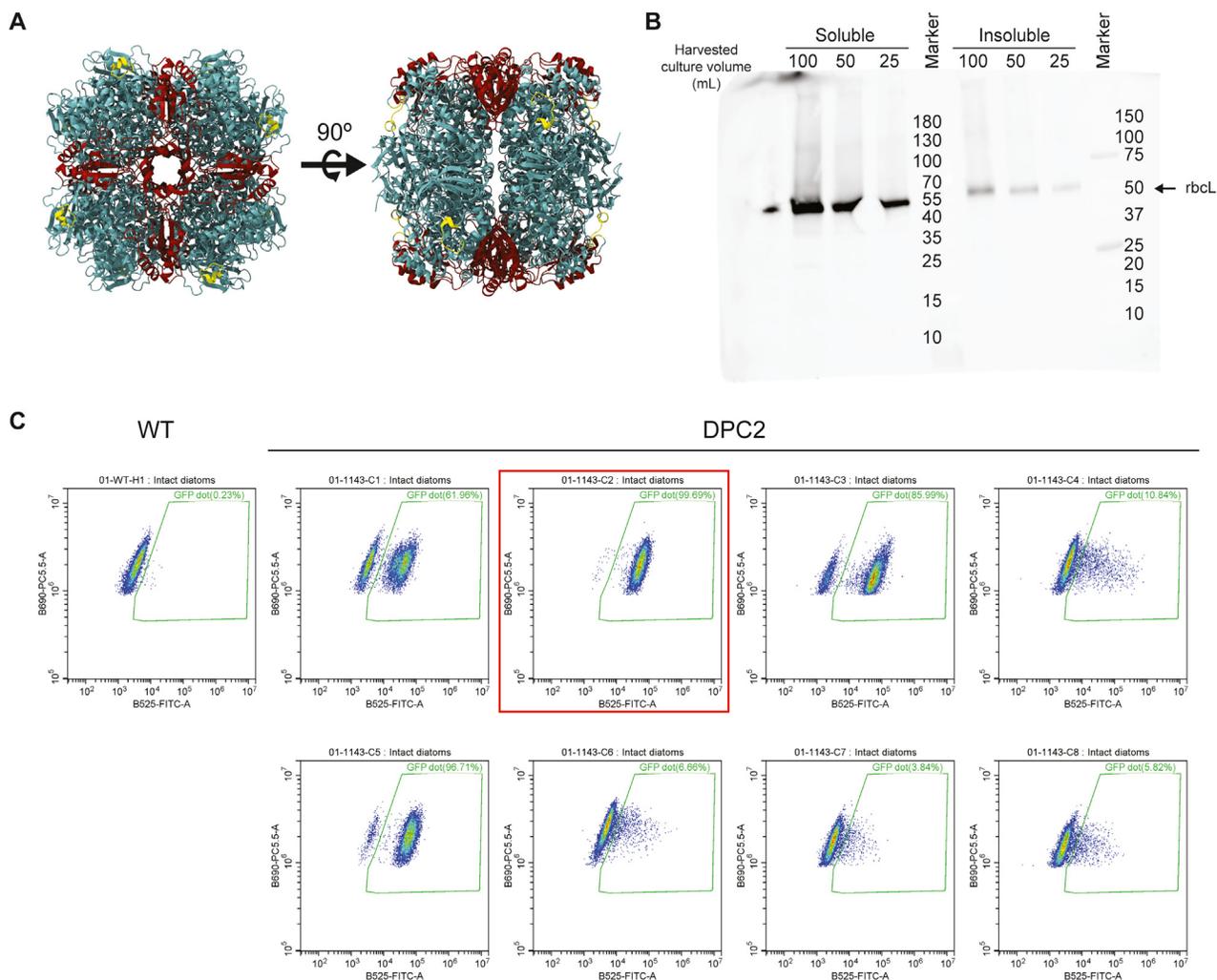


Figure S1. RbcL antibody and flow cytometry screening to identify fluorophore-fusion lines, related to Figures 1 and 2

(A) An antibody was raised against a surface-exposed region (yellow) of the rbcL (cyan). rbcS is shown in red.

(B) Immunoblot using the rbcL antibody against the soluble and insoluble cell fractions of *T. pseudonana*.

(C) Example screening data for diatom pyrenoid component 2 (DPC2). Eight independent lines expressing DPC2-mEGFP from an episome were screened via flow cytometry. The green box outlines the gate used to quantify the percentage of cells that show mEGFP fluorescence with the percentage indicated above. The red box denotes the cell line used for subsequent imaging and APMS.

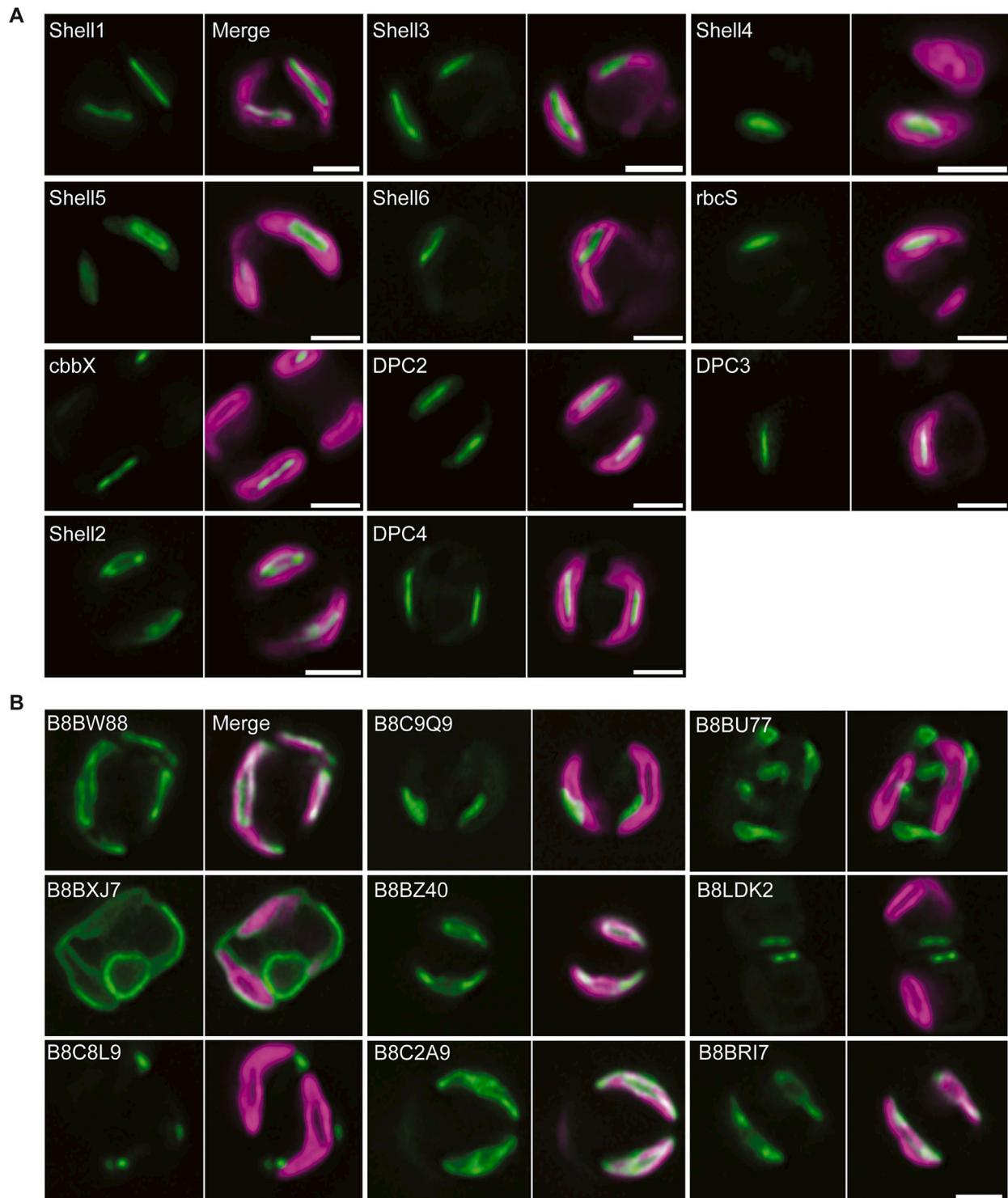


Figure S2. Images of additional mEGFP-tagged lines, related to Figures 2 and 3

(A) Images of additional pyrenoid-localized proteins and additional independently generated lines to those used in Figures 2 and 3.

(B) Images of mEGFP fusion proteins that did not localize to the pyrenoid. Predicted functions/annotations: B8BW88, aspartate/glutamate/uridylylase kinase domain-containing protein; B8BXJ7, uncharacterized protein; B8C8L9, Gfo/dh/MocA-like oxidoreductase N-terminal domain-containing protein; B8C9Q9, oxoglutarate/malate translocator; B8BZ40, phosphoribulokinase; B8C2A9, S-malonyltransferase; B8BU77, nucleoside-diphosphate kinase; B8LDK2, glycoside hydrolase family 5 domain-containing protein; B8BRI7, 2-isopropylmalate synthase.

Green: mEGFP fusion protein; magenta: chlorophyll. Scale bars: 2 μ m.

A

	Signal peptide	Transit peptide	
Consensus	MKFAAALXLAAXX-GASAFAPSANRXTSXLX-----X-XST-----XVRX--XVG--X---XGGX--PXKXGGTXR-XLAE		61
Shell1	MKF.AAL.LAA..-GA.AF.AP.A..T.L.....ST-----VR...VG-----GG.-P.YK.GGT.R-.LAE		67
Shell2	MKF.AA..L.A.-GASAF.AP.ANR..S.L.....ST-----VR...VG-----GG.-P.YK.GGT.R-.LAE		66
Shell3	MK..AAL.LAA..-GASAF..PS.....S.....ST-----VR...VG.....GG.....G.R-.L.E		71
Shell4	MKF.A.L.L.....-A.F.....P.....T.....		36
Shell5	MKFA..L.L.....-G.SAF.AP.....S.....T.....A		78
Shell6	MK.AAAL.L.A.....AF..S.NR.T.L.....GG.....		122
β-sheet			
Consensus	IWD---NSSPVIQGG-SXRTWSFXNPAIESVQVLLKTEGRPLDADVELWQGPNTPHKMRVYVEDGALRXFNAVIGTPRG---PNTVXRNIQLEFFLDAVVRPDRDDGL-----		166
Shell1	IWD---NSSPVIQGG-S.RTWSF.NPAIESVQVLLKTEGRPLDADVELWQGPNTPHKMRVYVEDGALR.FNAVIGTPRG---PNTV.IRNIQLEFFLDAVVRPDRDDGL-----		172
Shell2	IWD---NSSPVIQGG-S.RTWSF.NPAIESVQVLLKTEGRPLDADVELWQGPNTPHKMRVYVEDGALR.FNAVIGTPRG---PNTV.IRNIQLEFFLDAVVRPDRDDGL-----		171
Shell3	IW---NS.PVIVQGG-S.RTWSF.NP.I..VQVLLKTEGRPLDADVELWQGPNTPHKMRVY.EDGALR.FNAVIGTPRG---PNTV..RNIQLEFFLDAVVRPDR.DGL-----		176
Shell4	.WD.....V..G.-S...W.F.P..E.VQV...GRP...ELW.GPD.TP.K...Y.EDG..R...IGT.....I.N.G...FPL.A.....		144
Shell5N.....QG.....T.....VQ.....GRPL.A.VELW.GP...H.M...GA...F.A.....P.....I.N.G...FP.V.....		196
Shell6	.W.....VQGG-S..T.....E...V.L.T.GRPLD...ELW.GP.NTP.K.VY.EDG..R..NA.....P.....NT...RN.G..EFP..A.VR.....		249
β-sheet			
Consensus	AAXI-ASVATRSETIQGGALRTYPNXXVDSVXIIILKTD---GRPLNARIELLQGPNNKQVXELYTEDGLDRPFFAIETPGSGNVVVRXNTAPLEFPLYASV-DAYXVG-----		272
Shell1	AA.I-ASVATRSETIQGGALRTYPN..VDSV.IILKTD---GRPLNARIELLQGPNNKQV.ELYTEDGLDRPFFAI.ETPGSGNVVVR.NTAP.EFPLYASV-DAY.VG-----		278
Shell2	AA.I-ASVATRSETIQGGALRTYPN..VDSV.IILKTD---GRPLNARIELLQGPNNKQV.ELYTEDGLDRPFFAI.ETPGSGNVVVR.NTAP.EFPLYASV-DAY.VG-----		277
Shell3	A..I-ASV..RSETIQGGALRTYPN..VDSV..ILKTD---GRPLNARIELLQGPNNKQV.ELYTEDGLDRPFFAI.ETPG.GNVVVR.NTAPLEFPLYASV-DAY..G-----		282
Shell4	A.....A.....I.GGA.....F...VDSV...L.TD---R.LNA..ELL.GPNN.KQ..E..T..G.....TPGSGNVVVR.N.APLEFP..A.....		270
Shell5T.....R..P.....SV.....IELLQGPN...Q..EL...G...P..A..ETPG.G...RV.NT..L.....V...Y.V.....		324
Shell6R.ET.QGGALRT.P...V..V.I.L.L.D---G.P...IEL...KQV.E.Y...G..RFP.AI..TPG..N...V.N..P.E.P...V-----		355
Consensus	GXGDWDXGLMIGRSXXFXGSRSEFRNSPYGSGGRGQAIGGNW		315
Shell1	G.GDW.D.GLMIGR.....		294
Shell2	G..DW.D.GLMIGRS.....		293
Shell3	.G.WND.G..G.....		300
Shell4		248
Shell5	G.....G...G.S..P.GSRESFRNSPYGSGGRGQAIGGNW		367
Shell6	...WN.....G.S..P.....		373

B Protein (DNA) (%)

	Shell1	Shell2	Shell3	Shell4	Shell5	Shell6
Shell1	100 (100)	92 (93)	73 (72)	34 (46)	23 (39)	29 (42)
Shell2		100 (100)	71 (69)	35 (45)	23 (39)	30 (42)
Shell3			100 (100)	33 (41)	25 (37)	31 (43)
Shell4				100 (100)	23 (35)	25 (33)
Shell5					100 (100)	18 (35)
Shell6						100 (100)

High Low

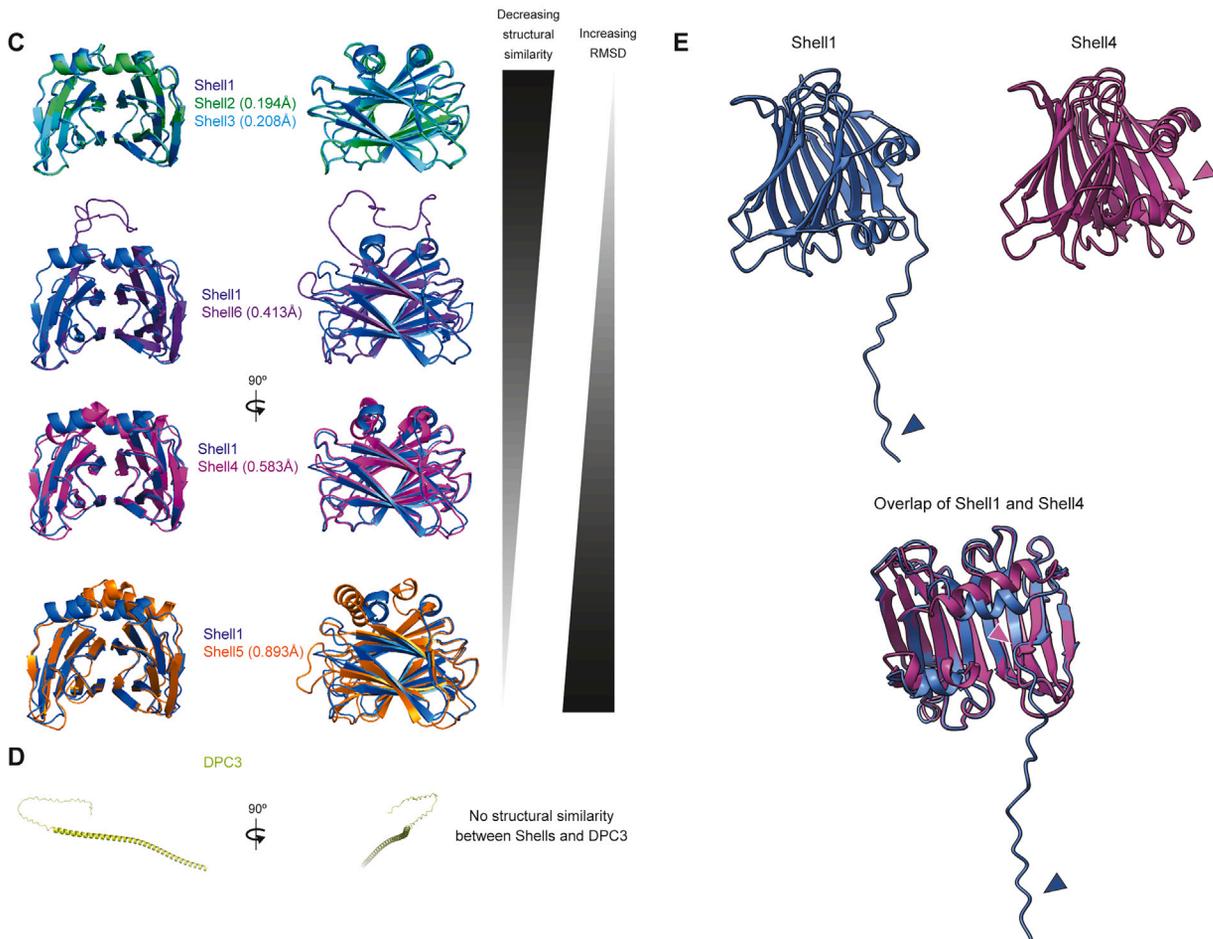


Figure S3. Shell protein amino acid alignment, sequence, and structural similarities, related to [Figure 4](#)

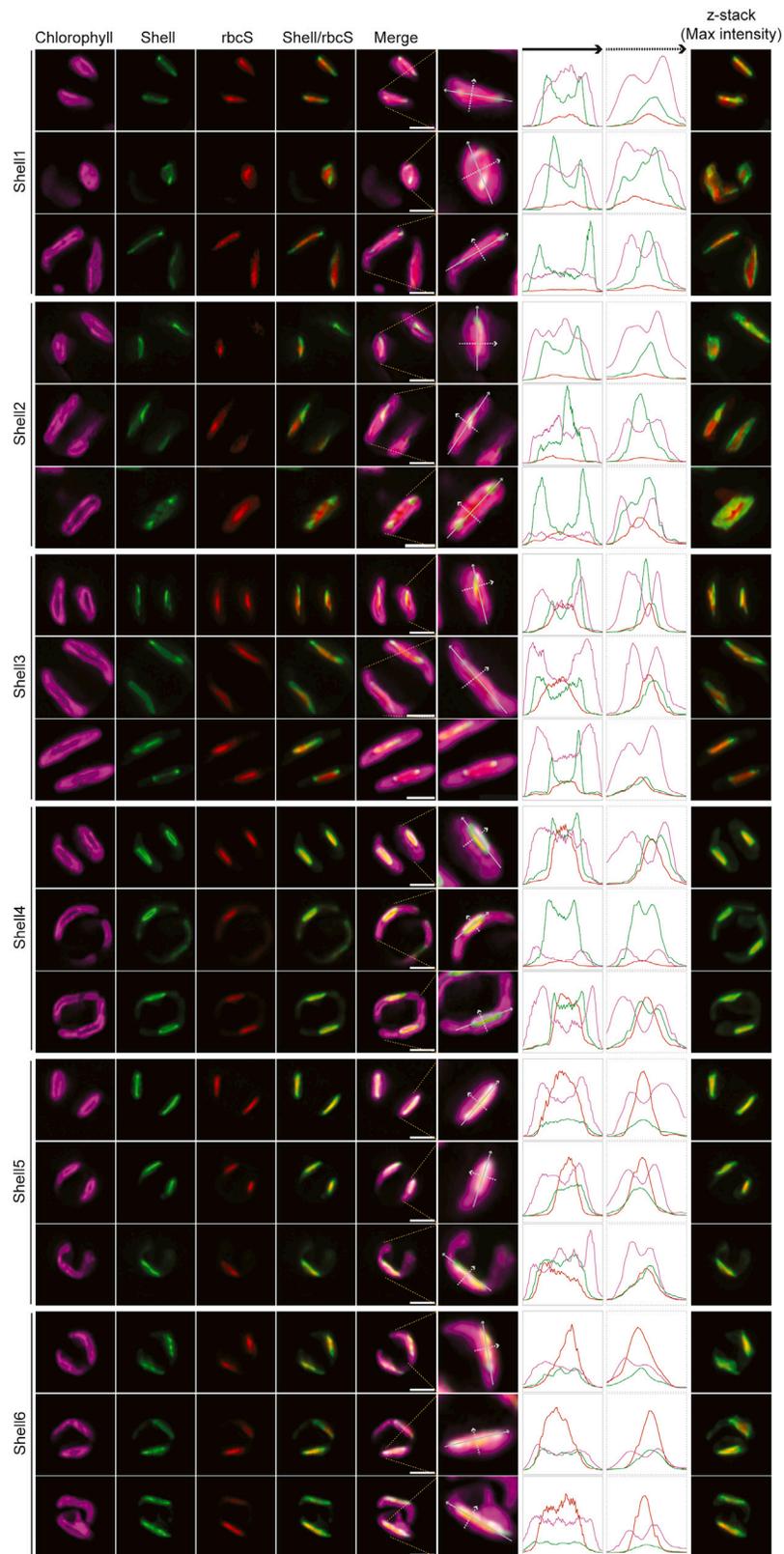
(A) Alignment of full-length Shell amino acid sequences.

(B) Protein and DNA sequence similarities between Shell homologs, including predicted signal and transit peptides.

(C) AlphaFold2 model comparisons of Shell proteins with both N- and C-terminals removed. Root-mean-square deviation (RMSD) values relative to Shell1 are shown.

(D) AlphaFold2 model of DPC3.

(E) AlphaFold3-predicted monomers of Shell1 and Shell4 with C-terminal domains indicated by arrows (predicted N-terminal disordered regions were removed). Overlap of Shell1 and Shell4 models was performed using ChimeraX MatchMaker.



(legend on next page)

Figure S4. Co-localization of Shell proteins with Rubisco, related to [Figure 5](#)

Image columns 1–4: additional images of co-localization of Shell proteins with Rubisco. Scale bars: 2 μm . Image columns 5–8: fluorescence intensity cross-sections of chlorophyll, Shell, and rbcS fluorescence across the pyrenoid. Level of zoom changes between images. Image column 9: z stack max intensity projections.

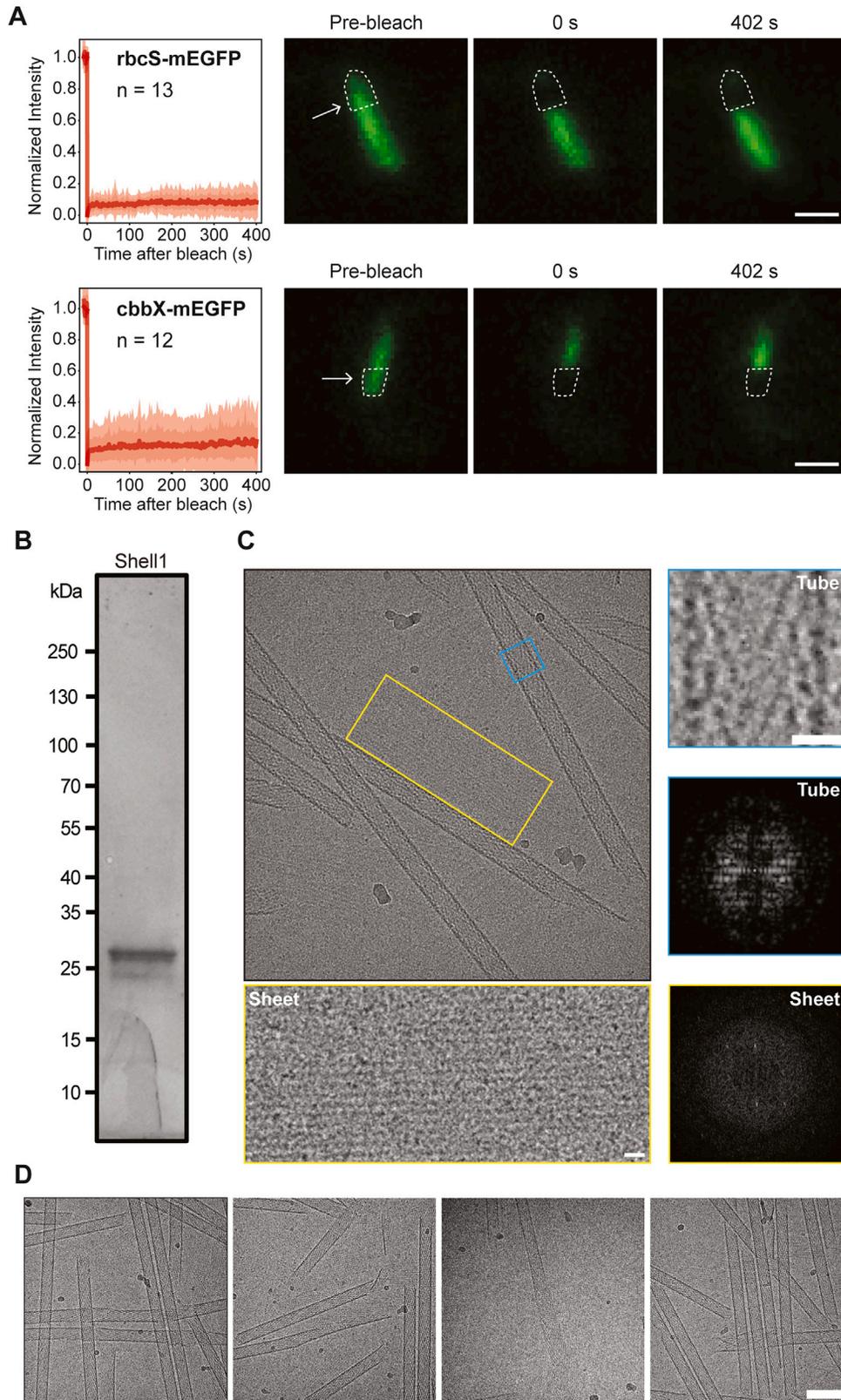


Figure S5. *rbcS* and *cbbX* are not mobile in the *T. pseudonana* pyrenoid, and Shell1 forms tubes and sheets *in vitro*, related to Figure 6

(A) Fluorescence recovery after photobleaching (FRAP) experiments for *rbcS* and *cbbX*. Arrows indicate photobleached regions. Scale bars: 1 μm . Plots show SEM and SD of the mean.

(B) Coomassie-stained SDS-PAGE gel of recombinant Shell1 used for cryo-EM.

(C) Fourier transform of zoomed images of tubes and sheets of self-assembled Shell1. Same micrograph as Figure 6C. Scale bar: 10 nm.

(D) Representative cryo-EM micrographs of Shell1 self-assembly into tubes and sheets. Scale bar: 100 nm.

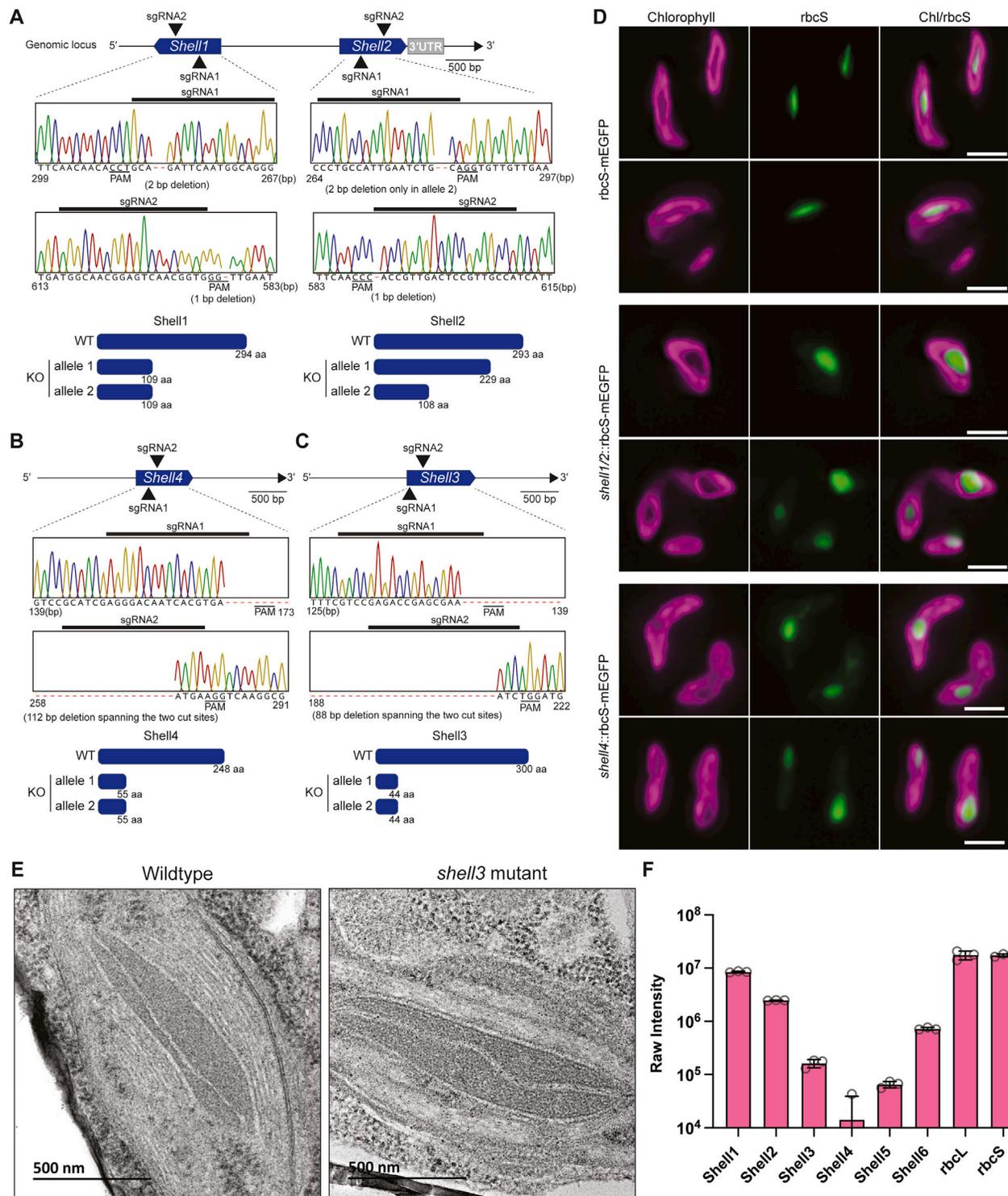


Figure S6. *Shell1/2*, *shell3*, and *shell4* mutant genotyping, pyrenoid morphology, and *shell4* proteomics, related to Figure 6

(A) Biallelic knockout genotyping of *shell1/2*. Due to 93% DNA sequence similarity between *Shell1* and *Shell2*, both genes were knocked out simultaneously by designing two sgRNAs that targeted conserved sequences in *Shell1* and *Shell2*.

(B) Biallelic knockout genotyping of *shell4*.

(C) Biallelic knockout genotyping of *shell3*.

(legend continued on next page)

(D) *Shell1/2* and *shell4* mutants have malformed pyrenoids. Additional images of WT::rbcS-mEGFP, *shell1/2*::rbcS-mEGFP, and *shell4*::rbcS-mEGFP. Green: mEGFP fusion protein; magenta: chlorophyll. Scale bars: 2 μ m.

(E) The *shell3* mutant has WT pyrenoid morphology. TEM images of WT and *shell3* mutant pyrenoids. The WT image is the same as used in [Figure 1](#).

(F) Whole-cell proteomics of the *shell4* mutant. Relative abundance of Shell proteins and Rubisco in the *shell4* mutant. Shell4 was detected at low levels in one of the three replicas. Error bars: SD of the mean.

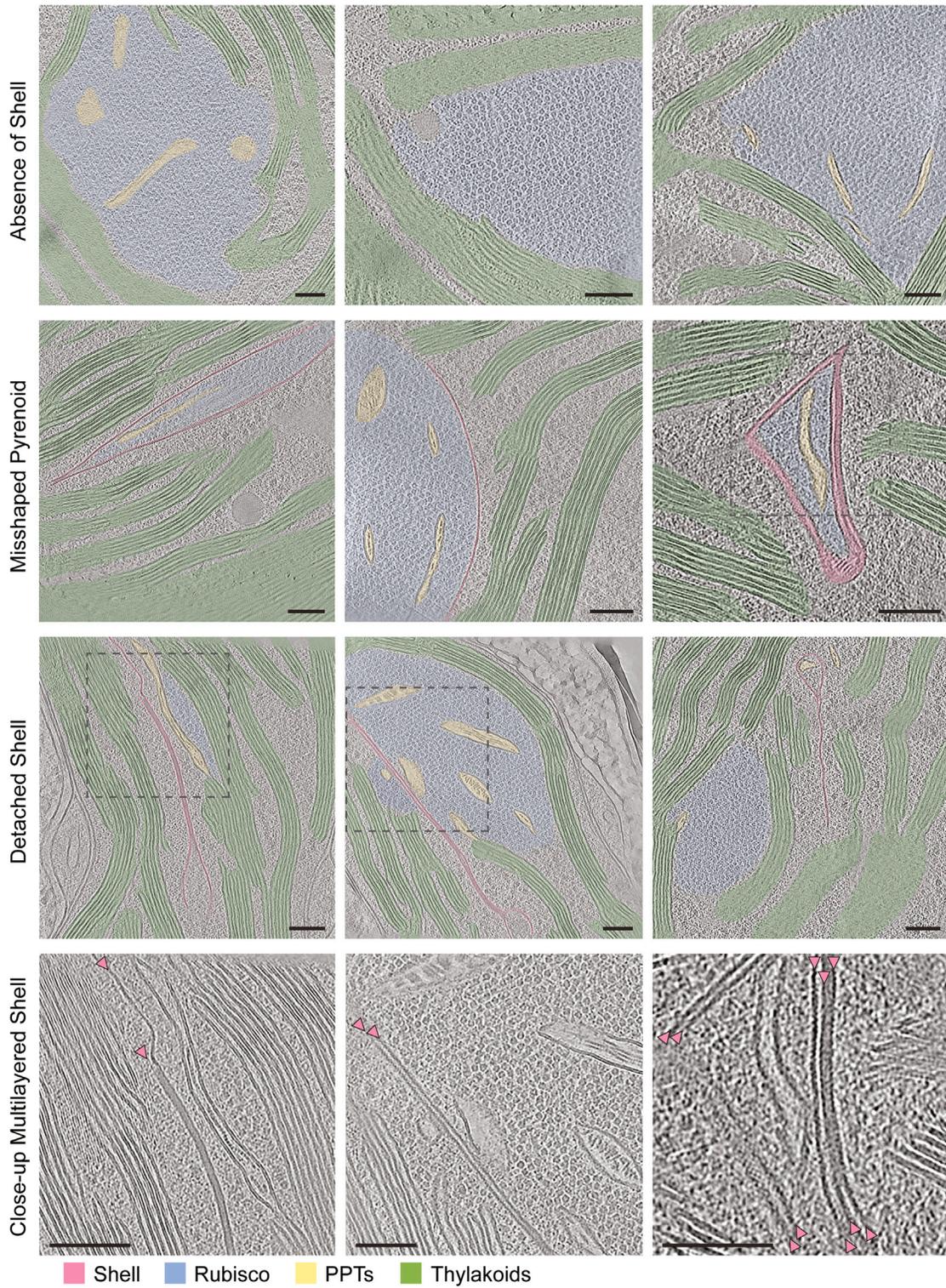


Figure S7. Additional tomograms of the *shell4* mutant pyrenoid morphologies, related to Figure 6

Additional tomogram slices of the different types of misassembled pyrenoids in the *shell4* mutant. The Shell (pink), Rubisco matrix (blue), thylakoids (green), and PPTs (yellow) are highlighted. Bottom row shows enlarged views of the boxed regions above. Scale bar: 100 nm.