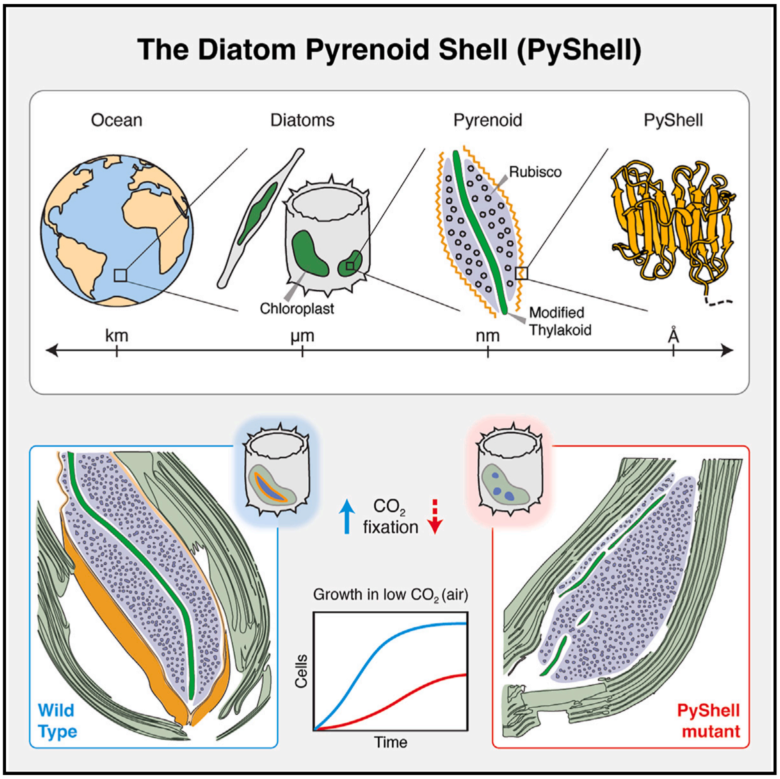
Cell

Diatom pyrenoids are encased in a protein shell that enables efficient CO_2 fixation

Graphical abstract



Authors

Ginga Shimakawa, Manon Demulder, Serena Flori, ..., Giovanni Finazzi, Benjamin D. Engel, Yusuke Matsuda

Correspondence

ben.engel@unibas.ch (B.D.E.),
yusuke@kwansei.ac.jp (Y.M.)

In brief

Identification and characterization of a protein lattice around the pyrenoid compartments of diatoms reveals that these prolific marine algae evolved a distinct pyrenoid architecture to promote Rubisco's CO_2-fixing activity.

Highlights

- Identification of the PyShell, a protein sheath that surrounds diatom pyrenoids
- Multiscale imaging of PyShell lattices from *in situ* architecture to *in vitro* structure
- PyShell knockout disrupts pyrenoid morphology and function, impairing cell growth
- The PyShell is widely conserved, enabling much of the ocean's CO_2 fixation

Shimakawa et al., 2024, Cell *187*, 5919–5934
October 17, 2024 © 2024 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.cell.2024.09.013

CellPress

Cell

CellPress
OPEN ACCESS

Article

Diatom pyrenoids are encased in a protein shell that enables efficient CO_2 fixation

Ginga Shimakawa,[1,13] Manon Demulder,[2,13] Serena Flori,[2,3,10,13] Akihiro Kawamoto,[4,13] Yoshinori Tsuji,[1,5]
Hermanus Nawaly,[1] Atsuko Tanaka,[6] Rei Tohda,[4,11] Tadayoshi Ota,[1] Hiroaki Matsui,[1] Natsumi Morishima,[1]
Ryosuke Okubo,[1] Wojciech Wietrzynski,[2] Lorenz Lamm,[2,7] Ricardo D. Righetto,[2] Clarisse Uwizeye,[3] Benoit Gallet,[8]
Pierre-Henri Jouneau,[9] Christoph Gerle,[4,12] Genji Kurisu,[4] Giovanni Finazzi,[3] Benjamin D. Engel,[2,*]
and Yusuke Matsuda[1,14,*]

[1]Department of Bioscience, School of Biological and Environmental Sciences, Kwansei Gakuin University, 1 Gakuen-Uegahara, Sanda, Hyogo 669-1330, Japan
[2]Biozentrum, University of Basel, Spitalstrasse 41, 4056 Basel, Switzerland
[3]Laboratoire de Physiologie Cellulaire et Végétale, Université Grenoble-Alpes, CNRS, CEA, INRAE, IRIG-DBSCI, 17 rue des Martyrs, 38000 Grenoble, France
[4]Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan
[5]Graduate School of Biostudies, Kyoto University, Kyoto 606-8502, Japan
[6]Department of Chemistry, Biology and Marine Science, Faculty of Science, University of the Ryukyus, Nishihara 903-0213, Japan
[7]HelmholtzAI, Helmholtz Munich, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany
[8]University of Grenoble Alpes, CEA, CNRS, IBS, 38000 Grenoble, France
[9]University of Grenoble Alpes, CEA, IRIG-MEM, 38000 Grenoble, France
[10]Present address: European Molecular Biology Laboratory, Meyerhofstraße 1, 69117 Heidelberg, Germany
[11]Present address: Department of Life Science, Graduate School of Science, University of Hyogo, Kouto, Hyogo, Japan
[12]Present address: Life Science Research Infrastructure Group, RIKEN SPring-8 Center, Kouto, Hyogo, Japan
[13]These authors contributed equally
[14]Lead contact
*Correspondence: ben.engel@unibas.ch (B.D.E.), yusuke@kwansei.ac.jp (Y.M.)
https://doi.org/10.1016/j.cell.2024.09.013

SUMMARY

Pyrenoids are subcompartments of algal chloroplasts that increase the efficiency of Rubisco-driven CO_2 fixation. Diatoms fix up to 20% of global CO_2, but their pyrenoids remain poorly characterized. Here, we used *in vivo* photo-crosslinking to identify pyrenoid shell (PyShell) proteins, which we localized to the pyrenoid periphery of model pennate and centric diatoms, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. *In situ* cryo-electron tomography revealed that pyrenoids of both diatom species are encased in a lattice-like protein sheath. Single-particle cryo-EM yielded a 2.4-Å-resolution structure of an *in vitro* TpPyShell1 lattice, which showed how protein subunits interlock. *T. pseudonana* TpPyShell1/2 knockout mutants had no PyShell sheath, altered pyrenoid morphology, and a high-CO_2 requiring phenotype, with reduced photosynthetic efficiency and impaired growth under standard atmospheric conditions. The structure and function of the diatom PyShell provide a molecular view of how CO_2 is assimilated in the ocean, a critical ecosystem undergoing rapid change.

INTRODUCTION

Diatoms are one of the most dominant groups of phytoplankton in the ocean. They are responsible for 15%–20% of annual global primary production,[1,2] powering the Earth's carbon cycle and feeding energy into vast marine food webs. Despite their importance, the underlying molecular mechanisms that enable diatoms to efficiently assimilate CO_2 remain poorly understood. Diverse clades of eukaryotic algae, including diatoms, rely on a biophysical CO_2-concentrating mechanism (CCM) to thrive in CO_2-limited aquatic environments. Algal CCMs use HCO_3^- transporters to accumulate dissolved inorganic carbon (DIC) in the chloroplast and then use carbonic anhydrases (CAs) to convert this HCO_3^- into a high local concentration of CO_2 in the pyrenoid, a chloroplast subcompartment packed with the carbon-fixing enzyme ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco). The pyrenoid thereby provides Rubisco with an abundance of its CO_2 substrate, while suppressing Rubisco's competitive oxygenase reaction, enabling rates of carbon fixation that exceed those of land plants.[3–6]

Pyrenoids are a general feature of algal CCMs. However, these chloroplast subcompartments have convergently evolved numerous times and exhibit a wide variety of morphologies,[7–9] indicating that pyrenoids in different clades may have distinct components and specialized mechanisms. To date, only the pyrenoid of the freshwater green alga *Chlamydomonas reinhardtii* has been characterized in molecular detail. The *C. reinhardtii* pyrenoid consists of a spherical matrix of densely packed Rubisco complexes, surrounded by a starch sheath and fenestrated by a network of membrane tubules.[10–12] These tubules contain an α-type CA that produces a source of CO_2 at the center of the pyrenoid.[13–16] The matrix is formed by liquid-liquid phase separation of Rubisco with its linker protein, essential pyrenoid component 1 (EPYC1),[17–20] and it dynamically condenses or disperses in response to changes in CO_2 concentration.[21–23]

In contrast to the *C. reinhardtii* pyrenoid, the Rubisco matrix of the diatom pyrenoid has an elongated shape and is typically traversed along its long axis by one or two specialized thylakoids.[24–26] Several proteins have been localized to the pyrenoid of the marine diatom, *Phaeodactylum tricornutum*. In addition to Rubisco, this pyrenoid contains β-type CAs,[27] fructose 1,6-bisphosphate aldolases (FBAs),[28] and a θ-type CA specifically localized in the lumen of the thylakoids at the center of the pyrenoid.[6,29] Although these observations strongly suggest that diatom pyrenoids increase CO_2 concentration around Rubisco in a similar fashion to green algae, the pyrenoid components in *P. tricornutum* have distinct origins and arose from endosymbiotic red algae, stramenopile host cells, or diatom-specific bacterial gene transfer.[28,30,31] In other words, the pyrenoids of diatoms and green algae may have convergently evolved similar functions from a different set of proteins.

In this study, we identify and characterize a distinct component of diatom pyrenoids that is not present in *C. reinhardtii*: the pyrenoid shell (PyShell). This proteinaceous sheath tightly encases the Rubisco matrix, is required for establishing pyrenoid architecture, and is essential for efficient photosynthesis and cell growth. We directly observe PyShells in both pennate diatoms (*P. tricornutum*) and centric diatoms (*Thalassiosira pseudonana*), while bioinformatic analysis suggests that PyShells are common in several clades of marine algae and, thus, likely play a major role in driving the ocean's carbon cycle.

RESULTS

Identification and localization of diatom PyShell proteins

To identify components of diatom pyrenoids, we performed *in vivo* photo-crosslinking,[32] then disrupted the cells and looked for proteins that co-migrated with the Rubisco large subunit (RbcL) by sucrose density gradient centrifugation and SDS-PAGE (diagrammed in Figure 1A). *P. tricornutum* cells were fed with L-photo-leucine and L-photo-methionine, synthetic amino acid derivatives with diazirine rings in their side chains. Because they are structurally similar to natural amino acids, these photo-reactive amino acids (pAAs) are taken up by the cells and incorporated during protein synthesis.[33] These *P. tricornutum* cells were then irradiated with UV light, causing the pAAs to form reactive carbenes that enable "zero-distance" photo-crosslinking with directly interacting proteins. Crude extracts from the photo-crosslinked cells were separated by SDS-PAGE and immunoblotted for RbcL ("procedure A," Figure 1B). In only the sample with both pAAs and UV irradiation, we observed a RbcL-positive band trapped in the stacking gel, which we analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). In an alternative approach, the crude extracts were separated by sucrose density gradient followed by SDS-PAGE ("procedure B," Figure 1C). The sample treated with both pAAs and UV irradiation showed RbcL bands in denser sucrose fractions, which we subjected to LC-MS/MS analysis.

From these two procedures, we identified more than 100 candidates for *P. tricornutum* pyrenoid proteins. We then filtered this list for the presence of the stramenopile-specific plastid-targeting sequence (ER+ASAFAP) at the N terminus,[34] yielding 22 and 36 candidate chloroplast proteins from procedures A and B, respectively (Tables S1A and S1B). Eleven of the candidate proteins were identified in both procedures (Figure S1A). In addition to known pyrenoid proteins such as Rubisco, β-CAs,[27] and FBAs (C1 and C5),[28] we identified several new candidates, including an abundant protein of unknown function encoded by the gene *Pt45465*. We chose to study this uncharacterized protein and its homologs in the model diatoms *P. tricornutum* and *T. pseudonana*.

The *P. tricornutum* gene *Pt45465* (*PtPyShell1a*) is located on chromosome 7 together with a paralog *Pt45466* (*PtPyShell2a*), which shares 74% identity. There are duplications of both genes on chromosome 28: *Pt50215* (*PtPyShell1b*) and *Pt50214* (*PtPyShell2b*). PtPyShell2 was also detected as a putative pyrenoid protein in procedure A (Table S1A). Analysis by quantitative PCR (qPCR) showed that *PtPyShell1a/b* and *PtPyShell2a/b* are expressed in wild-type (WT) *P. tricornutum* cells grown under normal atmospheric CO_2 (0.04%, hereafter "LC" for "low CO_2") and high CO_2 (1%, hereafter "HC") (Figure S1D). All PtPyShell proteins harbor stramenopile-type plastid-targeting sequences. These proteins contain no transmembrane helices and thus are likely localized to the stroma. We defined two conserved regions (CR1 and CR2) in the PyShell proteins (Figures S1B and S1C), which were composed of >50% hydrophobic amino acids. Using CR1 and CR2 as reference sequences, we searched for candidate PyShell genes and found homologs primarily in diatoms and haptophytes, but also in a few marine algae from other clades (Figure S1F).

In *T. pseudonana*, we identified three putative PyShell genes on chromosome 9: *Tp7881* (*TpPyShell1*), *Tp23918* (*TpPyShell2*), and *Tp7883* (*TpPyShell3*) (Figure S1B). *TpPyShell1* shares 92% identity with *TpPyShell2* and 75% identity with *TpPyShell3*. We also noted three additional putative PyShell "family members" that share around 30%–40% identity with *TpPyShell1*: *Tp24512* on chromosome 12 (*TpPyShell4*), *Tp3883* on chromosome 3 (*TpPyShell5*), and *Tp8449* on chromosome 10 (*TpPyShell6*). Expression of these six TpPyShell genes was analyzed by qPCR in *T. pseudonana* WT cells grown under LC and HC, indicating that TpPyShell1 and 2 are the most abundant isoforms (Figure S1D). We therefore decided to primarily focus on these two closely homologous proteins in our study of *T. pseudonana*.

We next checked the subcellular localization of PyShell proteins in *P. tricornutum* and *T. pseudonana* by fluorescence microscopy (Figures 1D and S2). We generated strains of these two diatom species expressing PtPyShell1a, PtPyShell2a, TpPyShell1, TpPyShell2, or TpPyShell3 fused to a C-terminal green fluorescent protein (GFP) tag. In *P. tricornutum*, PtPyShell1a:GFP and

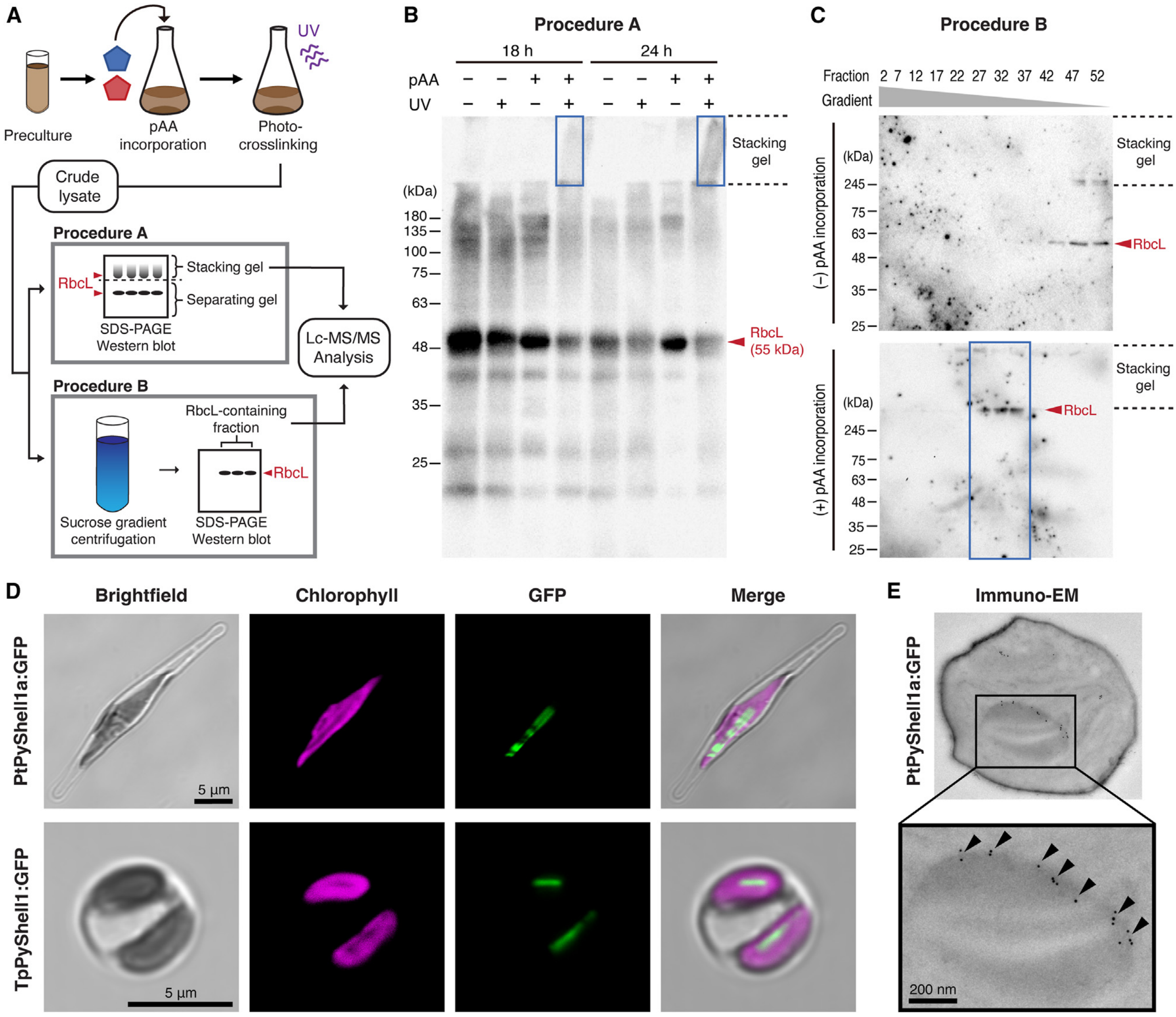


Figure 1. Identification of PyShell proteins in diatoms

(A) Proteomics-based workflow for detecting pyrenoid proteins in *P. tricornutum*. Cells were cultured with (+) or without (−) photo-reactive amino acids (pAAs), photo-crosslinked *in vivo* with UV irradiation, and then disrupted by sonication.

(B and C) The crude extracts were subjected to either (B) SDS-PAGE (“procedure A”) or (C) 22%–55% (w/v) sucrose density gradient centrifugation (“procedure B”). Gel shift of the crosslinked Rubisco was detected by immunoblotting against the Rubisco large subunit (RbcL). Rubisco-containing gels or collected fractions (indicated by blue boxes in B and C) were digested by trypsin and analyzed by LC-MS/MS (for list of candidate Rubisco interactors, see Table S1).

(D) Confocal images of PtPyShell1a:GFP in *P. tricornutum* (top row) and TpPyShell1:GFP in *T. pseudonana* (bottom row). See Figure S2 for additional examples. Scale bars: 5 μm.

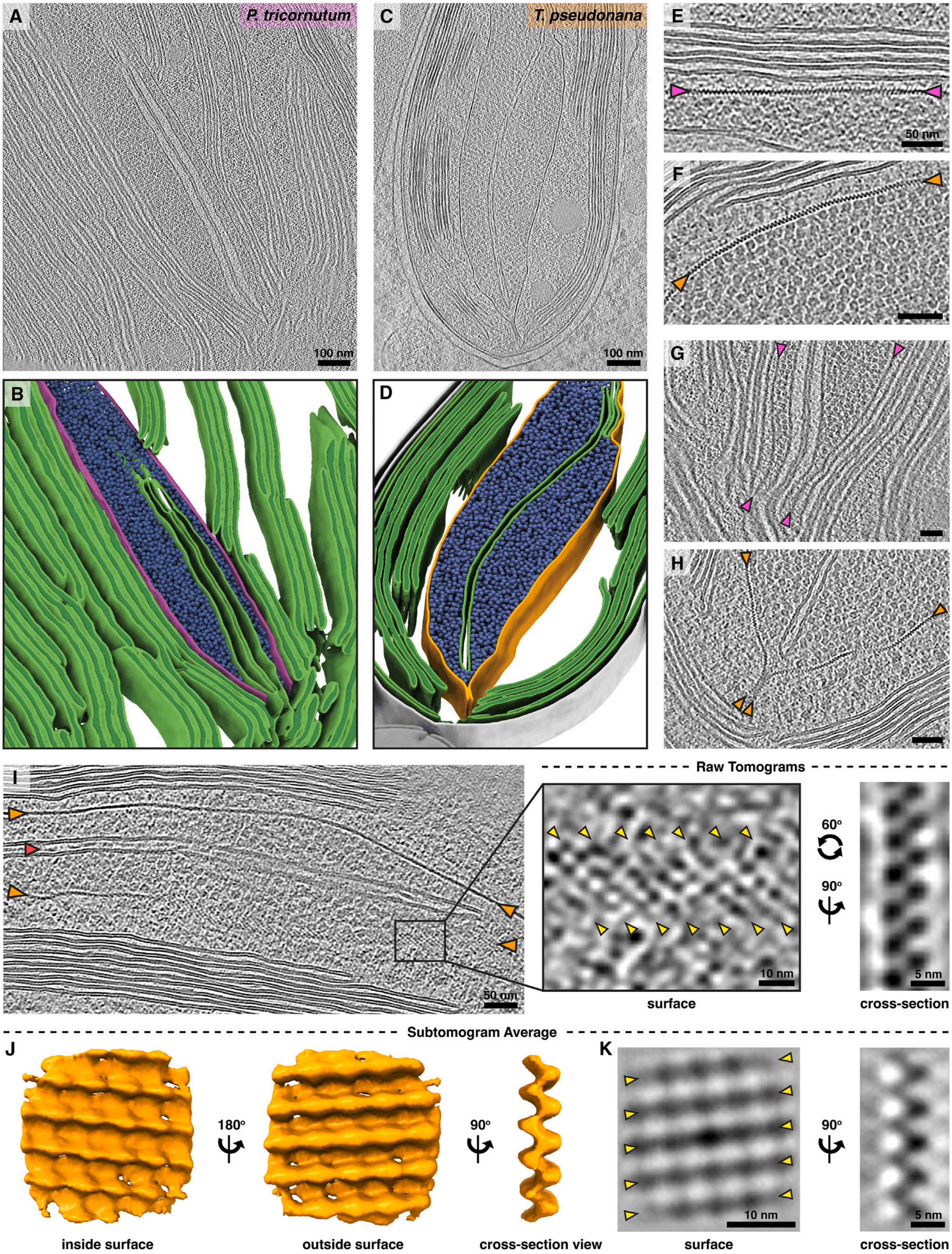
(E) Immunogold-labeling transmission electron microscopy (TEM) of a *P. tricornutum* PtPyshell1a:GFP transformant probed with an anti-GFP antibody. Gold particles are indicated by black arrowheads. The darker contrast region in the boxed inset corresponds the pyrenoid’s Rubisco matrix. Scale bar: 200 nm.

See also Figures S1 and S2.

PtPyShell2a:GFP signal was clearly detected in a “hollow rod” shape at the center of the chloroplast where the chlorophyll autofluorescence was dim, suggesting localization to the pyrenoid, perhaps surrounding the Rubisco matrix (Figures 1D, S2A, and S2B). In *T. pseudonana*, we similarly observed TpPyShell1:GFP, TpPyShell2:GFP, and TpPyShell3:GFP signal in a rod shape at the center of the chloroplast (Figures 1D and S2B). We further analyzed *P. tricornutum* by immunoelectron microscopy, with anti-GFP nanogold localization confirming that the PtPyShell1a:GFP proteins were accumulated along the peripheral regions of the pyrenoid (Figure 1E).

Molecular architecture of the PyShell lattice inside native diatom cells

To gain a higher-resolution view of the pyrenoid periphery, we turned to *in situ* cryo-electron tomography (cryo-ET).[35–37]



(legend on next page)

P. tricornutum and *T. pseudonana* cells were vitreously plunge-frozen on EM grids, thinned with a focused ion beam (FIB),[38,39] and imaged in three dimensions with cryo-ET. We observed that the pyrenoids of both diatom species are surrounded by a proteinaceous sheath, which tightly encloses the Rubisco matrix (Figures 2A–2D and S3). Closer inspection of these sheaths revealed that they are apparently formed from a repetitive lattice of protein subunits (Figures 2E, 2F, S3C, and S3G). We hereafter refer to this pyrenoid-encapsulating shell as "the PyShell"; its location is consistent with our observations of GFP-tagged PyShell proteins in *P. tricornutum* and *T. pseudonana*, while our structural and mutational analysis described later in this study definitively implicate PyShell proteins in sheath formation.

The native cellular views provided by cryo-ET revealed some species-specific differences in pyrenoid architecture. In *P. tricornutum*, the PyShell is relatively flat and straight. Two specialized thylakoids penetrate the Rubisco matrix and run the length of the pyrenoid (Figures 2A, 2G, and S3B). The luminal space of these traversing thylakoids is swollen and sometimes filled with dense particles (red arrowheads). At the two ends of the pyrenoid, the PyShell closely associates with these two traversing thylakoids, which exit the pyrenoid and connect to the rest of the thylakoid network (Figures 2A, 2B, 2G, and S3A). In *T. pseudonana*, the PyShell has more regions of high local curvature. This pyrenoid is also bisected by one or two specialized thylakoids (Figures 2C, 2D, 2H, 2I, and S3F) that frequently contain dense particles in their lumen; however, we never observed these thylakoids exiting the pyrenoid. Instead, at the two ends of the pyrenoid, the PyShell interacts with itself like a zipper to seal the Rubisco matrix (Figures 2C, 2D, 2H, and S3E). Due to the limited cell area visualized by cryo-ET, we cannot definitely conclude that the pyrenoid-traversing thylakoids in *T. pseudonana* are disconnected from the rest of the thylakoid network. However, if such connection sites exist, they are much rarer than in *P. tricornutum*. We observed enough cells in asynchronous cultures to conclude that these architectural differences are characteristic of the two diatom species and not due to cell division.

In our raw tomograms, the PyShell showed different features, depending on its orientation. When observed in cross-section, it resembled a solid line, a chain of dots, or a zig-zag (Figures 2I right, S3C, and S3G). However, when the PyShell twisted 90° to show its face, we could observe a lattice of subunits producing clear stripe patterns (Figures 2I middle, S3C, and S3G). To understand the three-dimensional (3D) structure of the PyShell, we performed subtomogram averaging (STA)[40,41] of subvolumes picked along PyShell sheaths from our highest quality tomograms of *T. pseudonana*. After iterative alignment and classification steps, we ultimately resolved an ∼20-Å-density map of the native *T. pseudonana* PyShell using 14,341 subvolumes from seven tomograms (Figures 2J and S4). The STA density contains the stripe and zig-zag features seen in the raw tomograms, and reveals the 3D architecture of a tightly packed pseudocrystalline protein lattice.

High-resolution *in vitro* structure of the *T. pseudonana* PyShell lattice

We required even higher resolution to determine precisely how individual PyShell proteins assemble to form a tight protein lattice. Therefore, we reconstituted the PyShell *in vitro* and performed single-particle cryoelectron microscopy (cryo-EM). We expressed and purified recombinant TpPyShell1, which when concentrated *in vitro* to 2 mg/mL, self-assembled to form both flat sheets and hollow tubes with an outer diameter of ∼32 nm (Figures S5A–S5C). Following cryo-EM imaging and particle picking along the tubes, we used single-particle analysis (SPA) and helical reconstruction to attain a 2.4-Å density map (Figures 3A, S5E, and S5F), enabling us to build an atomic model of TpPyShell1 assembled in a homo-oligomeric lattice (Figures 3B–3E and S5G).

The *in vitro* tube is an assembly of TpPyShell1 proteins in two alternating poses: half of the monomers face inward, while the other half face outward with a 90° in-plane rotation relative to the inward-facing monomers (Figures 3A, 3C, and 3D). Thus, the minimal building block of the tube is a homodimer of TpPyShell1 proteins adopting these two poses (Figure 3E).

Each TpPyShell1 monomer contains 16 β-strands arranged in two parallel β-sheets, one slightly more extended than the other (Figure 3B, teal and purple for more- and less-extended sheets, respectively). The TpPyShell1 monomer has an internal pseudo-2-fold symmetry that subdivides the two β-sheets into the two conserved regions that we previously identified by sequence homology: CR1 and CR2 (Figure S5H). A short α-helix spanning residues 169–181 (Figures 3B and S5H, pink) is positioned along the wall of the less-extended β-sheet and connects CR1 with CR2. Our TpPyShell1 model starts at Trp69 because the N-terminal domain is poorly resolved and likely flexible; a second short α-helix appears to be present in this region (approximately spanning residues 64–72) but could not be clearly modeled. The C-terminal domain extends from one monomer and inserts its

Figure 2. *In situ* cryo-ET reveals the native architecture of the PyShell inside diatom cells

(A–D) Magenta labels and arrowheads: *P. tricornutum*. Orange labels and arrowheads: *T. pseudonana*. (A and C) Two-dimensional (2D) overview slices through tomograms and (B and D) corresponding 3D segmentations (green: thylakoids, blue: Rubisco complexes, magenta or orange: PyShell). Scale bars: 100 nm.

(E and F) Close-up views of native PyShells (marked by arrowheads) in both diatom species. Scale bars: 50 nm.

(G and H) Comparison of pyrenoid ends. In *P. tricornutum*, there is a gap in the PyShell that allows entry of two specialized thylakoids into the pyrenoid. In *T. pseudonana*, two apposing sheets of the PyShell bind each other to seal the pyrenoid matrix. Scale bars: 50 nm.

(I) Molecular details of the PyShell in raw tomograms. Left: overview revealing a stripe pattern when the PyShell twists to show its surface view (red arrowhead: particles inside the lumen of traversing thylakoid). Center: zoom-in on the surface view, with the major stripes of the PyShell lattice marked with yellow arrowheads. Right: zoom-in on a cross-section view, showing an apparent lattice of dimers. Scale bars: 50 nm in left, 10 nm in middle, 5 nm in right.

(J and K) Subtomogram average (STA) of the PyShell from *T. pseudonana*, displayed in 3D isosurface view (J), as well as 2D slices (K) showing the surface view (yellow arrowheads: major stripes of lattice) and cross-section view. Scale bars: 10 nm in left, 5 nm in right. See Figure S3 for additional cryo-ET images from both species.

See also Figures S3 and S4 and Videos S1 and S2.

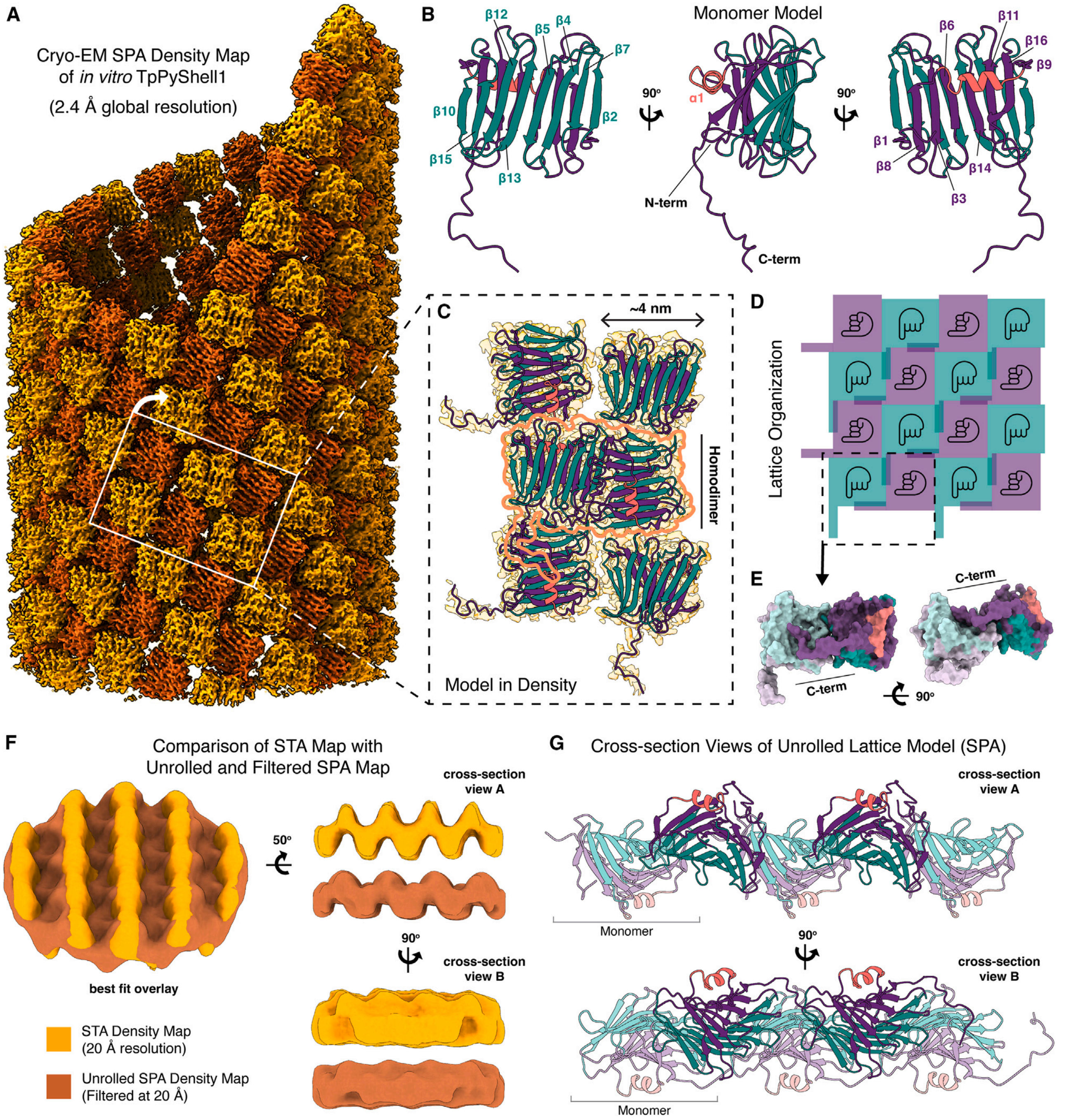


Figure 3. High-resolution *in vitro* structure of the *T. pseudonana* PyShell lattice

(A) Cryo-EM density map obtained by single-particle analysis (SPA) and helical reconstruction of TpPyShell1, which assembles into a tube *in vitro*. Global resolution: 2.4 Å (see Figures S5E and S5F).

(B) Cartoon model of the TpPyShell1 monomer. The two β-sheets (each composed of eight β-strands) and the adjacent α-helix are indicated in teal, purple, and pink, respectively. See Figure S5H for the monomer's pseudo-2-fold symmetry.

(C) Models of six TpPyShell1 monomers fit into the cryo-EM density map from (A) (yellow). The minimal building block of the tube's lattice is a homodimer of TpPyShell1 proteins (outlined in orange), which are flipped and rotated 90° relative to each other.

(D) Schematic representation of this lattice arrangement, with hands indicating the flipping and rotation of monomers. The pinky finger represents the C-terminal domain (C-term).

(E) Surface model representation of a homodimer unit from the lattice. The C-term extends and contacts a putative pocket in the adjacent monomer (see also Figure S5J).

(legend continued on next page)

positively charged Arg292 residue into a negatively charged pocket on the adjacent monomer of the opposite pose, possibly providing a stabilizing interaction for the lattice (Figures 3E and S5J). We also observed small ~1 nm gaps in the lattice at the junctions between four monomers; the residues surrounding these gaps do not have obvious surface charge or hydrophobic properties (Figure S5I).

We next compared the lattice architecture of *in vitro* TpPyShell1 to the architecture of the PyShell sheath from inside *T. pseudonana* cells (Figures 3F and S6). To do so, we unrolled the high-resolution SPA map of the TpPyShell1 tube to match the lateral subunit periodicity that we measured from TpPyShell1 flat sheets (Figures S5C and S6A–S6E). The unrolled SPA map was then filtered to match the 20-Å resolution of the *in situ* STA map (Figures 3F and S6E). The two maps were overlaid and then scored for correlation with an in-plane rotational search, determining an optimal fit from the scores and visual inspection (Figures S6F and S6G). The two 20-Å maps are not identical (perhaps reflecting additional structural heterogeneity *in situ*), but the overall features align well. From this analysis, we conclude that the SPA density of the *in vitro* TpPyShell1 lattice shows good correspondence with the STA density of the native PyShell sheath when compared at 20 Å resolution. This helps confirm that the pyrenoid sheath we observe inside diatom cells is indeed composed of PyShell proteins and also indicates that the overall subunit architecture of the *in vitro* TpPyShell1 lattice is likely physiologically relevant.

To better visualize the molecular architecture of a flat TpPyShell1 lattice, we fit models of the TpPyShell1 monomers into the full-resolution unrolled SPA map. Cross-section views through the flat lattice model (Figure 3G) show how the more-extended β-sheets (teal) align to form a continuous wall at the center of the PyShell. The less-extended β-sheets (purple) protrude inward and outward, positioning the short α-helices (pink) on each lobe as the most distant domain from the central wall of the PyShell lattice. Although this simplified model using only TpPyShell1 is consistent with the native PyShell architecture at the modest 20-Å resolution of our *in situ* average, our localization of GFP-tagged PyShell proteins (Figure S2) suggests that the native PyShell lattice is likely more heterogeneous and composed of multiple PyShell isoforms, which have similar structures that would require a much higher resolution *in situ* STA map to distinguish. Other factors that might bind the native PyShell lattice asymmetrically or substoichiometrically are also not resolved in our STA map (see "limitations of the study").

PyShell mutants have altered pyrenoid morphology and reduced CO_2 fixation

To understand the physiological role of the PyShell *in vivo*, we performed simultaneous gene disruptions of *TpPyShell1* and *TpPyShell2* in *T. pseudonana* using a CRISPR-Cas9 (D10A) nickase approach that we recently developed for diatom gene editing.[42] Because these two genes share high sequence identity (92%), we were able to design a single set of guide RNAs specifically targeting the CR1 domain of both genes (Figures S7A and S7B). Two independent biallelic double-knockout mutants were successfully obtained, denoted ΔTpPyShell1/2-1 (*m1*) and ΔTpPyShell1/2-2 (*m2*) (Figure S7C). Western blotting indicated that *m1* and *m2* lacked both the TpPyShell1 and TpPyShell2 proteins (Figure 4A). Expression of other PyShell family genes (*TpPyShell3*, *4*, *5*, and *6*) did not increase to compensate for the loss of TpPyShell1 and 2 (Figure S7D).

Mutants *m1* and *m2* both showed severely inhibited growth in normal atmospheric CO_2 (LC, Figure 4B). Compared with WT, the mutants had a longer lag phase, a slower growth rate, and took twice as long to reach stationary phase. In contrast, the growth profiles of WT, *m1*, and *m2* were similar when supplemented with 1% CO_2 (HC, Figure 4C). This indicates that disruption of the major PyShell genes in *T. pseudonana* gives a clear HC-requiring phenotype, presumably because of an impaired CCM. To test this hypothesis, we analyzed the photosynthetic affinity of WT and mutant cells for DIC by measuring net O_2 evolution rate at increasing DIC concentrations (Figures 4D and S7F). Whereas WT cells reached their maximum photosynthetic rate (P_{max}) at <0.5 mM [DIC], the mutants *m1* and *m2* required >10 mM [DIC] to reach their P_{max}. Other photosynthetic parameters, including the DIC compensation point and apparent photosynthetic conductance, also support the highly impaired photosynthetic activity in these mutants (Table S2), indicating that they cannot efficiently provide CO_2 to Rubisco in seawater with less than 10 mM [DIC]. Currently, the average [DIC] near the ocean surface is ~2 mM.[43]

To investigate how this HC-requiring mutant phenotype is related to pyrenoid morphology, we reconstructed the 3D architecture of WT, *m1*, and *m2* cells using FIB scanning electron microscopy (FIB-SEM). Cells were cryo-fixed at high pressure to improve sample preservation, then subjected to freeze substitution and resin embedding. Following FIB-SEM imaging (Figure 4E), we used 3D segmentation to quantify the volumes and shapes of chloroplast regions (Figures 4F–4J). In WT *T. pseudonana*, pyrenoids consisted of a single elongated compartment occupying around 10% of the chloroplast volume, similar to our previous measurements of *P. tricornutum*.[9] Conversely, *m1* and *m2* chloroplasts contained multiple ellipsoid pyrenoid-like structures of heterogeneous size and higher sphericity than WT, indicating that removal of the PyShell causes fragmentation of the Rubisco matrix and loss of its normal elongated architecture. Furthermore, the total measured pyrenoid volume decreased to <5% of the chloroplast in the mutants (Figure 4J), suggesting that a portion of the Rubisco may be dispersed in the stroma or present as small aggregates not detectable by our FIB-SEM imaging.

(F) Comparison of the *in situ* STA map from Figure 2J (yellow), with the *in vitro* SPA map (red), which has been unrolled to a flat lattice and filtered to the same 20-Å resolution as the STA map. Both isosurfaces displayed with thresholds of 1.5 standard deviations. Left: overlay of the two maps in their best fitting orientation (see Figure S6 for details). Right: side views of the two maps, corresponding to cross-section views through the PyShell lattice.

(G) Lattice model generated by fitting TpPyShell1 monomer structures into the full-resolution unrolled SPA map. The same cross-section views are shown as in (F).

See also Figures S5 and S6 and Video S3.

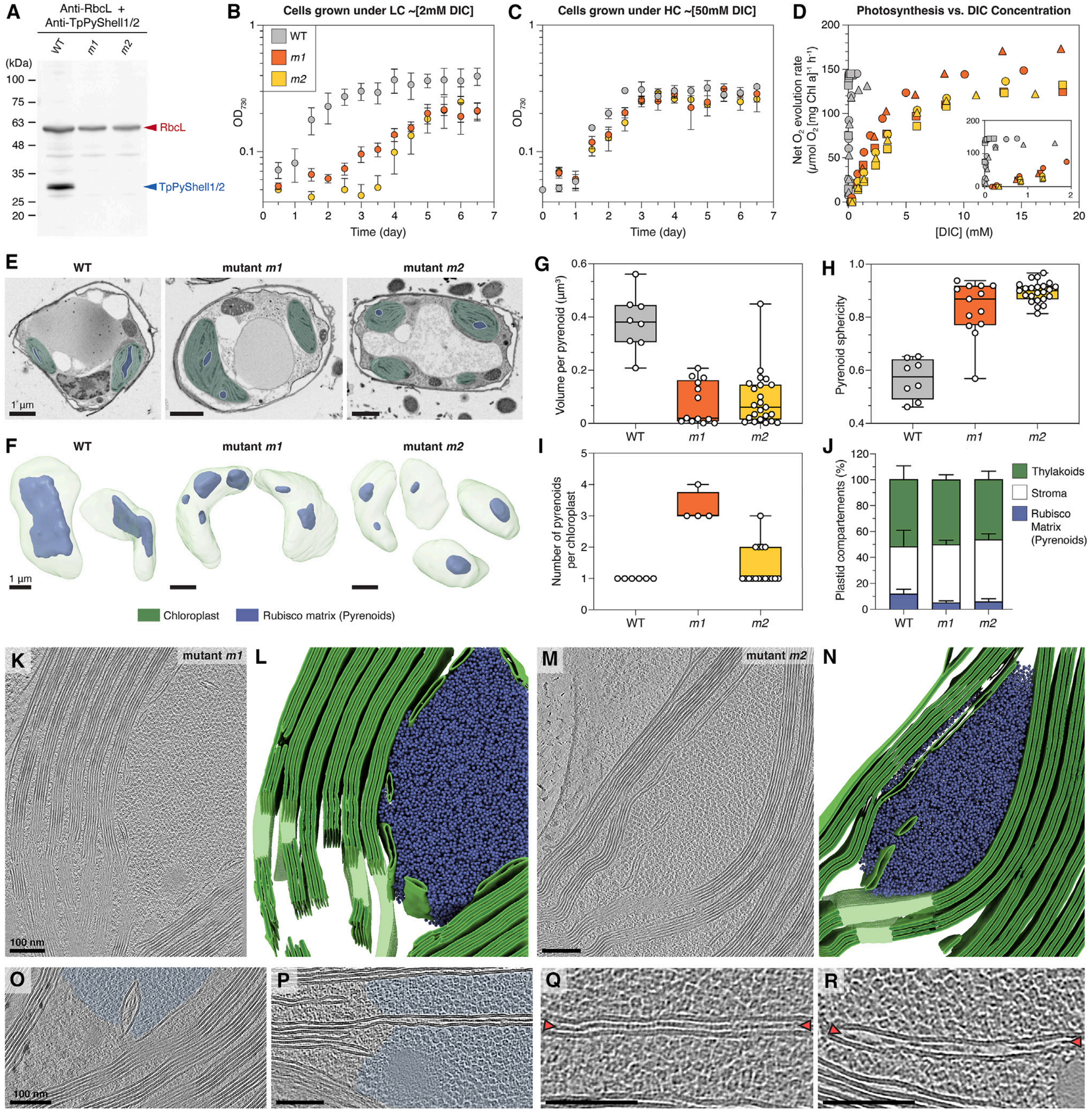


Figure 4. Phenotypes of *T. pseudonana* PyShell-deficient mutants

Mutants *m1* and *m2* (ΔTpPyShell1/2) have slightly different gene truncations (see Figures S7A–S7C).

(A) Western blot of crude cell extract (5 µg each) with anti-PyShell1/2 antibody, confirming the absence of TpPyShell1 and 2 proteins in the mutants.

(B and C) Growth of WT cells (gray), *m1* (orange), and *m2* (yellow) in air-level CO_2 (LC; 0.04%; equivalent to ~2 mM [DIC]) and high CO_2 (HC; 1%; equivalent to ~50 mM [DIC]) conditions, under continuous light (40 µmol photons m^{-2} s^{-1}). Points: mean, error bars: standard deviation ($n = 3$, biological replicates).

(D) Dependence of photosynthetic activity (measured by O_2 evolution under 900 µmol photons m^{-2} s^{-1} constant actinic light) on DIC concentration (set by supplementing with bicarbonate) in WT cells (gray), *m1* (orange), and *m2* (yellow). Cells were preconditioned in LC conditions. Data from three independent experiments are shown with different symbols.

(E and F) FIB-SEM imaging of WT, *m1*, and *m2* cells, shown as (E) 2D slices through the raw tomographic data and (F) 3D segmentations of the chloroplast (green) and pyrenoid Rubisco matrix (blue) volumes. Scale bars: 1 µm.

(legend continued on next page)

We next performed *in situ* cryo-ET of *m1* and *m2* cells to gain a molecular-resolution view of the mutant pyrenoids (Figures 4K–4R). Consistent with the FIB-SEM imaging, we observed pyrenoid-like aggregates of Rubisco matrix that were rounder than WT pyrenoids. With the resolution afforded by cryo-ET, we confirmed that these mutant pyrenoids were not encased in a PyShell sheath. Nevertheless, the cohesion of the Rubisco matrix was maintained without the PyShell, consistent with a recent report of a linker protein that mediates phase separation of Rubisco in diatoms.[44] The rounder shape of the mutant pyrenoids and clear boundary between the Rubisco matrix and the stroma (Figures 4O, 4P, and S7E) are also consistent with phase separation. Strikingly, the specialized thylakoids with luminal particles that normally traverse the long axis of WT pyrenoids were strongly disturbed in the mutants. Thylakoids containing luminal particles were frequently seen in the peripheral regions of mutant pyrenoids (Figures 4Q, 4R, and S7E) but often appeared fragmented and did not pass through the matrix center (Figures 4K–4N). Therefore, the PyShell seems to play important roles in maintaining the elongated shape and singular cohesiveness of *T. pseudonana* pyrenoids as well as in helping position the specialized thylakoids on an end-to-end trajectory bisecting the Rubisco matrix.

DISCUSSION

In this study, we discovered the PyShell, a protein sheath that encases the pyrenoids of *P. tricornutum* and *T. pseudonana*, model species of the pennate and centric diatom clades, respectively (Figure 1). We characterized the structure of the *T. pseudonana* PyShell lattice across scales, from near-atomic resolution *in vitro* to molecular resolution within native diatom cells (Figures 2 and 3). Our functional analysis of *T. pseudonana* PyShell deletion mutants showed that the PyShell sheath maintains pyrenoid architecture and is essential for efficient performance of the CCM and, thereby, the ability of diatoms to grow by assimilating environmental CO_2 (Figure 4).

How does the PyShell contribute so significantly to diatom CCM function? Comparison with the well-characterized pyrenoid of the green alga *C. reinhardtii* may help provide some insight. Reaction-diffusion modeling of *C. reinhardtii*[45] suggests that all pyrenoid-based CCMs require the following essential features: (1) aggregation of most of the chloroplast's Rubisco enzymes, (2) a local source of high CO_2 concentration at the center of this Rubisco aggregate, and (3) a diffusion barrier at the aggregate border to prevent CO_2 leakage. Our data indicate that the PyShell contributes to the first two essential pyrenoid features (Figure 5A), and we wonder whether the PyShell may directly perform the third (Figure 5B).

In *C. reinhardtii*, the molecular mechanism underlying aggregation of pyrenoid Rubisco (essential feature #1) has been well characterized. The multivalent linker protein EPYC1 is necessary and sufficient for Rubisco to undergo liquid-liquid phase separation, both *in vitro* and *in vivo*, giving rise to a condensed yet fluid pyrenoid matrix.[17–20] Rubisco phase separation appears to be common throughout evolution, as Rubisco has been observed to form liquid condensates during the biogenesis of cyanobacterial carboxysomes.[46–48] Evidence is now mounting that the Rubisco matrix of diatom pyrenoids also forms by phase separation. Recently, the multivalent linker protein PYCO1 was shown to localize to the *P. tricornutum* pyrenoid matrix and phase separate with Rubisco *in vitro*.[44] Furthermore, our FIB-SEM and cryo-ET imaging of the *T. pseudonana* ΔTpPyShell1/2 mutants (*m1* and *m2*) show that Rubisco continues to aggregate in the absence of the PyShell, forming rounder bodies with clear boundaries between Rubisco and stroma that are consistent with phase-separated condensates (Figures 4E, 4F, 4H, 4K–4P, and S7E). The PyShell is thus not required for phase separation of Rubisco. However, it may still play a role in concentrating the majority of Rubisco at one location. In the *m1* and *m2* mutants, Rubisco forms multiple smaller condensates throughout the chloroplast, and the combined volume of these dispersed condensates is less than a single WT pyrenoid, suggesting that some Rubisco may not enter the condensed phase (Figures 4E–4G, 4I, and 4J). We speculate that a direct or indirect interaction between the PyShell and Rubisco helps concentrate all the condensed Rubisco into the pyrenoid matrix. Indeed, in some cryo-ET images, it appears that a single layer of ordered Rubisco is lined up along the PyShell lattice (Figures 2F, S3D, and S3H). The precise mechanism of this direct or indirect PyShell-Rubisco interaction requires further investigation.

C. reinhardtii and diatoms use a common strategy to produce a local source of high CO_2 concentration at the center of the pyrenoid's Rubisco matrix (essential feature #2). Both algae have thylakoid-derived membrane systems that cross the center of the pyrenoids, taking the shape of cylindrical tubules in *C. reinhardtii*[12,49] and specialized thylakoid sheets in *P. tricornutum* and *T. pseudonana*.[25,50] Inside the luminal space of these pyrenoid-traversing tubules and thylakoids, there are CAs (α-type CAH3 in *C. reinhardtii*, θ-type Ptθ-CA1 in *P. tricornutum*, and Tpθ-CA2 in *T. pseudonana*)[6,15,29,51] that convert HCO_3^- into CO_2 at the center of the pyrenoid. This helps increase Rubisco efficiency by providing the enzyme with a high concentration of its CO_2 substrate, outcompeting unproductive binding to O_2. However, in the *T. pseudonana m1* and *m2* mutants, the specialized thylakoids are mislocalized to the periphery of the Rubisco condensate and, therefore, cannot provide a source of CO_2 at the center of the pyrenoid (Figure 5A). Our

(G–J) Morphometric quantification of pyrenoids (Rubisco matrix regions) from FIB-SEM data: (G) volume per pyrenoid, (H) pyrenoid sphericity, (I) number of pyrenoids per cell, (J) percent chloroplast volume occupied by pyrenoid, thylakoids, and stroma. Boxplots in (G)–(I) show median (center line), 75%–25% percentiles (box borders), and max-min values (whiskers). Error bars in (J): standard deviation (*n* chloroplasts = 8 WT; 4 *m1*; 17 *m2*).

(K–R) Cryo-ET of *m1* and *m2* cells. Overviews (K and M: tomographic slices, L and N: 3D segmentations) show higher sphericity of the Rubisco matrix and failure of specialized thylakoids to properly traverse the matrix. (O and P) Close-up tomographic slices showing the defined border of the Rubisco matrix (light blue) in the PyShell mutants. (Q and R) Close-up tomographic slices showing luminal particles (red arrowheads) inside the mislocalized specialized thylakoids. Scale bars: 100 nm. See Figure S7E for additional cryo-ET images of the mutants.

See also Figure S7.

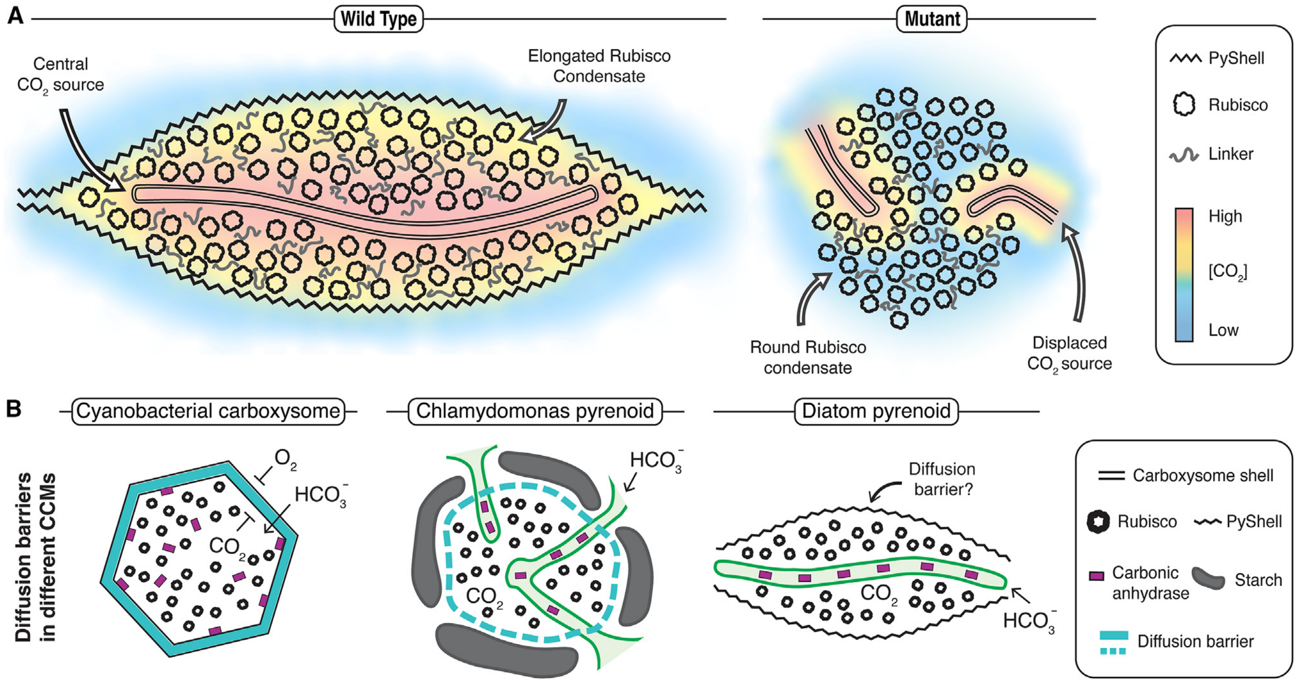


Figure 5. The PyShell's role in organizing diatom pyrenoid architecture, and open questions about PyShell permeability

(A) In wild-type cells, the PyShell encloses the Rubisco matrix, enforcing an elongated pyrenoid shape. One or two specialized thylakoids traverse the long axis of the pyrenoid. Carbonic anhydrase (CA) inside the lumen of these thylakoids generates CO_2, which diffuses through the thylakoid membranes and permeates the Rubisco matrix, enabling efficient carbon fixation. In the PyShell-deficient mutants (Figure 4), much of the Rubisco matrix remains aggregated by a linker protein but forms a rounder, more ellipsoid shape. The specialized thylakoids fail to bisect the matrix, delocalizing the source of CO_2 from the pyrenoid center. This defective CO_2-concentrating mechanism (CCM) underlies the mutants' high-CO_2 requiring phenotype.

(B) Diagrams of the CCMs in cyanobacterial carboxysomes and *C. reinhardtii* pyrenoids compared with our present understanding of diatom pyrenoids. All three compartments demonstrate Rubisco clustering and a local CA-generated source of CO_2. There is evidence that carboxysome shells and *C. reinhardtii* starch may serve as diffusion barriers that limit leakage of CO_2 and entry of O_2. It remains to be determined whether the PyShell serves a similar diffusion barrier function for diatom pyrenoids. Given how tightly the PyShell encapsulates the pyrenoid, it is also an open question how sugar substrates (ribulose 1,5-bisphosphate [RuBP]) and products (3-phosphoglyceric acid [3-PGA]) transit between the stroma and the Rubisco matrix.

cryo-ET observations clearly show that the PyShell is necessary to correctly position these specialized thylakoids along the long axis of the Rubisco matrix (Figures 4K–4N). By constricting the pyrenoid matrix into an elongated compartment surrounding the CA-containing thylakoids, the PyShell minimizes the diffusion distance of CO_2 from CA-source to Rubisco-sink.

Modeling indicates that CCM efficiency is greatly enhanced when a pyrenoid's Rubisco matrix is surrounded by a diffusion barrier to limit CO_2 leakage (essential feature #3). In *C. reinhardtii*, the identity of this barrier is debated, with candidates including a pyrenoid-peripheral CA and the pyrenoid's surrounding starch sheath.[21,45] In diatoms, the PyShell forms a tight protein sheath around the Rubisco matrix that larger molecules certainly cannot pass. In this way, the PyShell is analogous to the proteinaceous shells of cyanobacterial carboxysomes,[52,53] which similarly form a dense wall around an aggregate of encapsulated Rubisco. There is experimental evidence that carboxysome shells may be selectively permeable, blocking diffusion of CO_2 and O_2 while permitting passage of HCO_3^- through small pores in the wall.[54–58] In this way, HCO_3^- would diffuse into the carboxysome, where it is converted to CO_2 by a CA, trapping a high concentration of CO_2 with Rubisco inside the carboxysome shell. Similarly, there is evidence that carboxysome shell pores can be gated with an airlock-like mechanism to allow selective passage of sugar metabolite substrates and products of Rubisco CO_2 fixation.[59–61] Our high-resolution *in vitro* structure of a homo-oligomerized TpPyShell1 tube revealed ~1 nm gaps in the otherwise densely packed lattice (Figure S5I). However, to understand whether the PyShell functions as a selective diffusion barrier (Figure 5B), detailed investigations are required to probe the PyShell's permeability to small molecules, in particular CO_2, O_2, and the sugar metabolites ribulose 1,5-bisphosphate (RuBP) and 3-phosphoglyceric acid (3-PGA). These studies must take into consideration that the *in vivo* PyShell may be heterogeneous and also bound by interacting proteins.

In summary, the *m1* and *m2* PyShell deletion mutants exhibit such a strong inhibition of photosynthetic efficiency and growth (Figures 4B–4D) because they may have defects in all three essential features of pyrenoid CCMs: Rubisco aggregation into the pyrenoid matrix, a CO_2 source at the center of the matrix, and—speculatively—a barrier around the matrix to prevent CO_2 leakage. Our physiological measurements (Figures 4D and S7F) show that the PyShell-deficient mutants require ~10 mM [DIC] to saturate their photosynthesis, which is roughly 5-fold higher than the concentration of DIC in the ocean. Thus, the PyShell is likely essential to maintain efficient carbon fixation of diatoms in the wild.

Our bioinformatic analysis indicates that PyShell orthologs are widespread mainly in diatoms and haptophytes (Figure S1F). Although these two clades are not closely related phylogenetically, both have a plastid of red algal origin. Orthologs are also found in some non-diatom stramenopiles, including pelagophytes and dictyochophytes, as well as several species of alveolata. It is plausible that the PyShell originated from the photosynthetic ancestor of haptophytes and the SAR supergroup (stramenopiles, alveolates, and Rhizaria), and, thereafter, evolved independently in each clade. Even in our cryo-ET comparison between two diatom species, we noted substantial differences in PyShell architecture, in particular at the pyrenoid ends (Figures 2A–2D, 2G, 2H, S3A, and S3E). Capturing the variations in PyShell architecture across evolution will require extensive cryo-ET of diverse algae, including in non-model species from culture collections or sampled directly from the environment.[62]

Diatoms and haptophytes produce immense biomass through the photosynthetic uptake of CO_2, which provides an energy source for much of the life in the ocean. To understand the global prevalence of the PyShell, we queried *Tara* Oceans metagenomic environmental sampling data in the Ocean Gene Atlas[63,64] and found that PyShell genes are broadly detected in marine environments around the world (Figure S1E). The PyShell therefore plays a major role in the CCM of environmentally relevant marine algae, which account for roughly half of carbon fixation in the ocean and, by extension, one quarter of carbon fixation on our planet. To forecast the future of the global carbon cycle, it will be important to understand how well PyShell-mediated carbon fixation can adapt to rapidly accelerating climate change.

Major efforts are underway to engineer cyanobacterial carboxysomes and algal pyrenoids into plants to increase their carbon-fixation capacity.[5,65] It is estimated that introducing such a CCM could increase yield by up to 60%, while reducing water and fertilizer requirements.[66] Recent progress has been made in assembling components of the *C. reinhardtii* pyrenoid inside *Arabidopsis thaliana*, including the EPYC1 linker, which causes Rubisco to condense to form a pyrenoid-like matrix within these plant chloroplasts.[67,68] Our discovery and characterization of the diatom PyShell expands the molecular toolbox with a lattice-forming protein that may be able to encapsulate these Rubisco condensates, providing a boundary between the artificial pyrenoid and the surrounding chloroplast. These engineering efforts hold potential for designing crops that grow faster, consume less resources, and are more resistant to environmental stress, helping feed the world's growing population in regions of the planet that climate change is rapidly making less arable. Although reducing CO_2 emissions is the immediate priority,[69] improving biological carbon capture may even one day help slow the progression of global warming by removing more CO_2 from the atmosphere.

Limitations of the study

PyShell structure

The N-terminal 68 residues of TpPyShell1 were not well resolved in our *in vitro* SPA map, likely due to flexibility, and were not modeled. The high resolution of the map did reveal a putative electrostatic interaction between the extended C-terminal domain and a pocket on the neighboring PyShell subunit (Figure S5J), but the role of this interaction remains to be tested experimentally. The biggest mystery from our SPA data is that the inside and outside surfaces of the TpPyShell1 lattice are structurally interchangeable (Figures 3C and 3G). When filtered to 20 Å resolution, the *in vitro* SPA map corresponds well to the STA map of the native PyShell (Figures 3F and S6). However, additional studies are required to understand whether the two surfaces of the PyShell are equivalent or distinct *in vivo*. For example, in the pyrenoid, Rubisco is only bound to the PyShell's inner surface, but we cannot say from our data whether this is driven by asymmetry of the PyShell surfaces or rather is a consequence of pyrenoid biogenesis events. Comparison of the SPA and STA maps reveals some extra density in the STA map at the lobes extending furthest from the sheet; this could perhaps be due to additional proteins binding the native PyShell (Figures 3F and S6F), but we cannot make confident conclusions about a small region of density at 20 Å resolution. With the cryo-ET dataset we acquired for this study, we were unable to improve the STA resolution further due to the small size of the PyShell monomers,[70] preferred orientation of particles picked along PyShell sheets, and potential heterogeneity of PyShell proteins within the native lattice.

PyShell heterogeneity

There are multiple homologous PyShell genes in both *P. tricornutum* and *T. pseudonana* (Figure S1B), with the latter expressing both TpPyShell1 and TpPyShell2 at high levels, along with lower levels of TpPyShell3-6 (Figure S1D). To resolve a high-resolution PyShell structure, we assembled the *in vitro* tube from only TpPyShell1, which also proves that a homogeneous solution of this single protein is sufficient to assemble a lattice. It is quite possible that multiple PyShell isoforms hetero-oligomerize to make the PyShell lattice *in vivo*. Indeed, we observed that GFP-tagged TpPyShell1, 2, and 3 each localize to the pyrenoid (Figure S2), and a parallel study by Nam et al. in this issue of *Cell* localized all six TpPyShell proteins to the pyrenoid, with distinct functions suggested for TpPyShell1/2 and TpPyShell4.[71] However, our cryo-ET analysis lacks the resolution to distinguish this heterogeneity due to the high homology and predicted structural similarity of different PyShell proteins.

PyShell permeability

Our *in vitro* SPA structure shows small ~1 nm gaps at the junctions between four monomers in the TpPyShell1 lattice (Figures 3C and S5I). Although these gaps are potentially large enough to allow gas and some small metabolites to pass through the PyShell wall, in-depth functional studies would be required to assay PyShell permeability.

Pyrenoid-traversing thylakoids

In cryo-ET of both diatom species, we observed dense particles within the lumen of pyrenoid-traversing thylakoids (Figures 2I, 4Q, 4R, S3B, and S3F). This is the location thought to be occupied by a CA (Ptθ-CA1 and Tpθ-CA2), but we lack the resolution to assign an identity to these luminal particles. Our cryo-ET data of the ΔTpPyShell1/2 mutants does not explain how thylakoids in the proximity of the Rubisco matrix acquire these luminal particles yet fail to properly bisect the matrix.

Prevalence of the PyShell in other algae

Our bioinformatic analysis indicates that PyShell homologs are commonly found throughout diatoms and haptophytes, and they may also be present in other algal clades (Figure S1F). However, this analysis is biased toward sequenced species, so some clades may be underrepresented.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Yusuke Matsuda (yusuke@kwansei.ac.jp).

Materials availability

Diatom strains and plasmids generated in this study are available from the lead contact without restriction.

Data and code availability

- Single-particle cryo-EM maps (EMD-37751), cryo-ET subtomogram averages (EMD-18709), and cellular tomograms (EMD-18710 to EMD-18713) are available in the Electron Microscopy Data Bank. Atomic models of the PyShell structure are deposited at the Protein Data Bank (PDB: 8WQP). Raw cryo-EM data (-11724) and raw cryo-ET data (EMPIAR-11747) used to generate the density maps are available on the Electron Microscopy Public Image Archive. The PyShell nucleotide sequences reported in this paper have been deposited in DDBJ/EMBL/GenBank under accession numbers XM_002179781 (*PtPyShell1a*), XM_002185069 (*PtPyShell1b*), OR682719 (*PtPyShell2a*), OR682720 (*PtPyShell2b*), XM_002292359 (*TpPyShell1*), XM_002292147 (*TpPyShell2*), XM_002292148 (*TpPyShell3*), and XM_002287584 (*Tpδ-CA3*). The proteomics data are available at ProteomeXchange (PXD041920) and jPOST (JPST001940).
- All original code has been deposited at Zenodo and is publicly available. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank Vandana Tomar Deopa, Kohei Ueda, and Kohei Yoneda for helping establish the mutant strains, as well as Nobuko Higashiuchi and Eri Nakayama for technical assistance. We thank Cécile Giustini and Haythem Hsine for data analysis and strain maintenance, Glen Wheeler and Colin Brownlee (Marine Biological Association of the UK) for supporting the first cryo-ET investigations, and Caitlyn McCafferty for help with bioinformatic analysis. We thank Rikuya Morimoto, Kanae Fukushima, and Taiki Fukuzawa for their help establishing the preparation of recombinant PyShell protein, Takanori Nakane and Keitaro Yamashita (University of Tokyo) for helping refine the helical models, and Atsunori Oshima (Nagoya University) and Yoshinori Fujiyoshi (Tokyo Medical and Dental University) for TEM instrumentation access at the beginning of the SPA cryo-EM study. We thank Guy Schoehn and Christine Moriscot (IBS/ISBG electron microscope facility; Grenoble Instruct-ERIC center) for access and assistance with FIB/SEM instrumentation. We thank Sagar Khavnekar, Jürgen Plitzko, and Wolfgang Baumeister (Max Planck Institute of Biochemistry), as well as Daniel Böhringer, Miroslav Peterek, and Bilal M. Qureshi (ETH ScopeM facility), for access to FIB and TEM instruments and assistance with cryo-ET data acquisition. Cryo-ET analysis was performed at the sciCORE (http://scicore.unibas.ch/) scientific computing center at the University of Basel. Funding is as follows: this work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI (19H01153 to Y.M., 23H04958 to G.K.), JST CREST "Cell dynamics" (JPMJCR20E1 to G.K.), European Research Council (ERC) advanced grant "ChloroMito" (833184 to G.F.), European project Plankt-ON (101099192 to G.F.), ERC consolidator grant "cryOcean" (fulfilled by the Swiss State Secretariat for Education, Research and Innovation, M822.00045 to B.D.E.), an EMBO short-term fellowship (8015) and Royal Society international exchanges (IESR3183037) to S.F., an Alexander Von Humboldt Foundation fellowship and a non-stipendiary EMBO long-term fellowship to R.D.R., and a Boehringer Ingelheim Fonds fellowship to L.L.

AUTHOR CONTRIBUTIONS

Y.M. and B.D.E. initiated the project; Y.T. and N.M. identified PyShell proteins; H.N. and T.O. generated the PyShell mutants; H.M. performed qPCR analysis; G.S. and T.O. performed growth and photosynthesis analyses; N.M. and R.O. analyzed the localization of GFP-fused PyShell proteins; H.N. and M.D. performed bioinformatic analyses; A.T. performed immunoelectron microscopy; R.T. expressed and purified recombinant TpPyShell1 protein; C.G. confirmed the tube structure of TpPyShell1; A.K. prepared cryo-EM grids, acquired data, and calculated the SPA density map; G.K. performed modeling and refinement; S.F. and W.W. acquired cryo-ET data; M.D., L.L., and R.D.R. analyzed cryo-ET data and performed STA; S.F. and B.G. performed sample preparation for FIB-SEM imaging; P.-H.J. acquired FIB-SEM tomograms; S.F. and C.U. performed FIB-SEM tomogram segmentation and analysis; G.S., M.D., S.F., G.K., G.F., B.D.E., and Y.M. wrote the manuscript with support from all other 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 MODEL AND STUDY PARTICIPANT DETAILS
    - ○ Diatom Cultures
    - ○ Genome editing of *TpPyShell1/2*
    - ○ Expression of GFP fusion proteins in *P. tricornutum* and *T. pseudonana*
    - ○ Bacterial expression of recombinant TpPyShell1 protein for cryo-EM analysis
- METHOD DETAILS
    - ○ *In vivo* cross-linking with photo-reactive amino acids
    - ○ Western blotting
    - ○ Identification of proteins by LC-MS/MS
    - ○ qRT-PCR of PyShell transcripts
    - ○ Confocal fluorescence microscopy
    - ○ Immunoelectron microscopy
    - ○ Purification of recombinant TpPyShell1 protein
    - ○ Cryo-EM grid prep and data acquisition
    - ○ Cryo-EM image processing and model building
    - ○ Cryo-ET sample prep and data acquisition
    - ○ Cryo-ET data analysis
    - ○ Subtomogram averaging of *T. pseudonana* PyShell
    - ○ Comparison of density maps from *in vitro* SPA and *in situ* STA
    - ○ Photosynthesis measurements
    - ○ FIB-SEM data acquisition and analysis
    - ○ Phylogenetic analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2024.09.013.

Received: October 18, 2023
Revised: March 13, 2024
Accepted: September 6, 2024
Published: October 1, 2024

REFERENCES

1. Falkowski, P.G., Barber, R.T., and Smetacek, V. (1998). Biogeochemical controls and feedbacks on ocean primary production. Science *281*, 200–207. https://doi.org/10.1126/science.281.5374.200.
2. Smetacek, V. (1999). Diatoms and the ocean carbon cycle. Protist *150*, 25–32. https://doi.org/10.1016/S1434-4610(99)70006-4.
3. Giordano, M., Beardall, J., and Raven, J.A. (2005). CO2 concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. Annu. Rev. Plant Biol. *56*, 99–131. https://doi.org/10.1146/annurev.arplant.56.032604.144052.
4. Tsuji, Y., Nakajima, K., and Matsuda, Y. (2017). Molecular aspects of the biophysical CO_2-concentrating mechanism and its regulation in marine diatoms. J. Exp. Bot. *68*, 3763–3772. https://doi.org/10.1093/jxb/erx173.
5. Hennacy, J.H., and Jonikas, M.C. (2020). Prospects for engineering biophysical CO(2) concentrating mechanisms into land plants to enhance yields. Annu. Rev. Plant Biol. *71*, 461–485. https://doi.org/10.1146/annurev-arplant-081519-040100.
6. 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.
7. 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.
8. Barrett, J., Girr, P., and Mackinder, L.C.M. (2021). Pyrenoids: CO_2-fixing phase separated liquid organelles. Biochim. Biophys. Acta Mol. Cell Res. *1868*, 118949. https://doi.org/10.1016/j.bbamcr.2021.118949.
9. Uwizeye, C., Decelle, J., Jouneau, P.-H., Flori, S., Gallet, B., Keck, J.-B., Bo, D.D., Moriscot, C., Seydoux, C., Chevalier, F., et al. (2021). Morphological bases of phytoplankton energy management and physiological responses unveiled by 3D subcellular imaging. Nat. Commun. *12*, 1049. https://doi.org/10.1038/s41467-021-21314-0.
10. Griffiths, D.J. (1970). The pyrenoid. Bot. Rev. *36*, 29–58. https://doi.org/10.1007/BF02859154.
11. Lacoste-Royal, G., and Gibbs, S.P. (1987). Immunocytochemical localization of ribulose-1,5-bisphosphate carboxylase in the pyrenoid and thylakoid region of the chloroplast of Chlamydomonas reinhardtii. Plant Physiol. *83*, 602–606. https://doi.org/10.1104/pp.83.3.602.
12. Engel, B.D., Schaffer, M., Kuhn Cuellar, L., Villa, E., Plitzko, 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.
13. Funke, R.P., Kovar, J.L., and Weeks, D.P. (1997). Intracellular carbonic anhydrase is essential to photosynthesis in Chlamydomonas reinhardtii at atmospheric levels of CO2. Demonstration via genomic complementation of the high-CO_2-requiring mutant ca-1. Plant Physiol. *114*, 237–244. https://doi.org/10.1104/pp.114.1.237.
14. Raven, J.A. (1997). CO2-concentrating mechanisms: a direct role for thylakoid lumen acidification? Plant Cell Environ. *20*, 147–154. https://doi.org/10.1046/j.1365-3040.1997.d01-67.x.
15. Karlsson, J., Clarke, A.K., Chen, Z.Y., Hugghins, S.Y., Park, Y.I., Husic, H.D., Moroney, J.V., and Samuelsson, G. (1998). A novel alpha-type carbonic anhydrase associated with the thylakoid membrane in Chlamydomonas reinhardtii is required for growth at ambient CO_2. EMBO J. *17*, 1208–1216. https://doi.org/10.1093/emboj/17.5.1208.
16. Hanson, D.T., Franklin, L.A., Samuelsson, G., and Badger, M.R. (2003). The Chlamydomonas reinhardtii cia3 mutant lacking a thylakoid lumen-localized carbonic anhydrase is limited by CO2 supply to RuBisCO and not photosystem II function in vivo. Plant Physiol. *132*, 2267–2275. https://doi.org/10.1104/pp.103.023481.
17. 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.
18. 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.
19. 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.
20. He, S., Chou, H.-T., Matthies, D., Wunder, T., Meyer, M.T., Atkinson, N., Martinez-Sanchez, A., Jeffrey, P.D., Port, S.A., Patena, W., et al. (2020). The structural basis of RuBisCO phase separation in the pyrenoid. Nat. Plants *6*, 1480–1490. https://doi.org/10.1038/s41477-020-00811-y.
21. Ramazanov, Z., Rawat, M., Henk, M.C., Mason, C.B., Matthews, S.W., and Moroney, J.V. (1994). The induction of the CO_2-concentrating mechanism is correlated with the formation of the starch sheath around the pyrenoid of *Chlamydomonas reinhardtii*. Planta *195*, 210–216. https://doi.org/10.1007/BF00199681.
22. Morita, E., Kuroiwa, H., Kuroiwa, T., and Nozaki, H. (1997). High localization of ribulose-1,5-bisphosphate carboxulase/oxygenase in the pyrenoids of *Chlamydomonas reinhardtii* (Chlorophyta), as revealed by cryofixation and immunogold electron microscopy. J. Phycol. *33*, 68–72. https://doi.org/10.1111/j.0022-3646.1997.00068.x.
23. Borkhsenious, O.N., Mason, C.B., and Moroney, J.V. (1998). The intracellular localization of ribulose-1,5-bisphosphate Carboxylase/Oxygenase in Chlamydomonas reinhardtii. Plant Physiol. *116*, 1585–1591. https://doi.org/10.1104/pp.116.4.1585.
24. Jenks, A., and Gibbs, S.P. (2000). Immunolocalization and distribution of form II RuBisCO in the pyrenoid and chloroplast stroma of *Amphidinium carterae* and form I RuBisCO in the symbiont-derived plastids of *peridinium foliaceum* (Dinophyceae). J. Phycol. *36*, 127–138. https://doi.org/10.1046/j.1529-8817.2000.99114.x.
25. Bedoshvili, Y.D., Popkova, T.P., and Likhoshway, Y.V. (2009). Chloroplast structure of diatoms of different classes. Cell Tissue Biol. *3*, 297–310. https://doi.org/10.1134/S1990519X09030122.
26. Flori, S., Jouneau, P.-H., Bailleul, B., Gallet, B., Estrozi, L.F., Moriscot, C., Bastien, O., Eicke, S., Schober, A., Bártulos, C.R., et al. (2017). Plastid thylakoid architecture optimizes photosynthesis in diatoms. Nat. Commun. *8*, 15885. https://doi.org/10.1038/ncomms15885.
27. Tachibana, M., Allen, A.E., Kikutani, S., Endo, Y., Bowler, C., and Matsuda, Y. (2011). Localization of putative carbonic anhydrases in two marine diatoms, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. Photosynth. Res. *109*, 205–221. https://doi.org/10.1007/s11120-011-9634-4.
28. Allen, A.E., Moustafa, A., Montsant, A., Eckert, A., Kroth, P.G., and Bowler, C. (2012). Evolution and functional diversification of fructose bisphosphate aldolase genes in photosynthetic marine diatoms. Mol. Biol. Evol. *29*, 367–379. https://doi.org/10.1093/molbev/msr223.
29. 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 tricornutum. Proc. Natl. Acad. Sci. USA *113*, 9828–9833. https://doi.org/10.1073/pnas.1603112113.
30. Nonoyama, T., Kazamia, E., Nawaly, H., Gao, X., Tsuji, Y., Matsuda, Y., Bowler, C., Tanaka, T., and Dorrell, R.G. (2019). Metabolic innovations underpinning the origin and diversification of the diatom chloroplast. Biomolecules *9*, 322. https://doi.org/10.3390/biom9080322.
31. Kroth, P.G., and Matsuda, Y. (2022). Carbohydrate metabolism. In The Molecular Life of Diatoms (Springer), pp. 465–492.

32. Suchanek, M., Radzikowska, A., and Thiele, C. (2005). Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. Nat. Meth. *2*, 261–267. https://doi.org/10.1038/nmeth752.

33. Shiota, T., Nishikawa, S.-i., and Endo, T. (2013). Analyses of protein–protein interactions by in vivo photocrosslinking in budding yeast. In Membrane Biogenesis: Methods and Protocols, D. Rapaport and J.M. Herrmann, eds. (Humana Press), pp. 207–217.

34. Gruber, A., Vugrinec, S., Hempel, F., Gould, S.B., Maier, U.-G., and Kroth, P.G. (2007). Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. Plant Mol. Biol. *64*, 519–530. https://doi.org/10.1007/s11103-007-9171-x.

35. Turk, M., and Baumeister, W. (2020). The promise and the challenges of cryo-electron tomography. FEBS Lett. *594*, 3243–3261. https://doi.org/10.1002/1873-3468.13948.

36. Hylton, R.K., and Swulius, M.T. (2021). Challenges and triumphs in cryo-electron tomography. iScience *24*, 102959. https://doi.org/10.1016/j.isci.2021.102959.

37. McCafferty, C.L., Klumpe, S., Amaro, R.E., Kukulski, W., Collinson, L., and Engel, B.D. (2024). Integrating cellular electron microscopy with multimodal data to explore biology across space and time. Cell *187*, 563–584. https://doi.org/10.1016/j.cell.2024.01.005.

38. 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.

39. Klumpe, S., Fung, H.K.H., Goetz, S.K., Zagoriy, I., Hampoelz, B., Zhang, X., Erdmann, P.S., Baumbach, J., Müller, C.W., Beck, M., et al. (2021). A modular platform for automated cryo-FIB workflows. eLife *10*, e70506. https://doi.org/10.7554/eLife.70506.

40. Wan, W., and Briggs, J.A. (2016). Cryo-electron tomography and subtomogram averaging. Methods Enzymol. *579*, 329–367. https://doi.org/10.1016/bs.mie.2016.04.014.

41. Pyle, E., and Zanetti, G. (2021). Current data processing strategies for cryo-electron tomography and subtomogram averaging. Biochem. J. *478*, 1827–1845. https://doi.org/10.1042/BCJ20200715.

42. Nawaly, H., Tsuji, Y., and Matsuda, Y. (2020). Rapid and precise genome editing in a marine diatom, *Thalassiosira pseudonana* by Cas9 nickase (D10A). Algal Res. *47*, 101855. https://doi.org/10.1016/j.algal.2020.101855.

43. Cole, J.J., Hararuk, O., and Solomon, C.T. (2021). Chapter 7. The carbon cycle: with a brief introduction to global biogeochemistry. In Fundamentals of Ecosystem Science, Second Edition, K.C. Weathers, D.L. Strayer, and G.E. Likens, eds. (Academic Press), pp. 131–160.

44. 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.

45. Fei, C., Wilson, A.T., Mangan, N.M., Wingreen, N.S., and Jonikas, M.C. (2022). Modelling the pyrenoid-based CO(2)-concentrating mechanism provides insights into its operating principles and a roadmap for its engineering into crops. Nat. Plants *8*, 583–595. https://doi.org/10.1038/s41477-022-01153-7.

46. 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.

47. Oltrogge, L.M., Chaijarasphong, T., Chen, A.W., Bolin, E.R., Marqusee, S., and Savage, D.F. (2020). Multivalent interactions between CsoS2 and RuBisCO mediate α-carboxysome formation. Nat. Struct. Mol. Biol. *27*, 281–287. https://doi.org/10.1038/s41594-020-0387-7.

48. Zang, K., Wang, H., Hartl, F.U., and Hayer-Hartl, M. (2021). Scaffolding protein CcmM directs multiprotein phase separation in β-carboxysome biogenesis. Nat. Struct. Mol. Biol. *28*, 909–922. https://doi.org/10.1038/s41594-021-00676-5.

49. Ohad, I., Siekevitz, P., and Palade, G.E. (1967). Biogenesis of chloroplast membranes. II. Plastid differentiation during greening of a dark-grown algal mutant (Chlamydomonas reinhardi). J. Cell Biol. *35*, 553–584. https://doi.org/10.1083/jcb.35.3.553.

50. Pyszniak, A.M., and Gibbs, S.P. (1992). Immunocytochemical localization of photosystem I and the fucoxanthin-chlorophyll *a/c* light-harvesting complex in the diatom *Phaeodactylum tricornutum*. Protoplasma *166*, 208–217. https://doi.org/10.1007/BF01322783.

51. Nawaly, H., Tanaka, A., Toyoshima, Y., Tsuji, Y., and Matsuda, Y. (2023). Localization and characterization θ carbonic anhydrases in Thalassiosira pseudonana. Photosynth. Res. *156*, 217–229. https://doi.org/10.1007/s11120-023-01007-z.

52. Shively, J.M., Ball, F., Brown, D.H., and Saunders, R.E. (1973). Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of Thiobacillus neapolitanus. Science *182*, 584–586. https://doi.org/10.1126/science.182.4112.584.

53. Melnicki, M.R., Sutter, M., and Kerfeld, C.A. (2021). Evolutionary relationships among shell proteins of carboxysomes and metabolosomes. Curr. Opin. Microbiol. *63*, 1–9. https://doi.org/10.1016/j.mib.2021.05.011.

54. Dou, Z., Heinhorst, S., Williams, E.B., Murin, C.D., Shively, J.M., and Cannon, G.C. (2008). CO2 fixation kinetics of Halothiobacillus neapolitanus mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO2. J. Biol. Chem. *283*, 10377–10384. https://doi.org/10.1074/jbc.M709285200.

55. Cai, F., Menon, B.B., Cannon, G.C., Curry, K.J., Shively, J.M., and Heinhorst, S. (2009). The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO2 leakage barrier. PLoS One *4*, e7521. https://doi.org/10.1371/journal.pone.0007521.

56. Mahinthichaichan, P., Morris, D.M., Wang, Y., Jensen, G.J., and Tajkhorshid, E. (2018). Selective permeability of carboxysome shell pores to anionic molecules. J. Phys. Chem. B *122*, 9110–9118. https://doi.org/10.1021/acs.jpcb.8b06822.

57. Faulkner, M., Szabó, I., Weetman, S.L., Sicard, F., Huber, R.G., Bond, P.J., Rosta, E., and Liu, L.N. (2020). Molecular simulations unravel the molecular principles that mediate selective permeability of carboxysome shell protein. Sci. Rep. *10*, 17501. https://doi.org/10.1038/s41598-020-74536-5.

58. Huang, J., Jiang, Q., Yang, M., Dykes, G.F., Weetman, S.L., Xin, W., He, H.L., and Liu, L.N. (2022). Probing the internal pH and permeability of a carboxysome shell. Biomacromolecules *23*, 4339–4348. https://doi.org/10.1021/acs.biomac.2c00781.

59. Klein, M.G., Zwart, P., Bagby, S.C., Cai, F., Chisholm, S.W., Heinhorst, S., Cannon, G.C., and Kerfeld, C.A. (2009). Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. J. Mol. Biol. *392*, 319–333. https://doi.org/10.1016/j.jmb.2009.03.056.

60. Cai, F., Sutter, M., Cameron, J.C., Stanley, D.N., Kinney, J.N., and Kerfeld, C.A. (2013). The structure of CcmP, a tandem bacterial microcompartment domain protein from the β-carboxysome, forms a subcompartment within a microcompartment. J. Biol. Chem. *288*, 16055–16063. https://doi.org/10.1074/jbc.M113.456897.

61. Larsson, A.M., Hasse, D., Valegård, K., and Andersson, I. (2017). Crystal structures of β-carboxysome shell protein CcmP: ligand binding correlates with the closed or open central pore. J. Exp. Bot. *68*, 3857–3867. https://doi.org/10.1093/jxb/erx070.

62. Mocaer, K., Mizzon, G., Gunkel, M., Halavatyi, A., Steyer, A., Oorschot, V., Schorb, M., Le Kieffre, C., Yee, D.P., Chevalier, F., et al. (2023). Targeted volume correlative light and electron microscopy of an environmental marine microorganism. J. Cell Sci. *136*, jcs261355. https://doi.org/10.1242/jcs.261355.

63. Delmont, T.O., Gaia, M., Hinsinger, D.D., Frémont, P., Vanni, C., Fernandez-Guerra, A., Eren, A.M., Kourlaiev, A., d'Agata, L., Clayssen, Q., et al. (2022). Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. Cell Genomics *2*, 100123. https://doi.org/10.1016/j.xgen.2022.100123.

64. Vernette, C., Lecubin, J., Sánchez, P., Tara Oceans Coordinators, Sunagawa, S., Delmont, T.O., Acinas, S.G., Pelletier, E., Hingamp, P., and Lescot, M. (2022). The Ocean Gene Atlas v2.0: online exploration of the biogeography and phylogeny of plankton genes. Nucl. Acids Res. *50*, W516–W526.

65. Borden, J.S., and Savage, D.F. (2021). New discoveries expand possibilities for carboxysome engineering. Curr. Opin. Microbiol. *61*, 58–66. https://doi.org/10.1016/j.mib.2021.03.002.

66. McGrath, J.M., and Long, S.P. (2014). Can the cyanobacterial carbon-concentrating mechanism increase photosynthesis in crop species? A theoretical analysis. Plant Physiol. *164*, 2247–2261. https://doi.org/10.1104/pp.113.232611.

67. 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.

68. Atkinson, N., Stringer, R., Mitchell, S.R., Seung, D., and McCormick, A.J. (2024). SAGA1 and SAGA2 promote starch formation around proto-pyrenoids in Arabidopsis chloroplasts. Proc. Natl. Acad. Sci. USA *121*, e2311013121. https://doi.org/10.1073/pnas.2311013121.

69. Ho, D.T. (2023). Carbon dioxide removal is not a current climate solution - we need to change the narrative. Nature *616*, 9. https://doi.org/10.1038/d41586-023-00953-x.

70. Russo, C.J., Dickerson, J.L., and Naydenova, K. (2022). Cryomicroscopy in situ: what is the smallest molecule that can be directly identified without labels in a cell? Faraday Discuss. *240*, 277–302. https://doi.org/10.1039/d2fd00076h.

71. Nam, O., McKenzie, C., Dowle, A., Dowson, M., Barrett, J., and Mackinder, L.C.M. (2024). A protein blueprint of the diatom CO2-fixing organelle Cell *187*, 5935–5950.e18. https://doi.org/10.1016/j.cell.2024.09.025.

72. Zaslavskaia, L.A., Lippmeier, J.C., Kroth, P.G., Grossman, A.R., and Apt, K.E. (2000). Transformation of the diatom Phaeodactylum tricornutum (Bacillariophyceae) with a variety of selectable marker and reporter genes. J. Phycol. *36*, 379–386. https://doi.org/10.1046/j.1529-8817.2000.99164.x.

73. Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres, S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. eLife *7*, e42166. https://doi.org/10.7554/eLife.42166.

74. Sander, J.D., Maeder, M.L., Reyon, D., Voytas, D.F., Joung, J.K., and Dobbs, D. (2010). ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. Nucleic Acids Res. *38*, W462–W468. https://doi.org/10.1093/nar/gkq319.

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. 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.

77. Kimanius, D., Dong, L., Sharov, G., Nakane, T., and Scheres, S.H.W. (2021). New tools for automated cryo-EM single-particle analysis in RELION-4.0. Biochem. J. *478*, 4169–4185. https://doi.org/10.1042/BCJ20210708.

78. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods *14*, 290–296. https://doi.org/10.1038/nmeth.4169.

79. Zhang, K. (2016). Gctf: real-time CTF determination and correction. J. Struct. Biol. *193*, 1–12. https://doi.org/10.1016/j.jsb.2015.11.003.

80. Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A.J., and Berger, B. (2019). Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. Nat. Methods *16*, 1153–1160. https://doi.org/10.1038/s41592-019-0575-8.

81. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

82. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501. https://doi.org/10.1107/S0907444910007493.

83. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr. D Struct. Biol. *75*, 861–877. https://doi.org/10.1107/S2059798319011471.

84. Yamashita, K., Palmer, C.M., Burnley, T., and Murshudov, G.N. (2021). Cryo-EM single-particle structure refinement and map calculation using Servalcat. Acta Crystallogr. D Struct. Biol. *77*, 1282–1291. https://doi.org/10.1107/S2059798321009475.

85. 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. Meth. *14*, 331–332. https://doi.org/10.1038/nmeth.4193.

86. 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.

87. Kremer, J.R., Mastronarde, D.N., and McIntosh, J.R. (1996). Computer visualization of three-dimensional image data using IMOD. J. Struct. Biol. *116*, 71–76. https://doi.org/10.1006/jsbi.1996.0013.

88. Khavnekar, 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.

89. Wan, W., Khavnekar, S., and Wagner, J. (2023). STOPGAP, an open-source package for template matching, subtomogram alignment, and classification. Preprint at bioRxiv. https://doi.org/10.1101/2023.12.20.572665.

90. Buchholz, T.-O., Jordan, M., Pigino, G., and Jug, F. (2018). Cryo-CARE: content-aware image restoration for cryo-transmission electron microscopy data. Preprint at arXiv. https://doi.org/10.48550/arXiv.1810.05420.

91. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612. https://doi.org/10.1002/jcc.20084.

92. Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: meeting modern challenges in visualization and analysis. Protein Sci. *27*, 14–25. https://doi.org/10.1002/pro.3235.

93. 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.

94. 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.

95. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539. https://doi.org/10.1038/msb.2011.75.

96. Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., and Minh, B.Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. *44*, W232–W235. https://doi.org/10.1093/nar/gkw256.

97. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. *49*, W293–W296. https://doi.org/10.1093/nar/gkab301.

98. Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., and Ranzuglia, G. (2008). MeshLab: an Open-Source Mesh Processing Tool. In 6th Eurographics Italian Chapter Conference, pp. 129–136.

99. Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al. (2012). 3D Slicer as an image computing platform for the quantitative imaging network. Magn. Reson. Imaging *30*, 1323–1341. https://doi.org/10.1016/j.mri.2012.05.001.

100. Gallet, B., Moriscot, C., Schoehn, G., and Decelle, J. (2024). Cryo-fixation and Resin Embedding of Biological Samples for Electron Microscopy and Chemical Imaging. Protocolsio. https://www.protocols.io/view/cryo-fixation-and-resin-embedding-of-biological-sa-bp2l62kndgqe/v1.

101. Guillard, R.R.L., and Ryther, 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.

102. 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.

103. Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol. *32*, 347–355. https://doi.org/10.1038/nbt.2842.

104. Motohashi, K. (2015). A simple and efficient seamless DNA cloning method using SLiCE from *Escherichia coli* laboratory strains and its application to SLiP site-directed mutagenesis. BMC Biotechnol. *15*, 47. https://doi.org/10.1186/s12896-015-0162-8.

105. Matsui, H., Hopkinson, B.M., Nakajima, K., and Matsuda, Y. (2018). Plasma membrane-type aquaporins from marine diatoms function as CO(2)/NH(3) channels and provide photoprotection. Plant Physiol. *178*, 345–357. https://doi.org/10.1104/pp.18.00453.

106. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402–408. https://doi.org/10.1006/meth.2001.1262.

107. 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.

108. 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.

109. 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.

110. Lamm, L., Righetto, R.D., Wietrzynski, W., Pöge, M., Martinez-Sanchez, A., Peng, T., and Engel, B.D. (2022). MemBrain: A deep learning-aided pipeline for detection of membrane proteins in Cryo-electron tomograms. Comput. Methods Programs Biomed. *224*, 106990. https://doi.org/10.1016/j.cmpb.2022.106990.

111. Jeffrey, S.W., and Haxo, F.T. (1968). Photosynthetic pigments of symbiotic dinoflagellates (zooxanthellae) from corals and clams. Biol. Bull. *135*, 149–165. https://doi.org/10.2307/1539622.

112. Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., et al. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. *12*, e1001889. https://doi.org/10.1371/journal.pbio.1001889.

113. Villar, E., Vannier, T., Vernette, C., Lescot, M., Cuenca, M., Alexandre, A., Bachelerie, P., Rosnet, T., Pelletier, E., Sunagawa, S., et al. (2018). The Ocean Gene Atlas: exploring the biogeography of plankton genes online. Nucleic Acids Res. *46*, W289–W295. https://doi.org/10.1093/nar/gky376.

114. Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J. Mol. Biol. *333*, 721–745. https://doi.org/10.1016/j.jmb.2003.07.013.

115. Afonine, P.V., Klaholz, B.P., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C., Adams, P.D., and Urzhumtsev, A. (2018). New tools for the analysis and validation of cryo-EM maps and atomic models. Acta Crystallogr. D Struct. Biol. *74*, 814–840. https://doi.org/10.1107/S2059798318009324.

116. de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., et al. (2015). Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. Science *348*, 1261605. https://doi.org/10.1126/science.1261605.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
α-TpPyShell1 and 2 antiserum (antigen peptide: GTARDLAEIWDNSS)	Japan Bio Serum, This paper	N/A
α-RbcL antibody	Agrisera	Cat#AS03 037; RRID: AB_2175406
polyclonal α-GFP antibody	AnaSpec	Cat#AS-29779; RRID: AB_317930
Bacterial and virus strains		
E.coli strain BL21(DE3)	Thermo Fisher Scientific	Cat#EC0114
Chemicals, peptides, and recombinant proteins		
TpPyShell1	This paper	N/A
L-photo-methionine	Thermo Fisher Scientific	Cat#22615
L-photo-leucine	Thermo Fisher Scientific	Cat#22610
ImmunoStar Zeta	Wako	Cat#4548995003669
TI blue	Nisshin EM	Cat#335
10-nm colloidal gold particles	BBI Solutions	#SKU EM.GC10/4
Critical commercial assays		
Protein assay kit	Bio-Rad	Cat#5000006
ZipTip 0.6 µL C18	Merck Millipore	Cat#ZTC18S096
SMARTer RACE 5'/3' kit	TaKaRa	Cat#634858
Deposited data		
PyShell single particle cryo-EM maps	This paper	EMDB: EMD-37751
PyShell cryo-ET subtomogram average	This paper	EMDB: EMD-18709
Cellular tomograms	This paper	EMDB: EMD-18710 to EMD-18713
Atomic models of the PyShell structure	This paper	PDB: 8WQP
Raw cryo-EM data	This paper	EMPIAR: EMPIAR-11724
Raw cryo-ET data	This paper	EMPIAR: EMPIAR-11747
Proteomics data	This paper	ProteomeXchange: PXD041920; jPOST: JPST001940
PtPyShell1a nucleotide sequence	This paper	DDBJ/EMBL/GenBank: XM_002179781
PtPyShell1b nucleotide sequence	This paper	DDBJ/EMBL/GenBank: XM_002185069
PtPyShell2a nucleotide sequence	This paper	DDBJ/EMBL/GenBank: OR682719
PtPyShell2b nucleotide sequence	This paper	DDBJ/EMBL/GenBank: OR682720
TpPyShell1 nucleotide sequence	This paper	DDBJ/EMBL/GenBank: XM_002292359
TpPyShell2 nucleotide sequence	This paper	DDBJ/EMBL/GenBank: XM_002292147
TpPyShell3 nucleotide sequence	This paper	DDBJ/EMBL/Genbank: XM_002292148
Tpδ-CA3 nucleotide sequence	This paper	DDBJ/EMBL/Genbank: XM_002287584
Experimental models: Organisms / strains		
Thalasiossira pseudonana	Bigelow culture collection	CCMP 1335
Phaeodactylum tricornutum	UTEX culture collection	UTEX642
Oligonucleotides		
CRSPR/Cas9 guide RNAs *TpPyShell1* and *TpPyShell2* 5'-TAACGGCATTGAAGGTACGA-3'	This paper	N/A
CRSPR/Cas9 guide RNAs *TpPyShell1* and *TpPyShell2* 5'-TCCCCGCGGCCCCAACACCG-3'	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Primer used for purification of recombinant TpPyShell1 protein 5'-ctgtacttccaa ggcATGGGCGGAACT-3'	This paper	N/A
Primer used for purification of recombinant TpPyShell1 protein 5'-ctcgaattcggatcc TCAAAACGCACGGC-3'	This paper	N/A
Primers used for qRT-PCR in *P. tricornutum* and *T. pseudonana*	This paper, Table S3A	N/A
Primers used for expression of GFP fusion proteins in *P. tricornutum* and *T. pseudonana*	This paper, Table S3B	N/A
Recombinant DNA		
pPha-T1	Zaslavskaia et al.[72]	N/A
pTha-NR	This paper	N/A
pET28a	Novagen	Cat#69864
Software and algorithms		
RELION v3.1	Zivanov et al.[73]	https://www3.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page
Proteome Discoverer 1.4	Thermo Fisher Scientific	https://www.thermofisher.com/ch/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/proteome-discoverer-software.html
ZiFiT Targeter	Sander et al.[74]	http://bindr.gdcb.iastate.edu/ZiFiT/
Cas-Designer	Park et al.[75]	http://www.rgenome.net/cas-designer/
CRISPOR	Concordet and Haeussler[76]	http://crispor.tefor.net/
RELION v4.0	Kimanius et al.[77]	https://www3.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page
CryoSPARC v4.0.2	Punjani et al.[78]	https://cryosparc.com/
Gctf	Zhang[79]	https://www2.mrc-lmb.cam.ac.uk/research/locally-developed-software/zhang-software/#gctf
Topaz	Bepler et al.[80]	https://cb.csail.mit.edu/cb/topaz/
AlphaFold2	Jumper et al.[81]	https://alphafold.ebi.ac.uk
COOT	Emsley et al.[82]	https://www2.mrc-lmb.cam.ac.uk/coot/coot/
Phenix	Liebschner et al.[83]	https://phenix-online.org/
REFMAC5	Yamashita et al.[84]	https://www.ccp4.ac.uk/html/refmac5.html
MotionCorr v2.1	Zheng et al.[85]	https://emcore.ucsf.edu/ucsf-software
MemBrain v2	Lamm et al.[86]	https://github.com/teamtomo/membrain-seg
Amira v 2021.2	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	Kremer et al.[87]	https://bio3d.colorado.edu/imod/index.html
TOMOMAN v0.6.9	Khavnekar and Wan[88]	https://github.com/williamnwan/TOMOMAN/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
STOPGAP	Wan et al.[89]	https://github.com/wan-lab-vanderbilt/STOPGAP
Cryo-CARE v0.2.1	Buchholz et al.[90]	https://github.com/juglab/cryoCARE_T2T
UCSF Chimera	Pettersen et al.[91]	https://www.rbvi.ucsf.edu/chimera/
UCSF ChimeraX	Goddard et al.[92]	https://www.rbvi.ucsf.edu/chimerax/
Serial EM software	Mastronarde[93]	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
Fibics Atlas 3D software for tomography	N/A	https://www.zeiss.com/microscopy/en/products/software/zeiss-atlas-5.html
Fiji	Schindelin et al.[94]	https://fiji.sc/
Clustal Omega	Sievers et al.[95]	https://www.ebi.ac.uk/jdispatcher/msa/clustalo
iTOL	Trifinopoulos et al.[96]; Letunic and Bork[97]	https://itol.embl.de
Prism 9	GraphPad	https://www.graphpad.com/
Mechlab	Cignoni et al.[98]	https://www.meshlab.net/
3D Slicer	Fedorov et al.[99]	https://www.slicer.org
Code to compute sphericity	This paper	https://gitlab.com/clariaddy/compute-sphericity-of-3d-object https://doi.org/10.5281/zenodo.13361310
Other		
R 2/1 holey carbon-foil 200-mesh copper EM grids	Quantifoil Micro Tools	Cat#N1-c15ncu20-01
R 1/4 SiO2-foil 200-mesh gold EM grids	Quantifoil Micro Tools	Cat#N1-S13nAu20-01
R 1.2/1.3 holey carbon-foil 200-mesh copper EM grids	Quantifoil Micro Tools	Cat#N1-C14nCu20-01
Autogrid supports	Thermo Fisher Scientific	N/A
Aquilos 1 or 2	Thermo Fisher Scientific	N/A
Titan Krios 300kV	Thermo Fisher Scientific	N/A
Vitrobot plunger	Thermo Fisher Scientific	N/A
Luminograph I	ATTO	N/A
EASY-nLC 1000 connected to LTQ Orbitrap XL	Thermo Fisher Scientific	N/A
Particle delivery system PDS-1000/He	BioRad	N/A
Confocal laser microscopes A1	Nikon	N/A
TCS SP8	Leica	N/A
EM UC7 Ultramicrotome	Leica	N/A
JEM-1011 electron microscope	JEOL	N/A
Vitrobot Mark IV	Thermo Fisher Scientific	N/A
gas-chromatography flame ionization detector GC-8A	Shimadzu	N/A
CrossBeam 550 microscope	Zeiss	N/A
HPM-100 High pressure freezing machine	Leica microsystems	N/A
AFS2 freeze substitution machine	Leica microsystems	N/A
Protocol of sample preparation for FIB-SEM	Gallet et al.[100]	https://doi.org/10.17504/protocols.io.bp2l62kndgqe/v1

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Diatom Cultures

The marine diatoms *P. tricornutum* Bohlin (UTEX642) and *T. pseudonana* (Hustedt) Hasle et Heimdal (CCMP 1335) were axenically and photoautotrophically cultured in artificial seawater medium with the addition of half-strength Guillard's 'F' solution[101,102] supplemented with 10 nM sodium selenite under continuous light (20°C, 40 μmol photons m^{-2} s^{-1}, fluorescent lamp). The cultures were aerated with ambient air (0.04% CO_2) or 1% CO_2 gas for LC or HC conditions, respectively. For the culture of *T. pseudonana*, the concentration of NaCl was lowered to 270 mM in the medium. Diatom cells were first inoculated from the plate to 50 mL Erlenmeyer flasks with gentle agitation, then cultures were transferred to 100 mL test tubes that were aerated (300 mL min^{-1}) with normal atmospheric air or air enriched with 1% (v/v) CO_2. The mutant cultures *m1* and *m2* were grown in the same conditions but supplemented with 5 mM bicarbonate.

Genome editing of *TpPyShell1/2*

A pair of single guide RNA (sgRNA) sequences were constructed to target both *TpPyShell1* and *TpPyShell2* using ZiFiT Targeter (http://zifit.partners.org/ZiFiT/),[74] followed by Cas-Designer (http://www.rgenome.net/cas-designer/), which considers specificity and microhomology-mediated joining.[75] The nucleotides 5'-TAACGGCATTGAAGGTACGA-3' (396–415 in *TpPyShell1* and 423–442 in *TpPyShell2*) and 5'-TCCCCGCGGCCCCAACACCG-3' (393–412 in *TpPyShell1* and 420–439 in *TpPyShell2*) were chosen (Figure S7A). Off-target search (Figure S7A) was performed with CRISPOR (http://crispor.tefor.net/).[76] For appropriate RNA transcription, an additional G was inserted at the 5′ end of the DNA sequence encoding each sgRNA.[103] The dual sgRNA vector targeting *TpPyShell1/2* was generated and introduced together with Cas9 (D10A) nickase vector into the *T. pseudonana* cells by particle bombardment according to the previous work.[42] The mutant strains were screened on agar medium supplemented with 100 μg/mL nourseothricin (Jena Bioscience) under HC conditions. To ensure monoclonal strains, each colony was restreaked four times, each time picking a single colony from the streak. Mutations were verified by TIDE analysis and direct sequencing (Figure S7B). Furthermore, we confirmed the genome sequences for the other PyShell genes (*TpPyShell3*−*6*) were not edited in the mutants.

Expression of GFP fusion proteins in *P. tricornutum* and *T. pseudonana*

Correct full-length coding sequences for PyShell orthologues in *P. tricornutum* and *T. pseudonana* (PtPyShell1a, PtPyShell2a, TpPyShell1, TpPyShell2, and TpPyShell3) were determined by RACE using a SMARTer RACE 5'/3' kit (TaKaRa). Sequences were amplified by PCR and cloned into pPha-T1 or pTha-NR vectors containing a fragment of enhanced GFP by a seamless ligation cloning extract method.[104] The resulting plasmids were introduced into each WT cell using particle bombardment (PDS-1000/He, BioRad, Tokyo, Japan), and transformants expressing GFP were screened by fluorescence microscopy.[27] Primers used are listed in Table S3B.

Bacterial expression of recombinant TpPyShell1 protein for cryo-EM analysis

The expression plasmid for TpPyShell with an N-terminal His6-tag on pET28a was transformed into *Escherichia coli* strain BL21(DE3). Cells were cultured at 37°C in 6 L of LB medium containing 100 μg/mL Kanamycin. When the OD_{600} reached 0.5, an IPTG solution was added to a final concentration of 0.1 mM for induction of TpPyShell expression, and the culture was incubated overnight at 37°C.

METHOD DETAILS

In vivo cross-linking with photo-reactive amino acids

P. tricornutum cells grown in LC were harvested at logarithmic growth phase and resuspended in fresh medium at a concentration of OD_{730} = 1.0−1.2, in the presence of 1 mM L-photo-leucine and 2 mM L-photo-methionine (Thermo Fisher Scientific, Waltham, MA, USA). Incorporation of these photo-reactive amino acids (pAA) was performed under illumination with blue (455 nm) and red (635 nm) LED light (50 μmol photons m^{-2} s^{-1}) for 6−24 hours. Subsequently, the cell cultures were irradiated with UV light (365 nm) for 30−45 min to perform *in vivo* photo-cross-linking. Cells were harvested and resuspended in 25 mM Tris-HCl (pH 7.0), then disrupted by sonication. Insoluble debris was removed by centrifugation, and the resulting supernatant was either subjected to SDS-PAGE (for gel digestion, "procedure A") or centrifuged on a 25−55% (w/v) linear sucrose gradient in 25 mM Tris-HCl (pH 7.0) at 210,000 × *g* for 4 h at 4°C (for solution digestion, "procedure B"). Aliquots (200 μL) of each fraction were collected, and the protein concentrations were determined with a protein assay kit (Bio-Rad, Hercules, CA, USA) using bovine serum albumin as a standard.

Western blotting

Proteins extracted as described above were electrophoretically separated by SDS-PAGE, transferred to PVDF membrane, and blocked with 1% (w/v) skim milk dissolved in phosphate-buffered saline (PBS) containing 0.05% (v/v) Tween 20. For detection of RbcL, a rabbit anti-RbcL antiserum generated against *P. tricornutum* RbcL partial peptide (Japan Bio Serum, Hiroshima, Japan) was used as the primary antibody (diluted 1:1000). Goat anti-rabbit IgG conjugated with horseradish peroxidase was used as the secondary antibody (diluted 1:10000). Immunoreactive signals were detected by an enhanced chemiluminescence reagent (ImmunoStar Zeta, Wako, Osaka, Japan) with a high sensitivity CCD imaging system (Luminograph I, ATTO, Tokyo, Japan).

For the confirmation of the deletions of TpPyShell1 and 2 in *m1* and *m2*, we disrupted the *T. pseudonana* WT and mutants grown under HC by sonication in 50 mM HEPES (pH 7.5) with a protease inhibitor cocktail (nacalai, Kyoto, Japan) to obtain the crude extracts. Each 5 μg of protein were loaded and analyzed by immunoblotting as mentioned above. A rabbit anti-TpPyShell1 and 2 antiserum (Japan Bio Serum) targeting the conserved peptide sequence "GTARDLAEIWDNSS" (residues 60-73 of TpPyShell1 and 59-72 of TpPyShell2) was used as the primary antibody. The anti-RbcL antibody was also used as a loading control.

Identification of proteins by LC-MS/MS

Proteins either in acrylamide gel ("Procedure A") or solution ("Procedure B") were subjected to reduction, alkylation, and digested by trypsin before injection into the LC-MS/MS. For gel samples, gel blocks (ca. 1 mm^3) were dehydrated in 100 μL acetonitrile for 10 min at room temperature. After the removal of acetonitrile, the gel block was dried in an evaporator and incubated in 25 mM NH_4HCO_3 containing 10 mM dithiothreitol for 1 h at 56°C. The gel block was washed in 100 μL of 25 mM NH_4HCO_3 and incubated with 55 mM iodoacetamide for 45 min at room temperature. After washing twice with 100 μL of 25 mM of NH_4HCO_3, the gel block was dehydrated in acetonitrile. The dried gel blocks were soaked in 50 mM NH_4HCO_3 containing 10 ng μL^{-1} trypsin at 37°C for 16−20 h. Digested peptides were extracted with 50% (v/v) acetonitrile containing 5% (v/v) formic acid, concentrated by an evaporator, and dissolved in 1% (v/v) formic acid. The solutions were desalted by ZipTip C18 (Merck Millipore). For solution samples, disulfide bonds were reduced in 50 mM Tris-HCl (pH 8.5) containing 10 mM dithiothreitol at 37°C for 1.5 h, and subsequently alkylated with 50 mM iodoacetamide for 30 min at room temperature. Proteins were digested with 2 ng μL^{-1} trypsin in 50 mM NH_4HCO_3 for 16−20 h at 37°C, and then the reaction was stopped by addition of 0.18% (v/v) formic acid. Peptide samples were concentrated in an evaporator, dissolved in 0.1% (v/v) formic acid, and then desalted by ZipTip C18 (Merck Millipore, Burlington, MA, USA). The digested samples were injected into EASY-nLC 1000 connected to LTQ Orbitrap XL (Thermo Fisher Scientific). Data of LC-MS/MS were analyzed by the software Proteome Discoverer 1.4 (Thermo Fisher Scientific) with the open genome data resource for *P. tricornutum* from JGI (Phatr2). Homologs of PtPyShell1 were identified by BLAST search on the reference genomes of *P. tricornutum* CCAP1055/1 (Phatr2) and *T. pseudonana* CCMP1335 (Thaps3). Two conserved regions (CR1 and CR2) were identified in all PyShell proteins using ClustalW sequence alignment.

qRT-PCR of PyShell transcripts

Transcript levels of PyShell genes were quantified by qRT-PCR in *P. tricornutum* and *T. pseudonana* (Figures S1D and S6D). In *T. pseudonana*, *TpδCA3* (Tp233) was also analyzed as a control LC-inducible gene. The internal controls were *Actin* (*Pt51157*), *GapC2* (*Pt51129*), and *Histone 4* (*Pt26896*) in *P. tricornutum* and *Actin* (*Tp25772*), *GapC3* (*Tp28241*), and *Histone H4* (*Tp3184*) in *T. pseudonana*, which were confirmed to be unresponsive to different CO_2 concentrations.[105] Transcript levels were calculated with the $2^{-\Delta\Delta Ct}$ method against each internal control separately.[106] Then, the average values of ΔΔCt were calculated for each replicate. Primers used are listed in Table S3A.

Confocal fluorescence microscopy

To observe subcellular localizations of GFP fusion proteins, we used confocal laser microscopes A1 (Nikon, Tokyo, Japan) or SP8 (Leica, Wetzlar, Germany). When imaging with the A1, chlorophyll autofluorescence was detected at 662–737 nm after excitation with a 638 nm laser, and GFP fluorescence was monitored at 500–550 nm following excitation at 488 nm. When imaging with the SP8, chlorophyll was excited by a 552 nm laser and detected at 600–750 nm. GFP was excited by a 488 nm laser and detected at 500–520 nm.

Immunoelectron microscopy

The strain of *P. tricornutum* expressing PtPyShell1a:GFP was fixed by as previously described[29] with small modifications in the polymerization step; the samples immersed in resin were polymerized at −30°C for 5 days under UV light. Thin sections cut with a Leica EM UC7 were mounted on nickel slot grids, followed by an edging step with 1% (w/v) sodium periodate. After the blocking step, the sections were reacted with polyclonal anti-GFP antibody (AnaSpec, Fremont, CA, US) diluted 1:500 in 3% (w/v) BSA in PBS at 25°C overnight. After rinsing with PBS, they were incubated for 60 min at room temperature with a goat anti-rabbit IgG conjugated to 10-nm colloidal gold particles (1:50 diluted in PBS; BBI Solutions, Crumlin, UK). The thin sections were stained with TI blue (Nisshin EM, Aichi, Japan), following washing with distilled water. The sections were observed with a JEM-1011 electron microscope (JEOL, Tokyo, Japan).

Purification of recombinant TpPyShell1 protein

Escherichia coli BL21(DE3) cells expressing TpPyShell1 protein (see culture conditions above) were harvested by two rounds of centrifugation (6,000 rpm, 10 min, 4°C, JLA-9.1000 rotor; Beckman), resuspended in Buffer A containing 50 mM Tris-HCl pH8.0, 0.3 M NaCl, 1 mM EDTA and 1 mM DTT, and then frozen for storage at -80°C or used immediately for cell disruption. Frozen cells were resuspended in 200 mL of freshly prepared pre-cooled Buffer A with 0.25 mM PMSF and disrupted by sonication. After cell debris and undisrupted cells were removed by centrifugation (45,000 rpm, 30 min, 4°C, 70Ti rotor; Beckman), supernatant was applied to an open column with Ni-IMAC resin (BIO-RAD). The column was washed with wash buffer (50 mM Tris-HCl pH8.0, 0.3 M NaCl, 10 mM Imidazole, 1 mM EDTA, 1 mM DTT), and then TpPyShell1 protein was eluted from the column using elution buffer (50 mM Tris-HCl pH8.0, 0.3 M NaCl, 300 mM Imidazole, 1 mM EDTA and 1 mM DTT). TEV protease equivalent to 3% of the TpPyShell1 concentration was added to the fractionated solution. His-tagged TpPyShell1 together with TEV protease was dialyzed overnight at 4°C in SnakeSkin dialysis tube (Thermo Fisher Scientific) against dialysis buffer (50 mM Tris-HCl pH8.0, 0.3 M NaCl, 1 mM EDTA and 1 mM DTT). The dialyzed solution was applied to an open column with Ni-IMAC resin equilibrated by the dialysis buffer, and the His-tag free TpPyShell1 was eluted. Collected TpPyShell1 was applied to Superdex75 16/60 equilibrated with the dialysis buffer, and the fraction containing TpPyShell1 was collected. The fraction containing TpPyShell1 was further concentrated by centrifugation (4000 g, 15 min, 4°C, SX4400 rotor; Beckman) with Amicon Ultra (M.W.4000) to reach a TpPyShell1 concentration of 2.0 mg/mL.

Cryo-EM grid prep and data acquisition

3.0 μL of TpPyShell1 protein solution (purified as described above) was applied to a glow-discharged Quantifoil holey carbon grid (R1.2/1.3, Cu, 200 mech), blotted for 3.5 sec at 4°C and plunge-frozen into liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific). The grid was inserted into a Titan Krios (Thermo Fisher Scientific) operating at an acceleration voltage of 300 kV and equipped with a Cs corrector (CEOS, GmbH). Images were recorded with a K3 direct electron detector (Gatan) in CDS mode with an energy filter at a slit width of 20 eV. Data were automatically collected using SerialEM software[93] at a physical pixel size of 0.87 Å, with 52 frames at a dose of 0.96 $e^{-}/Å^{2}$ per frame, an exposure time of 2.63 sec per movie, and defocus ranging from -0.5 to -1.7 μm. A total of 5,951 movies were collected.

Cryo-EM image processing and model building

The movie frames were subjected to beam-induced motion correction using MotionCorr2.1,[85] and the contrast transfer function (CTF) was evaluated using Gctf.[79] Motion correction and CTF estimation were processed using RELION 3.1.[73] The motion corrected micrographs were imported into cryoSPARC ver.4.0.2,[78] and approximately 500 particles were manually selected from 10 micrographs to perform two-dimensional (2D) classification. Using a good 2D class average image as a template, a total of 800,778 particle images were automatically picked from all micrographs in a filament tracer job and were extracted with a box size of 150 pixels with 4x binning. After two rounds of 2D classification, 299,145 particles were selected and extracted with a box size of 600 pixels. The re-extracted particles were subjected to refinement without helical parameters. Even after refinement, symmetry search for the helical parameters were not convincing. Therefore, preliminarily modelled structures were manually fitted into an EM density map to estimate two helical parameters, and the helical parameters were determined to be rise of 25 Å and twist of -30 degrees. Based on these helical parameters, 444,736 particle images were automatically picked from all micrographs using Topaz[80] and were extracted with a box size of 170 pixels with 4x binning using RELION 4.0.[77] After 2D classification (Figures S5B and S5C), selected particles were sorted into 4 classes by 3D classification (Figure S5D). A total of 322,887 particles were re-extracted at a pixel size of 1.09 Å and subjected to five rounds of helical refinement, three rounds of CTF refinement, and Bayesian polishing. The 3D refinement and post-processing yielded a map with global resolution of 2.6 Å, according to Fourier shell correlation (FSC) with the 0.143 criterion. The refined values of helical rise and twist were 25.13 Å and -32.47 degrees, respectively. From the 3D refinement, the dimer of TpPyShell1 proteins was found to be the asymmetric unit; the helical parameters of the dimer were determined to be rise of 4 Å and twist of -56 degrees. The final 3D refinement and post-processing yielded a map with global resolution of 2.4 Å, according to Fourier shell correlation (FSC) with the 0.143 criterion (Figure S5E). The final refined values of helical rise and twist for the dimer were 3.59 Å and -56.07 degrees, respectively. Local resolution was estimated using RELION 4.0 (Figure S5F). The processing workflow is outlined in Figure S5D.

The model of TpPyShell1 from Trp69 to Phe321 (excluding the chloroplast targeting sequence) (Figures 3B and S5G) was built starting from the predicted AlphaFold2 model.[81] After manually fitting this predicted model into the EM density map using UCSF Chimera,[91] each domain was manually remodeled and refined iteratively using COOT,[82] Phenix,[83] and the Servalcat pipeline in REFMAC5.[84] In one monomer of the dimer model, the N-terminus modeling was extended by 5 residues, revealing the shape of a putative short α-helix, but the side chains of these residues could not be assigned. All figures were prepared using UCSF ChimeraX.[92] The statics of the 3D reconstruction and model refinement are summarized in Table S4.

Cryo-ET sample prep and data acquisition

T. pseudonana and *P. tricornutum* were grown under normal atmosphere in F/2 artificial seawater at 18°C and 40 μmol photons m^{-2} s^{-1} light without shaking. For the *T. pseudonana m1* and *m2* mutants (ΔTpPyShell1/2), 5 mM Na_2CO_3 was supplemented in the medium. Cells were sedimented at 800 × *g* for 5 min prior to vitrification, 4 μL of cell suspension was applied on 200-mesh R1/4 SiO_2-film covered gold grids or 200-mesh R2/1 carbon-film covered copper grids (Quantifoil Micro Tools) (for *P. tricornutum* and *T. pseudonana* cells, respectively) and plunge frozen using a Vitrobot Mark IV (Thermo Fisher Scientific). EM grids were clipped into Autogrid supports (Thermo Fisher Scientific) and loaded into Aquilos 1 or 2 FIB-SEM instruments (Thermo Fisher Scientific), where they were thinned with a Gallium ion beam as previously described.[38] The resulting EM grids with thin lamellae were transferred to a transmission electron microscope for tomographic imaging.

For *P. tricornutum*, cryo-ET data was acquired on a 300 kV Titan Krios G2 microscope (Thermo Fisher Scientific), equipped with a post-column energy filter (Quantum, Gatan) and a direct detector camera (K2 summit, Gatan) (‘‘M1’’), as well as on 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’’). For *T. pseudonana*, data was acquired on microscope M2, as well as on a 300 kV Titan Krios G4i microscope (Thermo Fisher Scientific), equipped with a post-column energy filter (Selectris X, Thermo Fisher Scientific) and a direct electron detector (Falcon 4, Thermo Fisher Scientific) (‘‘M3’’).

For microscopes M1 and M2, tilt-series were obtained using SerialEM 3.8 software.[93] In all cases, tilt-series were acquired using a dose-symmetric tilt scheme,[107] with 2° steps totaling 60 tilts per series. Each image was recorded in counting mode with ten frames per second. The target defocus of individual tilt-series ranged from -2 to -5 μm. Total dose per tilt series was approximately 120 $e^-/\text{Å}^2$. Image pixel sizes for microscopes M1 and M2 were 3.52 and 2.143 Å, respectively. For microscope M3, tilt-series were obtained using the Tomography 5.11 software (Thermo Fisher Scientific), using the same acquisition scheme as above, except for the use of multi-shot acquisition. Data was acquired in EER mode with a calibrated image pixel size of 2.93 Å.

In total, we acquired 21 tomograms for WT *P. tricornutum*, 40 tomograms for WT *T. pseudonana*, 8 tomograms for mutant *m1*, and 7 tomograms for mutant *m2*. Each tomogram was acquired from a different chloroplast. All cells were in asynchronous culture, and multiple cultures were imaged for each strain.

Cryo-ET data analysis

TOMOMAN Matlab scripts (version 0.6.9)[88] were used to preprocess the tomographic tilt series data. Raw frames were aligned using MotionCor2 (version 1.5.0),[85] then tilt-series were dose-weighted[108] followed by manual removal of bad tilts. The resulting tilt-series (binned 4 times, pixel sizes: 14.08 Å for M1, 8.57 Å for M2, 11.6 Å for M3) were aligned in IMOD (version 4.11)[109] using patch tracking and were reconstructed by weighted back projection. Cryo-CARE (version 0.2.1)[90] 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.[86] The resulting segmentations were manually curated in Amira (version 2021.2).

Subtomogram averaging of *T. pseudonana* PyShell

For subtomogram averaging, only data from microscope M2 was used. Segmented surfaces corresponding to the PyShell were used as input to determine initial normal vectors in MemBrain’s point and normal sampling module.[110] Vectors were sampled densely on the surface with a spacing of 1.5 voxels at bin4 (8.572 Å/px). The resulting positions were used as initial coordinates to extract subvolumes (box size of 32 pixels) from bin4 tomograms corrected for the contrast transfer function (CTF) using phase flipping in IMOD. Starting from the normal vectors determined in MemBrain, multiple rounds of subtomogram alignment, averaging and classification were carried out in STOPGAP software.[89] False positives and poorly aligning particles were removed by classification steps. A second round of extraction, alignment and classification was performed at bin2 (4.286 Å/px), starting with the coordinates from the previous round of averaging in bin4. During particle alignment a maximum resolution of 16 Å was allowed to prevent overfitting. For resolution estimation and map filtering, the half-maps from the final round of alignment were postprocessed in RELION using a soft disk-shaped mask.

Comparison of density maps from *in vitro* SPA and *in situ* STA

The 2.4 Å-resolution SPA density map of TpPyShell1 was unrolled from a tube to a flat sheet using the unroll command in ChimeraX.[92] Different inner radii for the unroll operation were tested with the aim to match the lateral periodicity measured in the 2D class average of TpPyShell1 flat sheets (Figures S6B–S6D). The flattened maps obtained using the measured inner tube radius and the measured sheet periodicity were then rigid-body fit into the STA density and resampled on the grid of the STA map to place them at the same box and pixel size. The same disk-shaped soft mask used for postprocessing the STA map was then applied to both unrolled, resampled versions of the SPA map. Finally, these two maps had their power spectrum matched to that of the STA map using the relion_image_handler program. The correlation scores between the matched SPA and STA maps were measured at each 10° in-plane rotation angle in ChimeraX; this was performed for both versions of the unrolled SPA map. For Figure 3G, the unrolled 2.4 Å-resolution SPA density map was fit with models of individual TpPyShell1 monomers (using ChimeraX rigid body fitting) to obtain a model of a flat TpPyShell1 lattice.

Photosynthesis measurements

T. pseudonana WT, *m1*, and *m2* cells were cultured under LC and HC (as described above in ‘‘Diatom Cultures’’) with 40 μmol photons m^{-2} s^{-1} light, harvested at the logarithmic growth phase, and resuspended in freshly prepared DIC-free F/2 artificial seawater. Chlorophyll a concentration of the samples was determined in 100% (v/v) methanol,[111] and the cell samples were applied to an oxygen electrode (Hansatech, King’s Lynn, U.K.) at 10 μg chlorophyll a mL^{-1} in the DIC-free F/2 artificial seawater (pH 8.1). Simultaneous measurement of net O_2 evolution rate with total DIC concentration in the sample mixture was achieved by an oxygen electrode and a gas-chromatography flame ionization detector (GC-8A, Shimadzu, Kyoto, Japan) during stepwise addition of $NaHCO_3$, as previously reported.[29] Measurements used constant actinic light of 900 μmol photons m^{-2} s^{-1}. The photosynthetic parameters were calculated from the plot of O_2 evolution rate against DIC concentration by curve fitting with the non-linear least squares method: P_{max}, maximum net O_2 evolution rate; $K_{0.5}$, DIC concentration giving half of P_{max}; $[DIC]_{comp}$, [DIC] giving no net O_2 evolution; and APC, apparent photosynthetic conductance. See Table S2.

FIB-SEM data acquisition and analysis

Cells were grown under the same conditions as for the cryo-ET analysis. Sample preparation was performed as in.[9] FIB-SEM tomography was performed with a Zeiss CrossBeam 550 microscope, equipped with Fibics Atlas 3D software for tomography. The voxel size was 16x16x16 nm for the WT, 6x6x6 nm for the *m1* mutant, and 8x8x8 nm for the *m2* mutant. The whole volumes were imaged with an average of 300 frames for WT and 1000 frames for the mutants. Single cells were isolated by cropping in 3D using the open software Fiji.[94] Image misalignment was corrected using the "StackReg" plugin in Fiji. We used 3D Slicer[99] for segmentation and 3D reconstruction, and Meshlab[98] to reduce noise and enhance contours of reconstructed objects. The quantitative measurements of chloroplasts and pyrenoids organelles (volume, diameter, sphericity) were implemented in python using libraries including trimesh, stl and scikit-image.

Phylogenetic analysis

Homologs of *T. pseudonana* TpPyShell1 were retrieved from the National Centre for Biotechnology Information and the Marine Microbial Eukaryote Transcriptome Sequence Project (MMETSP).[112] The highest scoring sequences per species were selected (E-value cutoff = 1e−35). Gaps and non-conserved regions were removed, and the protein sequences were subsequently aligned using Clustal Omega.[95] The alignment was used to generate a maximum likelihood tree, using IQTREE with standard settings and visualized with iTOL.[96,97] Taxonomic distribution of the TpPyShell1 protein sequence in the ocean was queried against the EUK_SMAGs dataset[63] in the Ocean Gene Atlas v2.0 webserver (https://tara-oceans.mio.osupytheas.fr).[64,113]

QUANTIFICATION AND STATISTICAL ANALYSIS

LC-MS/MS data were quantified with Proteome Discoverer 1.4 (Thermo Fisher Scientific) (Table S1). For qRT-PCR experiments, $\Delta\Delta$Ct values[106] were calculated as described in the STAR Methods section above. The resolution of cryo-EM SPA maps and cryo-ET STA maps were estimated by “gold-standard” Fourier shell correlation (FSC), using independent half-maps and the 0.143 criterion[114] (Figures S4B and S5E). All statistics of the atomic model built into the cryo-EM map were analyzed using Phenix[83,115] (Table S4). Correlation scoring between the SPA and STA maps (Figure S6G) was performed with ChimeraX.[92] Quantitative measurements of FIB-SEM data were implemented in python, with scripts deposited on GitHub and Zenodo (key resources table); the number of chloroplasts used for each calculation (*n*) is detailed in the legend of Figures 4G–4J. The number of acquired cryo-EM movies and cryo-ET tilt-series are noted in the relevant STAR Methods sections. Photosynthetic measurements and growth curves (Figures 4B–4D; Table S2) were quantified using three independent biological replicates.

Supplemental figures

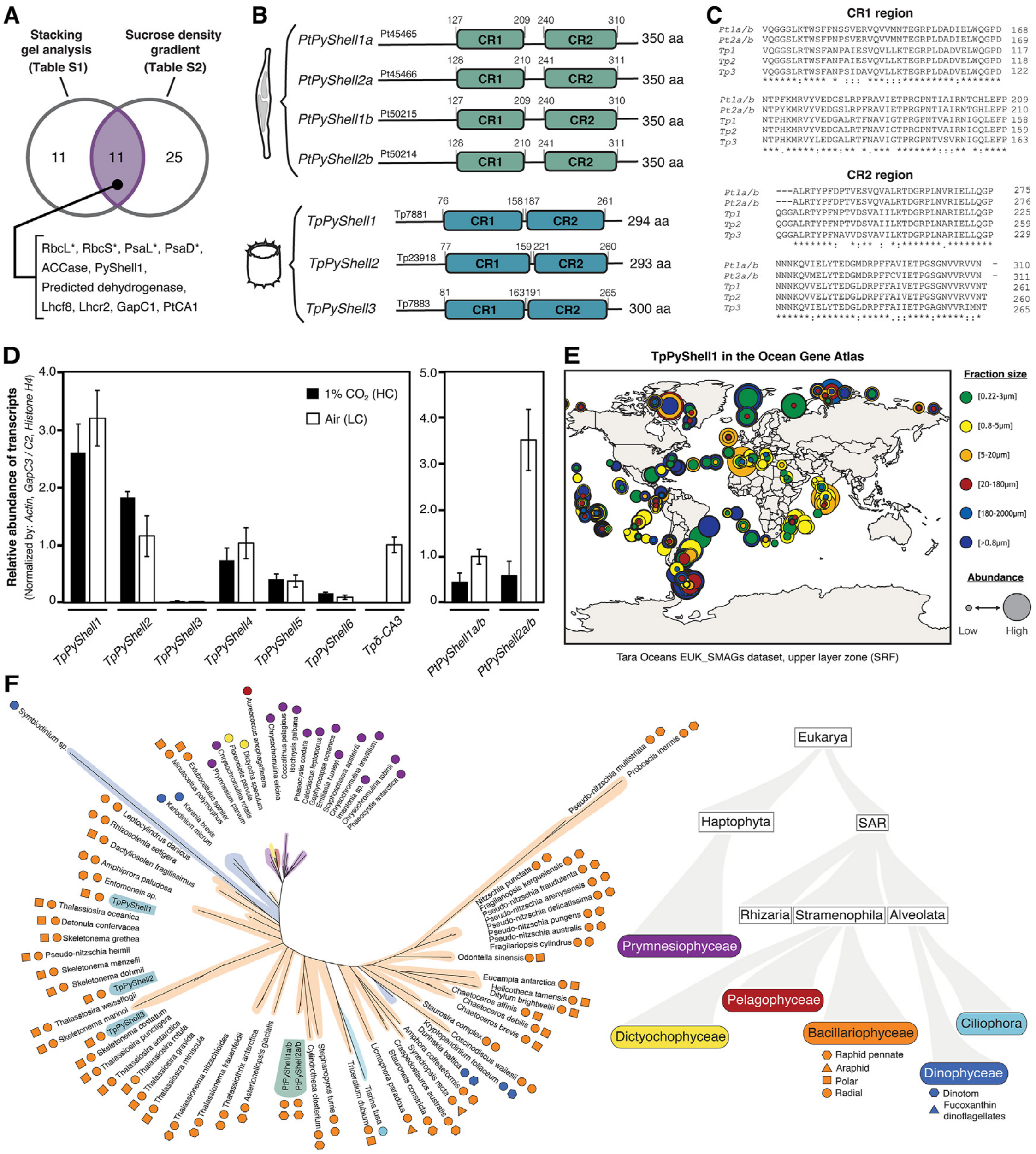


Figure S1. PyShell genes, expression, and phylogeny, related to Figures 1A–1C

(A) Venn diagram showing the number of potential pyrenoid proteins identified by LC-MS/MS analysis of the stacking gel (“procedure A,” listed in Table S1A) and of the sucrose density gradient fractions (“procedure B,” listed in Table S1B). The 11 proteins in the purple intersection were identified by both procedures. Proteins encoded by chloroplast genes are indicated by an asterisk.

(B) Domain architecture of the PyShell genes in *P. tricornutum* and *T. pseudonana*. The conserved regions CR1 and CR2 are indicated with colored bars.

(C) Sequence alignments of CR1 and CR2 from the PyShell genes listed in (B).

(D) Quantitative PCR (qPCR) analysis of PyShell gene expression. Left: transcript levels of six PyShell family genes identified in *T. pseudonana* (TpPyShell1–6) under both normal atmosphere (LC, 0.04% CO_2) and high CO_2 (HC, 1%), compared with the expression of the LC-induced carbonic anhydrase δ-CA3. Right: transcript levels of PtPyShell1a/b and 2a/b in *P. tricornutum* under LC and HC. Error bars: standard deviation.

(legend continued on next page)

(E) Global distribution of TpPyShell1 homologous sequences in fractions from *Tara* Oceans sampling,[116] identified in the Ocean Gene Atlas v2.0[64] by searching the EUK_SMAGs dataset,[63] which contains over 700 eukaryotic environmental genomes of diverse lineages (not just diatoms), built from 280 billion metagenomic reads from sunlit oceans in polar, temperate, and tropical regions.

(F) Maximum likelihood unrooted gene tree of TpPyShell1 (left), constructed with IQ-TREE, and an algal phylogenetic tree (right). The color in the phylogenetic tree indicates the clade to which each species in the gene tree belongs. Shapes and colors in the gene tree correspond to clades in the phylogenetic tree. The PyShell genes of *T. pseudonana* and *P. tricornutum* described in this study (TpPyshell1, 2, 3; PtPyShell1a/1b, 2a/2b) are highlighted in blue and teal, respectively.

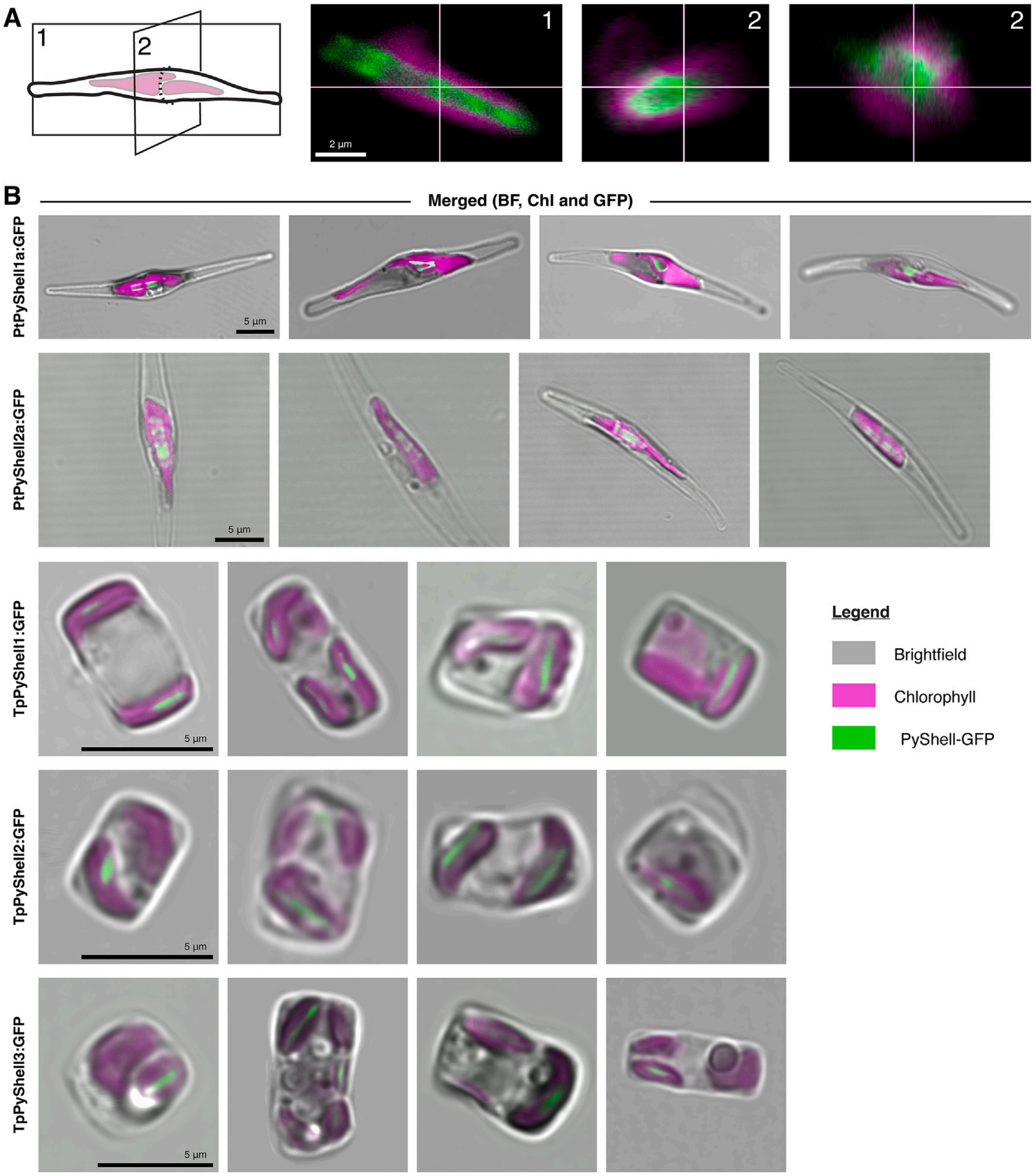


Figure S2. Additional fluorescence images of GFP-tagged PyShell proteins in *P. tricornutum* and *T. pseudonana*, related to Figure 1D

(A) 3D confocal images of PtPyShell1a:GFP in *P. tricornutum*. Diagram on the left shows the orientations of the virtual slices displayed on the right. Scale bar: 2 µm.

(B) Maximum intensity projections of PtPyShell1a:GFP and PtPyShell2a:GFP in *P. tricornutum*, as well as TpPyShell1:GFP, TpPyShell2:GFP, and TpPyShell3:GFP in *T. pseudonana*. Gray: bright field (BF), magenta: chlorophyll autofluorescence (Chl), green: GFP. Scale bars: 5 µm.

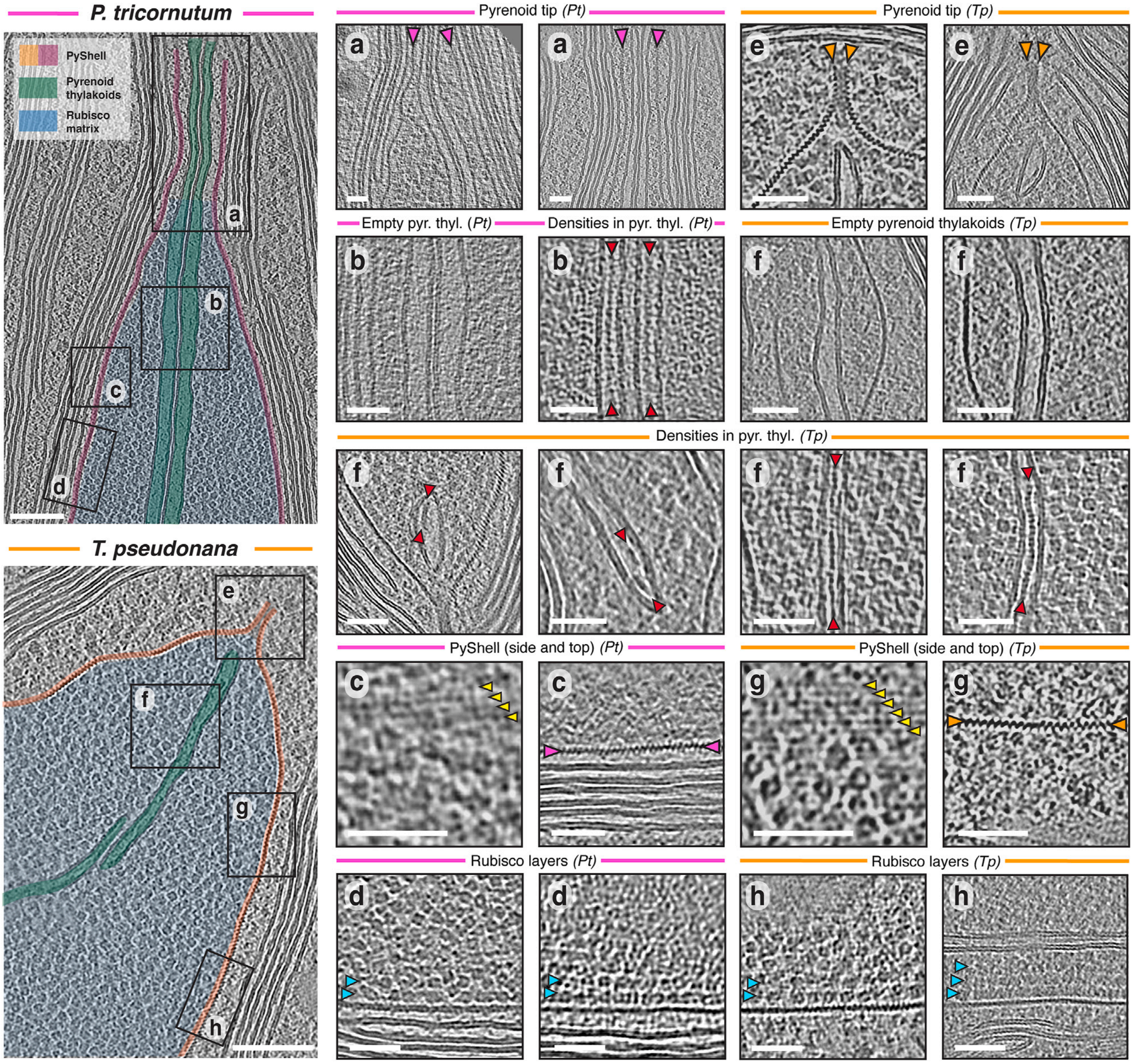


Figure S3. Additional cryo-ET images of pyrenoids inside native *P. tricornutum* and *T. pseudonana* cells, related to Figures 2A–2I

(A–H) Left panels show pyrenoid overviews of *P. tricornutum* (pink) and *T. pseudonana* (orange), with lettered boxes indicating corresponding pyrenoid regions detailed in panels to the right. For *P. tricornutum* and *T. pseudonana*, respectively: (A and E) examples of the pyrenoid ends (tips), which differ between species (PyShell: orange and pink arrowheads); (B and F) pyrenoid-traversing thylakoids (pyr. thyl.), which sometimes have dense particles in the lumen (red arrowheads); (C and G) PyShell surface (side) and cross-section (top) views. Yellow arrowheads: major stripes of the PyShell lattice; (D and H) ordered layers of Rubisco (blue arrowheads) adjacent to the PyShell. Scale bars: 100 nm in overviews, 50 nm in all others.

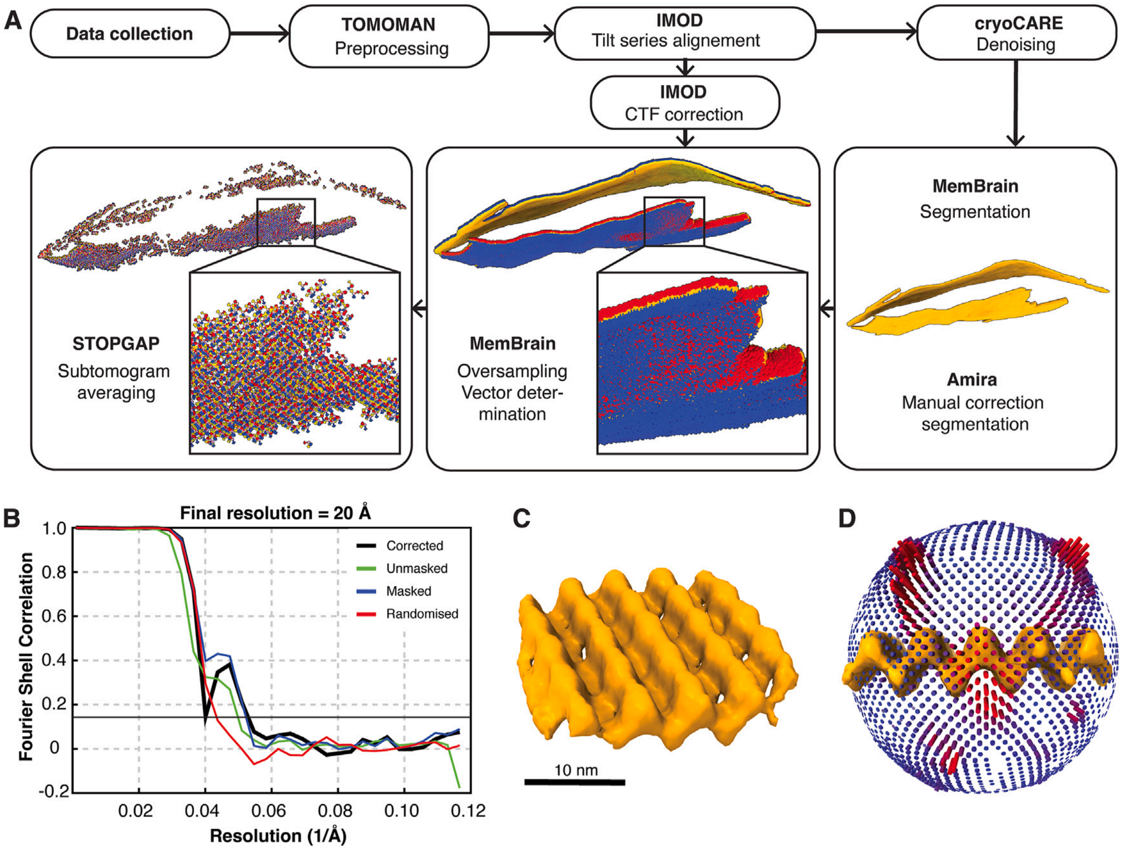


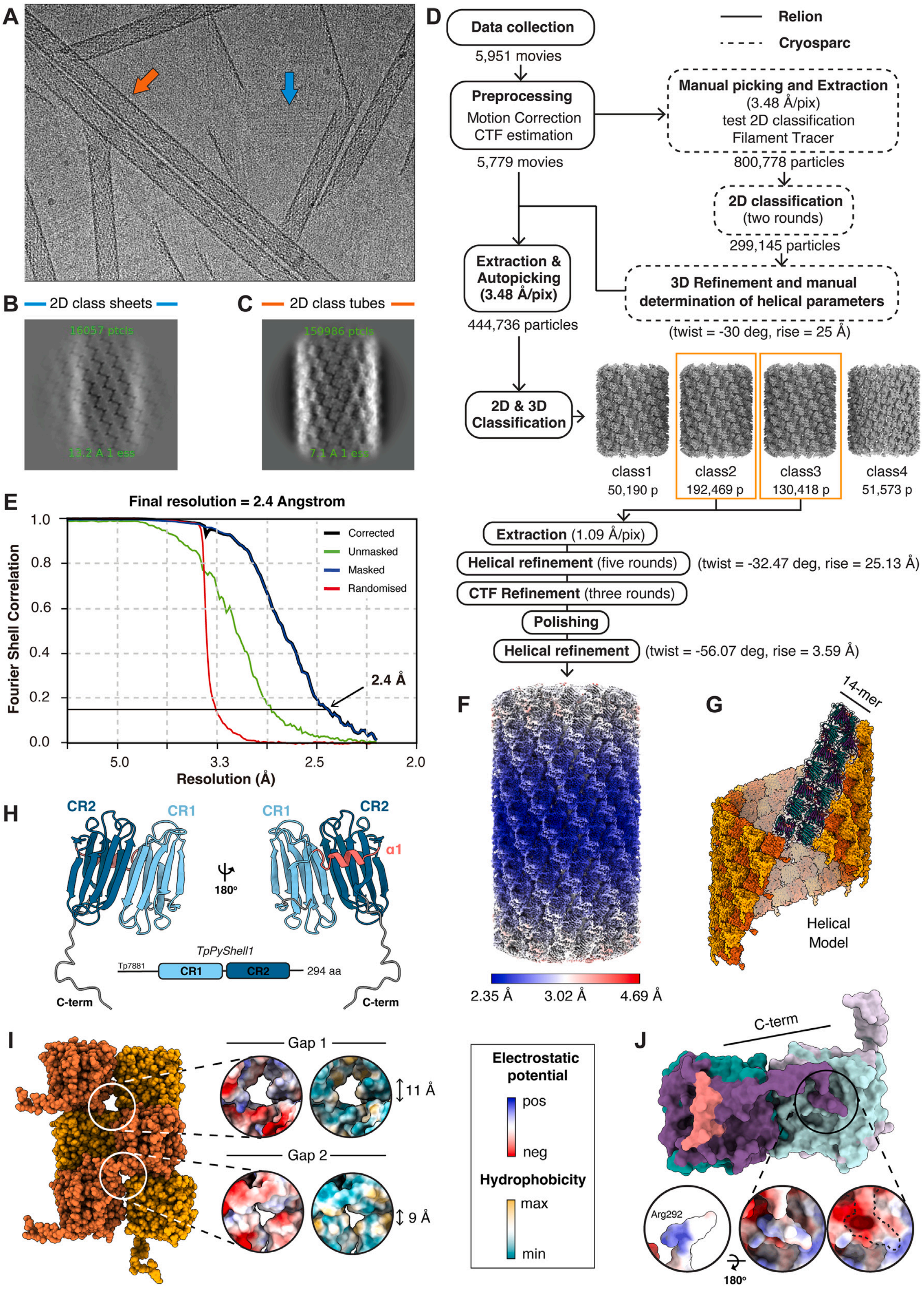
Figure S4. Cryo-ET subtomogram averaging supplement, related to Figure 2J

(A) Cryo-ET data processing workflow. Determination of initial and final coordinates and vectors (blue/red/yellow arrows) is shown for an example tomogram. Coordinates were initially oversampled along the PyShell segmentation in MemBrain.[110] After subtomogram averaging in STOPGAP,[89] the coordinates converged to the repeat of the PyShell lattice. Only one subvolume per coordinate was retained in the final average.

(B) Fourier shell correlation (FSC) resolution determination of the resulting STA map, using the 0.143 cutoff.

(C) Inclined view of the *in situ* PyShell STA density map. Scale bar: 10 nm.

(D) Angular distribution of particles contributing to the STA map. Red: more populated orientations, blue: less populated orientations.



(legend on next page)

Figure S5. Single-particle cryo-EM supplement, related to Figures 3A–3E

(A) Example micrograph. Sheets and tubes are indicated with blue and orange arrows, respectively.

(B) 2D class average obtained for the TpPyShell1 sheets.

(C) 2D class average obtained for the TpPyShell1 tubes. The homodimers of PyShell proteins can be seen in both the sheet and tube assemblies.

(D) Cryo-EM single-particle analysis (SPA) processing workflow for high-resolution helical reconstruction of the TpPyShell1 tubes, displaying the four classes resulting from 2D and 3D classification. The initial symmetric unit consisted of seven homodimers (14-mer displayed in G), with the helical parameters of twist = -32.47° and rise = 25.13 Å. After polishing the map, the new symmetric unit contained a homodimer with the helical parameters of twist = -56.07° and rise = 3.59 Å.

(E) FSC resolution determination of the resulting SPA map, using the 0.143 cutoff.

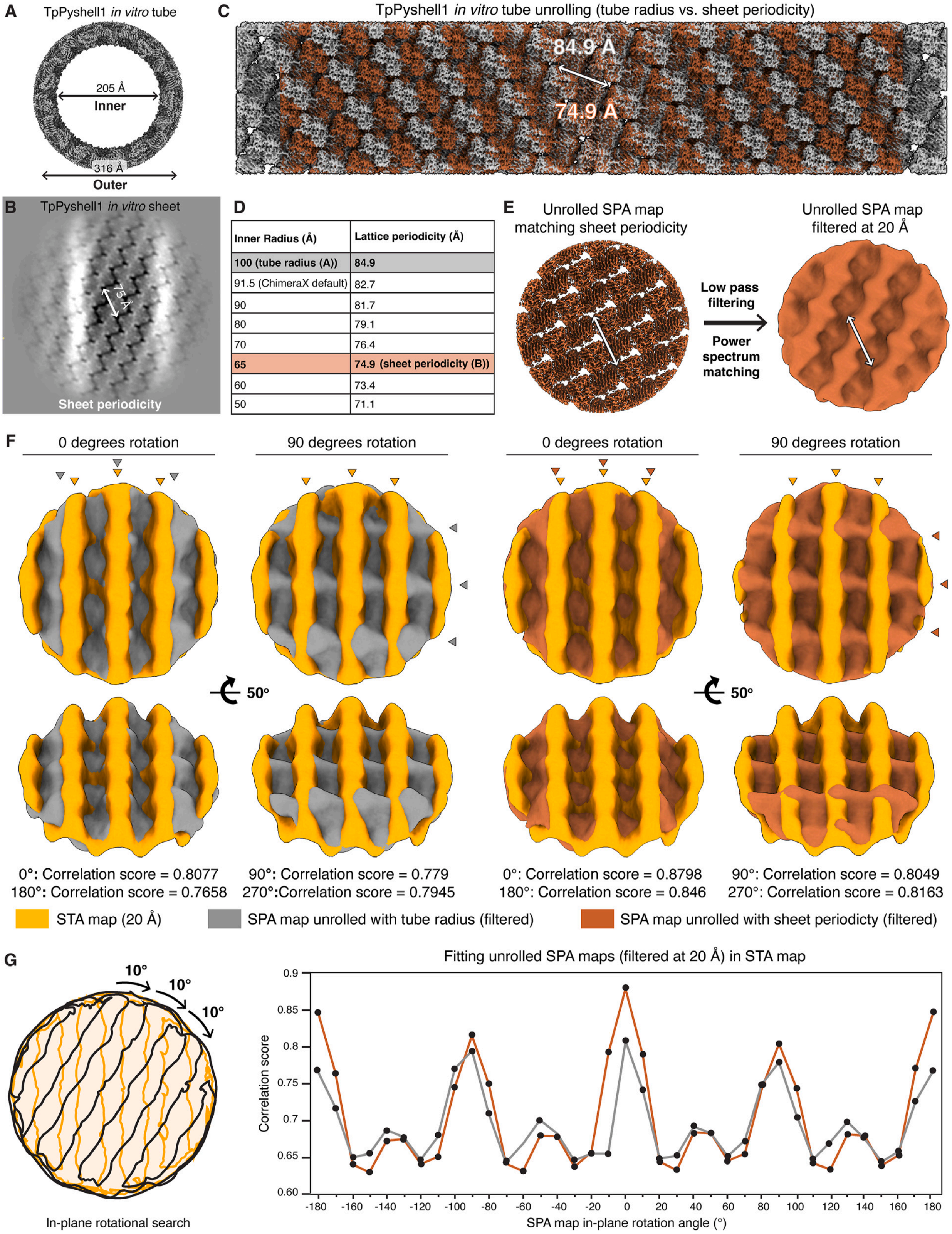
(F) The SPA cryo-EM density map displayed with local resolution.

(G) Resulting helical protein model (orange and yellow surface representation) that was initially built using a 14-mer repeating unit (purple and teal cartoon representation), and later refined using a dimeric repeating unit.

(H) The internal pseudo-2-fold symmetry the TpPyShell1 monomer, subdividing the β-sheets between the β-5/-6 and β-13/-14 strands. Conserved regions CR1 and CR2 are colored (shades of blue), along with the short α-helix connecting CR1 with CR2 (pink).

(I) Two potential gaps are present within the TpPyShell1 lattice. Orange and yellow surface representations correspond to the inward and outward facing monomers of the homodimer repeating unit. Circle insets: electrostatic potential and hydrophobicity are displayed on surface representations.

(J) The C-terminal domain (C-term) of each PyShell monomer reaches toward a putative pocket on the neighboring PyShell monomer. Above: colors on surface representations correspond to Figure 3E. One monomer is displayed in lighter colors. Circle insets: electrostatic potential is displayed on surface representations. Left: C-term shown alone, with positively charged Arg292 residue (blue), Middle: C-term shown attached to neighbor subunit, Right: neighbor subunit with this bound C-term removed (indicated with outline), revealing a negatively charged pocket (red).



Inner Radius (Å)	Lattice periodicity (Å)
100 (tube radius (A))	84.9
91.5 (ChimeraX default)	82.7
90	81.7
80	79.1
70	76.4
65	74.9 (sheet periodicity (B))
60	73.4
50	71.1

(legend on next page)

Figure S6. Comparison of the *in vitro* SPA map with the *in situ* STA map, related to Figures 3F and 3G

(A) SPA 3D map of *in vitro* TpPyShell1 assembled in a tube (Figures 3A and S5F). Shown in cross-section view, with measured inner and outer radius indicated.

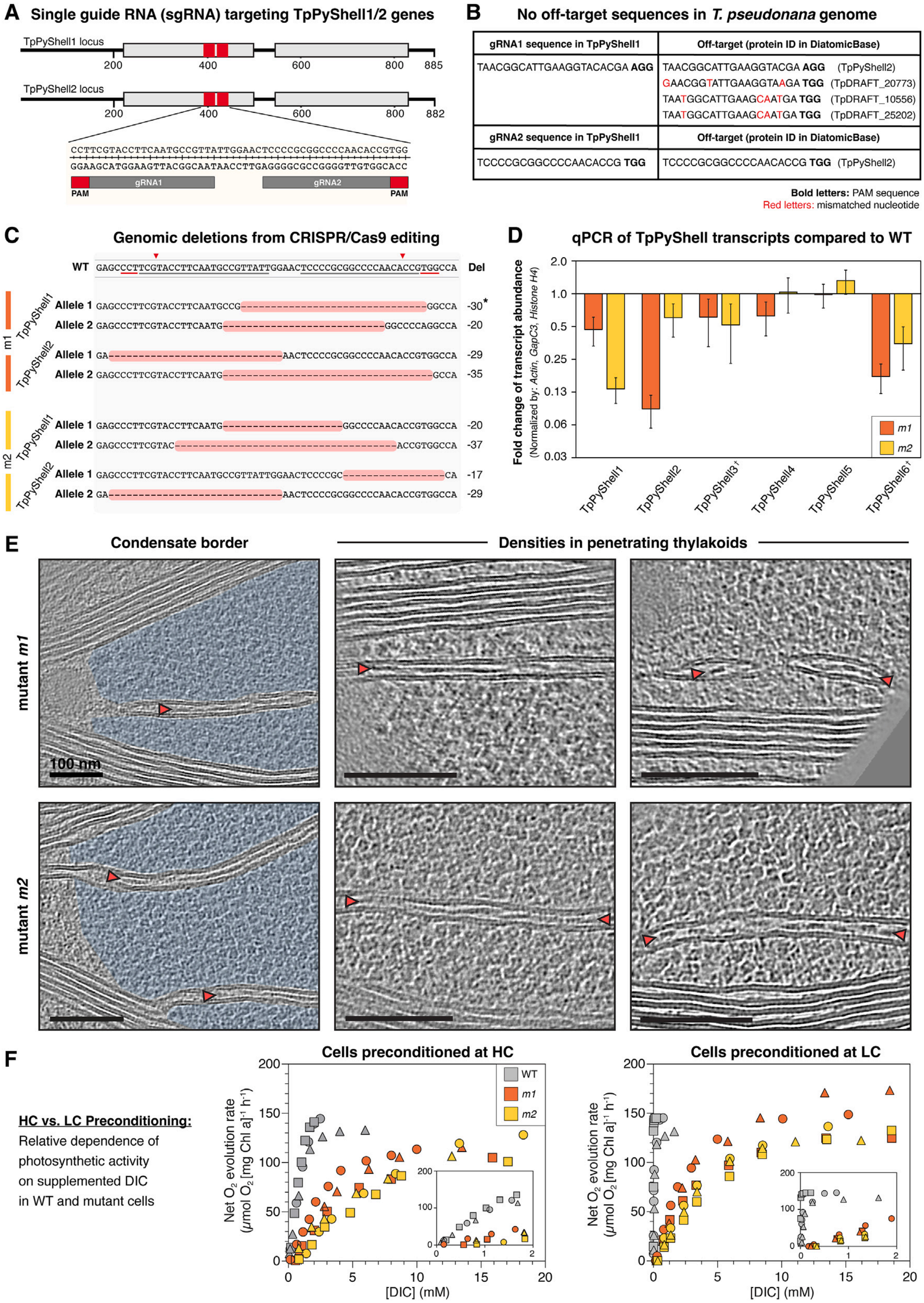
(B) SPA 2D class average of *in vitro* TpPyShell1 assembled in a flat sheet (Figure S5B). Shown in surface view, with measured lateral repeat indicated.

(C) SPA 3D map unrolled with ChimeraX[92] at full resolution, using an unrolling inner radius of 100 Å (gray, approximate inner radius of the TpPyShell1 tube) or 65 Å (red, resulting in a lateral periodicity matching the TpPyShell1 flat sheet). Lateral periodicities of the two unrolled maps are indicated.

(D) Table showing the measured lateral periodicities of unrolling the SPA tube map using different inner radii. The two radii selected for further analysis are indicated in gray and red.

(E) The unrolled SPA maps were filtered to 20 Å resolution by matching to the power spectrum of the *in situ* STA map. Isosurface threshold here and in subsequent panels = 1 standard deviation.

(F and G) In-plane rotational searches were performed in ChimeraX to compare the unrolled and filtered SPA maps (gray: unrolled with a 100-Å radius, red: unrolled with a 65-Å radius) to the postprocessed 20-Å resolution STA map. All maps were masked with a soft disk-shaped mask to facilitate the rotation and correlation operations in an unbiased manner. The initial 0° orientation was determined by a rigid-body fitting in ChimeraX. (F) Views of overlaying STA and SPA filtered maps at 0° and 90° offset, with correlation scores reported below. (G) Left: diagram of the rotational search. Right: correlation scores for the full 360° search of both filtered SPA maps against the STA map. The best scoring fit was obtained using the SPA density map unrolled with the measured sheet periodicity (65 Å radius) and 0° in-plane rotation relative to the STA map.



gRNA1 sequence in TpPyShell1	Off-target (protein ID in DiatomicBase)
TAACGGCATTGAAGGTACACGA **AGG**	TAACGGCATTGAAGGTACGA **AGG** (TpPyShell2) GAACGGTATTGAAGGTAAGA **TGG** (TpDRAFT_20773) TAATGGCATTGAAGCAATGA **TGG** (TpDRAFT_10556) TAATGGCATTGAAGCAATGA **TGG** (TpDRAFT_25202)
gRNA2 sequence in TpPyShell1	**Off-target (protein ID in DiatomicBase)**
TCCCCGCGGCCCCAACACCG **TGG**	TCCCCGCGGCCCCAACACCG **TGG** (TpPyShell2)

Bold letters: PAM sequence
Red letters: mismatched nucleotide

(legend on next page)

Figure S7. Generation of ΔTpPyShell1/2 mutants, O_2 evolution, and additional cryo-ET, related to Figures 4A, 4D, and 4O–4R

(A) Schematic representation of simultaneous CRISPR-Cas9 targeting in both TpPyShell loci. PAM, protospacer adjacent motif.

(B) Prediction of potential off-target sites of the sgRNAs in the *T. pseudonana* genome.

(C) Resulting genomic deletions in the different TpPyShell1/2 alleles. Asterisk: this is an in-frame deletion; however, no protein product was detected in the mutant (Figure 4A).

(D) qPCR analysis of the six PyShell family genes identified in *T. pseudonana*, showing relative expression of transcripts in the mutant strains *m1* and *m2* compared with expression in WT cells. Error bars: standard deviation. Cross: expression of TpPyShell3 and TpPyShell6 are already low in the WT background (Figure S1D).

(E) Additional cryo-ET of Rubisco condensates in *m1* and *m2* cells. Light blue: Rubisco matrix; red arrowheads: densities in thylakoid lumen. Scale bars: 100 nm.

(F) Dependence of photosynthetic activity (measured by O_2 evolution under 900 μmol photons m^{-2} s^{-1} constant actinic light) on DIC concentration (set by supplementing with bicarbonate) in WT cells (gray), *m1* (orange), and *m2* (yellow). Different symbols: three independent experiments. Cells were either preconditioned in HC or LC conditions; the WT cells preconditioned in LC have a more robust O_2 evolution response in low DIC concentrations, likely due to full activation of their CCM (compare inset panels).