

Structural basis for molecular assembly of fucoxanthin chlorophyll *a/c*-binding proteins in a diatom photosystem I supercomplex

Reviewed Preprint

v2 • October 9, 2024

Revised by authors

Reviewed Preprint

v1 • August 20, 2024

Koji Kato, Yoshiki Nakajima, Jian Xing, Minoru Kumazawa, Haruya Ogawa, Jian-Ren Shen ,
Kentaro Ifuku , Ryo Nagao 

Research Institute for Interdisciplinary Science and Graduate School of Environmental, Life, Natural Science and Technology, Okayama University, Okayama, Japan • Graduate School of Agriculture, Kyoto University, Kyoto, Japan
• Faculty of Agriculture, Shizuoka University, Shizuoka, Japan

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

Abstract

Photosynthetic organisms exhibit remarkable diversity in their light-harvesting complexes (LHCs). LHCs are associated with photosystem I (PSI), forming a PSI-LHCI supercomplex. The number of LHCI subunits, along with their protein sequences and pigment compositions, has been found to differ greatly among the PSI-LHCI structures. However, the mechanisms by which LHCs recognize their specific binding sites within the PSI core remain unclear. In this study, we determined the cryo-electron microscopy structure of a PSI supercomplex incorporating fucoxanthin chlorophyll *a/c*-binding proteins (FCPs), designated as PSI-FCPI, isolated from the diatom *Thalassiosira pseudonana* CCMP1335. Structural analysis of PSI-FCPI revealed five FCPI subunits associated with a PSI monomer; these subunits were identified as RedCAP, Lhcr3, Lhcq10, Lhcf10, and Lhcq8. Through structural and sequence analyses, we identified specific protein-protein interactions at the interfaces between FCPI and PSI subunits, as well as among FCPI subunits themselves. Comparative structural analyses of PSI-FCPI supercomplexes, combined with phylogenetic analysis of FCPs from *T. pseudonana* and the diatom *Chaetoceros gracilis*, underscore the evolutionary conservation of protein motifs crucial for the selective binding of individual FCPI subunits. These findings provide significant insights into the molecular mechanisms underlying the assembly and selective binding of FCPIs in diatoms.

eLife Assessment

This **important** study presents a high-resolution cryoEM structure of the supercomplex between photosystem I (PSI) and fucoxanthin chlorophyll a/c-binding proteins (FCPs) from the model diatom *Thalassiosira pseudonana* CCMP1335, revealing subunits, protein:protein interactions and pigments not previously seen in other diatoms or red/green photosynthetic lineages. Combining structural, sequence and phylogenetic analyses, the authors provide **convincing** evidence of conserved motifs crucial for the binding of FCPs, accompanied by interesting speculation about the mechanisms governing the assembly of PSI-FCP supercomplexes in diatoms and their implications for related PSI-LHC supercomplexes in plants. The findings set the stage for functional experiments that will further advance the fields of photosynthesis, bioenergy, ocean biogeochemistry and evolutionary relationships between photosynthetic organisms.

<https://doi.org/10.7554/eLife.99858.2.sa2>

Introduction

Oxygenic photosynthesis in cyanobacteria, algae, and land plants converts solar energy into chemical energy and releases molecular oxygen into the atmosphere (1). The conversion of light energy takes place within two multi-subunit membrane protein complexes, known as photosystem I (PSI) and photosystem II (PSII), which perform light harvesting, charge separation, and electron transfer reactions (2–5). To optimize light energy capture, numerous light-harvesting antenna subunits are associated with the periphery of the PSI and PSII core complexes, transferring excitation energy to the respective photosystem cores (1). These light-harvesting antennae exhibit significant diversity among photosynthetic organisms, both in protein sequences and pigment compositions, and can be broadly categorized into two major groups: membrane proteins and water-soluble proteins (1).

The membrane protein category primarily consists of the light-harvesting complex (LHC) protein superfamily (6, 7), which absorbs light energy through chlorophylls (Chls) and carotenoids (Cars). The number and types of Chls and Cars vary significantly among LHCs, which can be grouped into green and red lineages, leading to color diversity in photosynthetic organisms (8). The green lineage includes green algae and land plants, while the red lineage encompasses red algae, diatoms, haptophytes, cryptophytes, and dinoflagellates (8). LHCs specific to PSI (LHCIs) bind to a eukaryotic PSI monomer, forming a PSI-LHCI supercomplex (9, 10), the structures of which have been revealed by cryo-electron microscopy (cryo-EM) in various eukaryotes (9, 10). In the red lineage, the number of LHCIs and their protein sequences and pigment compositions exhibit considerable variation among the PSI-LHCI structures of red algae (11–14), a diatom (15, 16), a cryptophyte (17), and dinoflagellates (18, 19).

Recently, we demonstrated the conservation and diversity of LHCIs among red-lineage algae through structural and phylogenetic analyses of PSI-LHCI supercomplexes (14). This study revealed that while the binding sites of LHCIs to PSI were conserved to some extent among red-lineage algae, their evolutionary relationships were weak. It is known that LHCs have similar overall protein structures across photosynthetic organisms, with particular similarity in their three-transmembrane helices, regardless of whether they belong to the green or red lineages (9),

10 [10](#)). However, individual LHCI have altered their sequences and structures to adapt their respective binding sites to the PSI cores during the assembly of PSI-LHCI supercomplexes. These observations raise a critical question: how do LHCI recognize their binding sites in the PSI core?

Diatoms are among the most essential phytoplankton in aquatic environments, playing a crucial role in the global carbon cycle, supporting marine food webs, and contributing significantly to nutrient cycling, thus ensuring the health and sustainability of marine ecosystems (20 [20](#)). Diatoms possess unique LHCI known as fucoxanthin Chl *a/c*-binding proteins (FCPIs), which differ in pigment composition and amino acid sequences from the LHCI of land plants (21 [21](#)–23 [23](#)). Previous studies have reported the isolation and structural characterization of PSI-FCPI supercomplexes isolated from the diatom *Chaetoceros gracilis* (15 [15](#), 16 [16](#), 24 [24](#)–29 [29](#)). Kumazawa et al. showed significant diversity in FCPIs between *C. gracilis* and *Thalassiosira pseudonana*, with 46 and 44 FCPIs identified, respectively (30 [30](#)). These FCPIs are categorized into multiple, closely related subgroups (30 [30](#)), and their amino acid sequences are not entirely identical between the two diatoms. Consequently, comparing FCPIs, including their amino acid residues and protein structures at similar binding sites in PSI-FCPIs, may provide molecular insights into how FCPIs interact with PSI. However, an overall structure of the *T. pseudonana* PSI-FCPI supercomplex has yet to be solved.

In this study, we solved the structure of the PSI-FCPI supercomplex from *T. pseudonana* CCMP1335 at a resolution of 2.30 Å by cryo-EM single-particle analysis. The structure reveals a PSI-monomer core and five FCPI subunits. Structural and sequence comparisons highlight unique protein-protein interactions between each FCPI subunit and PSI. Based on these findings, we discuss the molecular assembly and selective binding mechanisms of FCPI subunits in diatom species.

Results and discussion

Overall structure of the *T. pseudonana* PSI-FCPI supercomplex

The PSI-FCPI supercomplexes were purified from the diatom *T. pseudonana* CCMP1335 and analyzed by biochemical and spectroscopic techniques (Fig. S1). Notably, the protein bands of PSI-FCPI closely resembled those reported in a previous study (31 [31](#)). Cryo-EM images of the PSI-FCPI supercomplex were obtained using a JEOL CRYO ARM 300 electron microscope operated at 300 kV. The final cryo-EM map was determined at a resolution of 2.30 Å with a C1 symmetry (Fig. S2, S3, Table S1), based on the “gold standard” Fourier shell correlation (FSC) = 0.143 criterion (Fig. S3A).

The atomic model of PSI-FCPI was built based on the cryo-EM map obtained (see Methods; Fig. S3, Table S1–S3). The structure reveals a monomeric PSI core associated with five FCPI subunits (Fig. 1A, B [1A](#), B [1A](#)). The five FCPI subunits were named FCPI-1 to 5 (Fig. 1A [1A](#)), following the nomenclature of LHCI subunits in the PSI-LHCI structure of *Cyanidium caldarium* RK-1 (NIES-2137) (14 [14](#)). Specifically, the positions of FCPI-1 and FCPI-2 in the *T. pseudonana* PSI-FCPI structure (Fig. 1A [1A](#)) correspond to those of LHCI-1 and LHCI-2 in the *C. caldarium* PSI-LHCI structure. The PSI core comprises 94 Chls *a*, 18 β-carotenes (BCRs), 1 zeaxanthin (ZXT), 3 [4Fe-4S] clusters, 2 phylloquinones, and 6 lipid molecules, whereas the five FCPI subunits include 45 Chls *a*, 7 Chls *c*, 2 BCRs, 15 fucoxanthins (Fxs), 7 diadinoxanthins (Ddxs), and 3 lipid molecules (Table S3).

Structure of the *T. pseudonana* PSI core

The PSI core contains 12 subunits, 11 of which are identified as PsaA, PsaB, PsaC, PsaD, PsaE, PsaF, PsaI, PsaJ, PsaL, PsaM, and Psa29 (Fig. 1B [1B](#)). The remaining subunit could not be assigned due to insufficient map resolution and was therefore modeled as polyanilines (Fig. S4A). This unidentified subunit, designated as Unknown, occupies the same site as Psa28 in the *C. gracilis* PSI-FCPI (15 [15](#)). The structural comparison reveals that Unknown closely resembles Psa28 in the *C.*

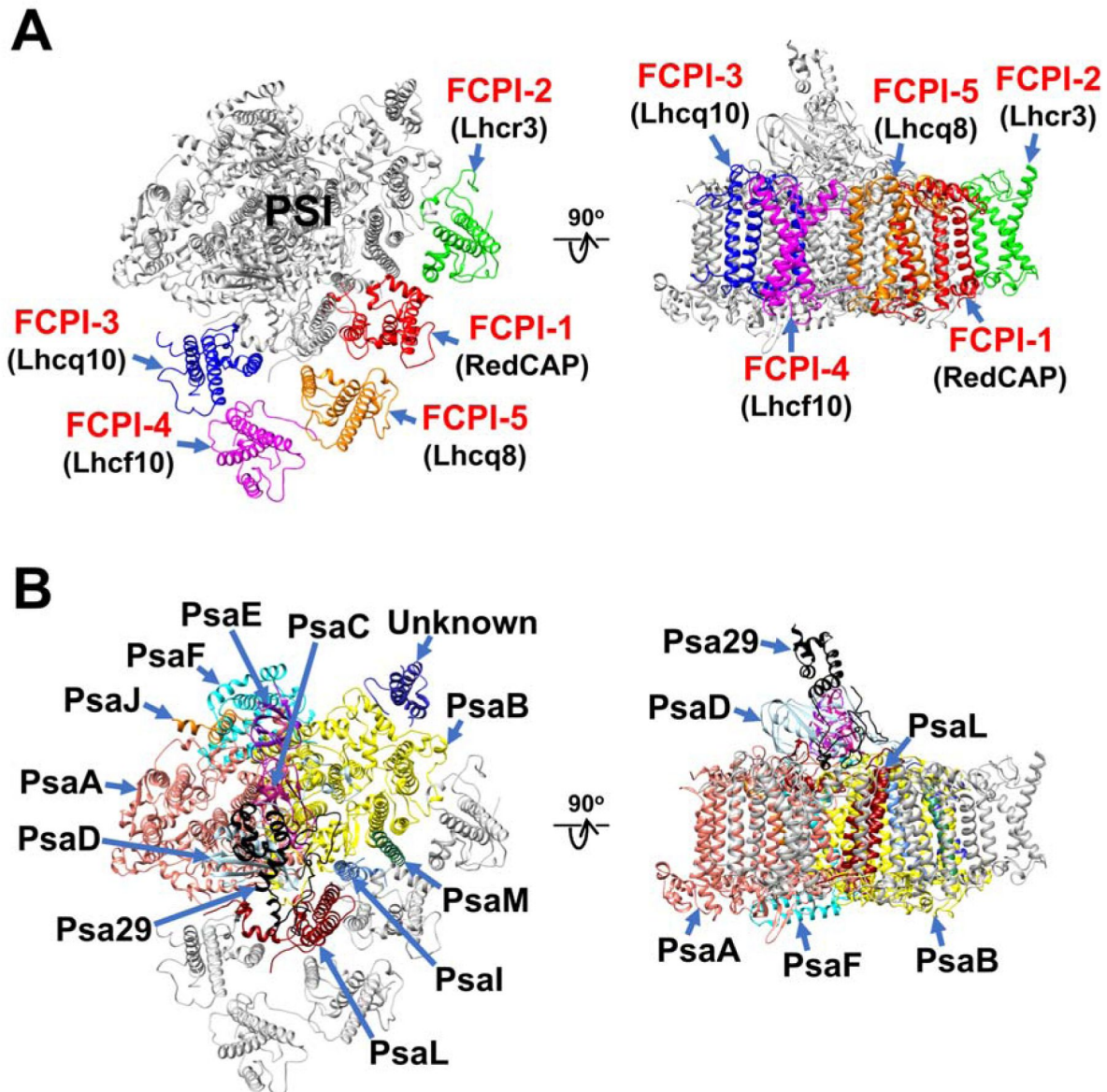


Fig. 1

Overall structure of the PSI-FCPI supercomplex from *T. pseudonana*. Structures are viewed from the stromal side (left panels) and from the direction perpendicular to the membrane normal (right panels). Only protein structures are depicted, with cofactors omitted for clarity. The FCPI (**A**) and PSI-core (**B**) subunits are labeled and colored distinctly. The five FCPI subunits are labeled as FCPI-1 to 5 (red), with their corresponding gene products indicated in parentheses (black) in panel **A**.

gracilis PSI-FCPI (Fig. S4B). Psa28, a novel subunit identified in the *C. gracilis* PSI-FCPI structure (15), follows the previously established nomenclature rule (32). Historically, genes encoding PSI proteins have been designated as *psaA*, *psaB*, and so forth. PsaZ was identified in the PSI cores of *Gloeobacter violaceus* PCC 7421 (33, 34). Subsequent discoveries led to the designation of a new subunit as Psa27, which was identified in the PSI cores of *Acaryochloris marina* MBIC11017 (35–37). Consequently, we designated this novel subunit as Psa28 (15). However, Xu et al. referred to this subunit as PsaR in the PSI-FCPI structure of *C. gracilis* (16).

Psa29 is newly identified in the *T. pseudonana* PSI-FCPI structure using ModelAngelo (38) and the NCBI database (<https://www.ncbi.nlm.nih.gov/>) (Fig. 2). The subunit corresponding to Psa29 was also observed previously in the *C. gracilis* PSI-FCPI structures (15, 16), where it was modeled as polyalanines and referred to as either Unknown1 (15) or PsaS (16). Psa29 exhibits a unique structure distinct from the other PSI subunits in the *T. pseudonana* PSI-FCPI (Fig. 2A, Fig. S4C) and engages in multiple interactions with PsaB, PsaC, PsaD, and PsaL at distances of 2.5–3.2 Å (Fig. 2B–G). Sequence analyses suggest that Psa29 has undergone evolutionary divergence between Bacillariophyceae (diatoms) and Bolidophyceae, the latter of which is a sister group of diatoms within Stramenopiles (Fig. 2H), although this subunit has not been found in other organisms. The arrangement of PSI subunits in the *T. pseudonana* PSI-FCPI is virtually identical to that in the *C. gracilis* PSI-FCPI structures already reported (15, 16). However, the functional and physiological roles of Psa29 remain unclear at present. It is evident that Psa29 does not have any pigments, quinones, or metal complexes, suggesting no contribution of Psa29 to electron transfer reactions within PSI. Further mutagenesis studies will be necessary to investigate the role of Psa29 in diatom photosynthesis.

The number and arrangement of Chls and Cars within the PSI core in the *T. pseudonana* PSI-FCPI structure (Fig. S4D, E) are largely similar to those in the *C. gracilis* PSI-FCPI structure (15). However, Chl a102 of PsaI was found in the *T. pseudonana* PSI-FCPI structure but not in the *C. gracilis* PSI-FCPI structure (15), whereas a844 of PsaA and BCR843 of PsaB were identified in the *C. gracilis* PSI-FCPI structure (15) but not in the *T. pseudonana* PSI-FCPI structure. One of the Car molecules in PsaJ was identified as ZXT103 in the *T. pseudonana* PSI-FCPI structure, while it is BCR103 in the *C. gracilis* PSI-FCPI structure (15).

Structure of the *T. pseudonana* FCPIs

Kumazawa et al. classified 44 *Lhc* genes in *T. pseudonana*, designating them as *Lhcf*, *Lhcq*, *Lhcr*, *Lhcx*, *Lhcz*, and *CgLhcr9* homologs (30). Based on this classification, the five FCPI subunits in the PSI-FCPI structure were identified using five genes: *RedCAP*, *Lhcr3*, *Lhcq10*, *Lhcf10*, and *Lhcq8*, corresponding to FCPI-1 to 5, respectively (Fig. 1A). It is important to note that *RedCAP* is not included among the 44 *Lhc* genes (30) but is classified within the LHC protein superfamily (6, 7). For the assignment of each FCPI subunit, we focused on characteristic amino acid residues derived from their cryo-EM map, especially S61/V62/Q63 in FCPI-1; A70/R71/W72 in FCPI-2; Y64/R65/E66 in FCPI-3; M63/R64/Y65 in FCPI-4; and A62/R63/R64 in FCPI-5 (Fig. S5). The root mean square deviations (RMSDs) of the structures between FCPI-4 and the other four FCPIs range from 1.91 to 3.73 Å (Table S4).

Each FCPI subunit binds several Chl and Car molecules: 7 Chls *a*/1 Chl *c*/2 Fxs/3 Ddxs/2 BCRs in FCPI-1; 10 Chls *a*/1 Chl *c*/3 Fxs/1 Ddx in FCPI-2; 7 Chls *a*/3 Chls *c*/2 Fxs/2 Ddxs in FCPI-3; 11 Chls *a*/2 Chls *c*/4 Fxs in FCPI-4; and 10 Chls *a*/4 Fxs/1 Ddx in FCPI-5 (Fig. S6A–E, Table S3). The axial ligands of the central Mg atoms of Chls within each FCPI are primarily provided by the main and side chains of amino acid residues (Table S5). Potential excitation-energy-transfer pathways can be proposed based on the close physical interactions among Chls between FCPI-3 and PsaA, between FCPI-3 and PsaL, between FCPI-1 and PsaI, and between FCPI-2 and PsaB (Fig. S7).

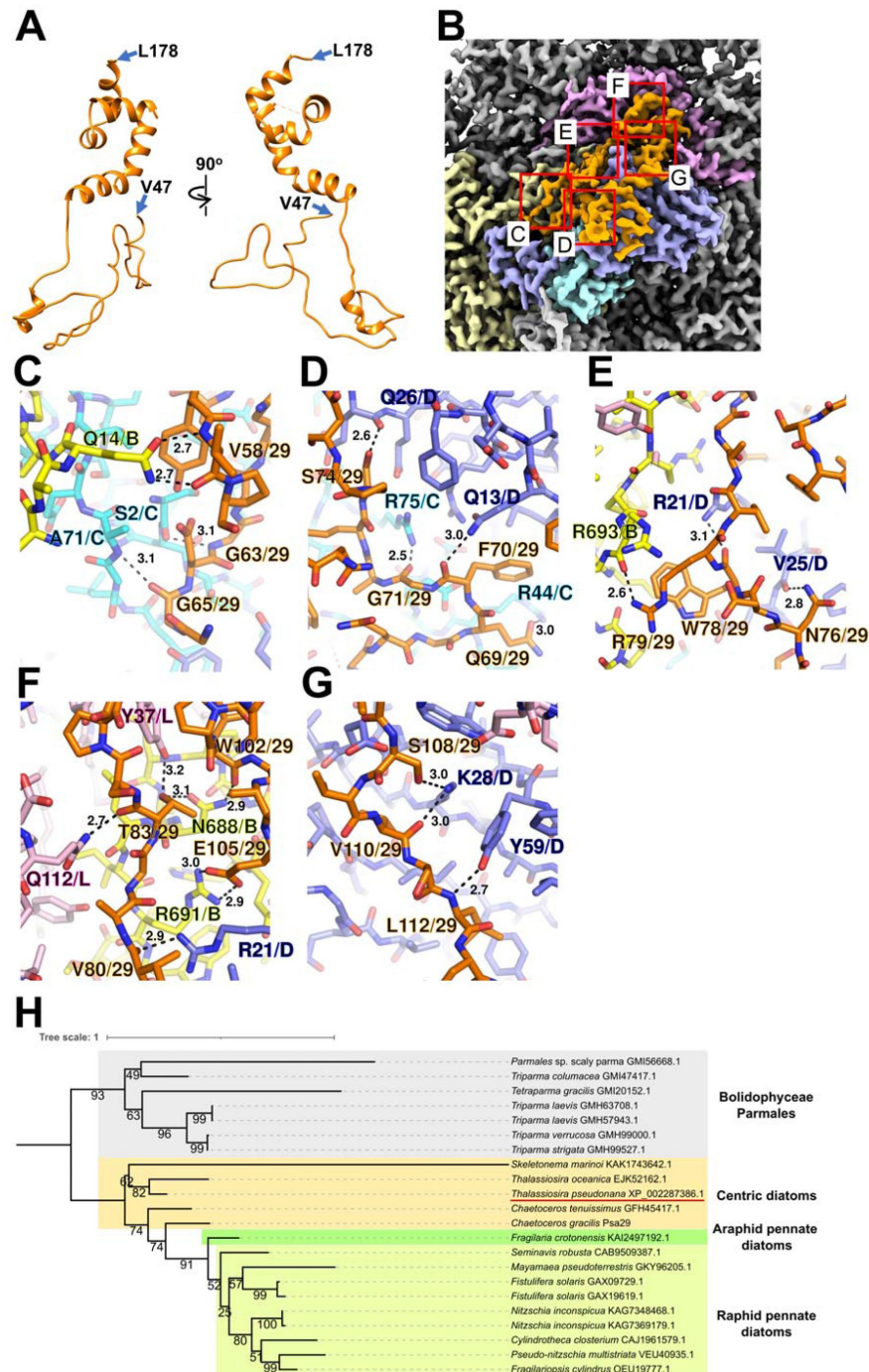


Fig. 2

Structure and diversity of PsA29. **(A)** Structure of PsA29 depicted as cartoons. PsA29 was modeled from V47 to L178. **(B)** Cryo-EM map of PsA29 and its surrounding environment, viewed from the stromal side. The red-squared areas are enlarged in panels **C–G**. Yellow, PsA/B; cyan, PsA/C; blue, PsA/D; magenta, PsA/L; orange, PsA29. **(C–G)** Protein-protein interactions of PsA29 with PsA/B/PsA/C (**C**), PsA/C/PsA/D (**D**), PsA/B/PsA/D (**E**), PsA/B/PsA/D/PsA/L (**F**), and PsA/D (**G**). Interactions are indicated by dashed lines, and the numbers are distances in Å. Amino acid residues participating in the interactions are labeled; for example, A71/C indicates Ala71 of PsA/C. B, PsA/B; C, PsA/C; D, PsA/D; L, PsA/L; 29, PsA29. **(H)** Phylogenetic analysis of PsA29 in photosynthetic organisms. A maximum-likelihood tree of PsA29 proteins was inferred using IQ-TREE v2.2.2.7 with the WAG+F+G4 model and a trimmed alignment of 22 sequences comprising 245 amino acid residues. Numbers at the nodes represent ultrafast bootstrap support (%) (1000 replicates). The tree was mid-point rooted between diatoms and Bolidophyceae Parmales. PsA29 of *T. pseudonana* CCMP1335 is indicated by a red underline.

Structural characteristics of RedCAP and its evolutionary implications

Among the FCPI subunits, only FCPI-1 contains two BCRs in addition to Fxs and Ddxs (Fig. S6A, F). This is the first report of BCR binding to FCPIs in diatoms. FCPI-1 is identified as RedCAP, a member of the LHC protein superfamily but distinct from the LHC protein family (6, 7); however, the functional and physiological roles of RedCAP remain unknown. FCPI-1 is positioned near PsaB, PsaI, and PsaL through protein-protein interactions with these subunits on both the stromal and luminal sides (Fig. 3A). At the stromal side, I138 and S139 of FCPI-1 interact with K121, G122, and F125 of PsaL (Fig. 3B), whereas at the luminal side, multiple interactions occur between I109 of FCPI-1 and F5 of PsaI, between T105/L106/T108 of FCPI-1 and W92/P94/F96 of PsaB, and between E102/W103 of FCPI-1 and S71/I73 of PsaL (Fig. 3C). The protein-protein interactions at the luminal side (Fig. 3C) appear to be caused by a loop structure of FCPI-1 from Q96 to T116 (pink in Fig. 3D), which is unique to FCPI-1 but absent in the other four FCPI subunits (pink in Fig. 3E). This loop structure is inserted into a cavity formed by PsaB, PsaI, and PsaL (Fig. 3C, D). These findings indicate that the Q96–T116 loop of FCPI-1 specifically recognizes and binds to the cavity provided by the PSI subunits.

RedCAP of *C. gracilis* (CgRedCAP) was not identified in the *C. gracilis* PSI-FCPI structures (15, 16). As previously discussed (14), we proposed that CgRedCAP may bind to the *C. gracilis* PSI core at a site similar to LHCI-1 in the red alga *C. caldarium* PSI-LHCI through sequence analysis. This site corresponds to the FCPI-1 site in the PSI-FCPI of *T. pseudonana* in this study. A sequence alignment between RedCAP of *T. pseudonana* (TpRedCAP) and CgRedCAP is shown in Fig. S8A, exhibiting a 72% sequence similarity. CgRedCAP contains a protein motif, Q106–I113 (QWGTLATI), corresponding to E102–I109 (EWGTLATI) in TpRedCAP (Fig. 3C). These findings suggest the potential binding of CgRedCAP to PSI in *C. gracilis* at a position similar to FCPI-1 in the *T. pseudonana* PSI-FCPI structure. However, it remains unclear (i) whether CgRedCAP is indeed bound to the *C. gracilis* PSI-FCPI supercomplex and (ii) if a loop structure corresponding to the Q96–T116 loop of TpRedCAP exists in CgRedCAP. Further structural studies of the *C. gracilis* PSI-FCPI are required to elucidate the molecular assembly mechanism of diatom RedCAPs.

RedCAPs have been found in the PSI-LHCI structures of the red alga *Porphyridium purpureum* (13) and a PSI supercomplex with alloxanthin Chl *a/c*-binding proteins (PSI-ACPI) in the cryptophyte *Chroomonas placoidea* (17), as summarized in our previous study (14). Both *P. purpureum* RedCAP (PpRedCAP) and *C. placoidea* RedCAP (CpRedCAP) exhibit loop structures similar to the Q96–T116 loop in TpRedCAP observed in the present study (Fig. S8B). Multiple sequence alignments of TpRedCAP with PpRedCAP and CpRedCAP are shown in Fig. S8C, revealing sequence similarities of 39% and 60%, respectively. PpRedCAP contains a protein motif of V105–L112 (VWGPLAQL), while CpRedCAP has a protein motif of Q117–A124 (QWGPLASA). These motifs correspond to E102–I109 (EWGTLATI) in TpRedCAP; however, the sequence conservation between TpRedCAP and PpRedCAP/CpRedCAP is lower than between TpRedCAP and CgRedCAP. Among the four RedCAPs, the amino acids Trp, Gly, Leu, and Ala are conserved in the protein motifs (xWGXLAxx), implying that this conserved loop structure contributes to the binding of RedCAP to PSI across the red-lineage algae.

Protein-protein interactions of the other FCPI subunits

FCPI-2 (Lhcr3) is positioned near PsaB and PsaM, engaging in protein-protein interactions with these subunits at distances of 3.0–4.3 Å at both the stromal and luminal sides (Fig. 4). The amino acid residues I63/T65/D66/Y69/W134/Y138/D140 of FCPI-2 are associated with W153/L154/K159/F160/W166 of PsaB at the stromal side (Fig. 4B), while F116 and F120 of FCPI-2 interact with F5/I9/M12 of PsaM at the luminal side (Fig. 4C). The amino acid sequences

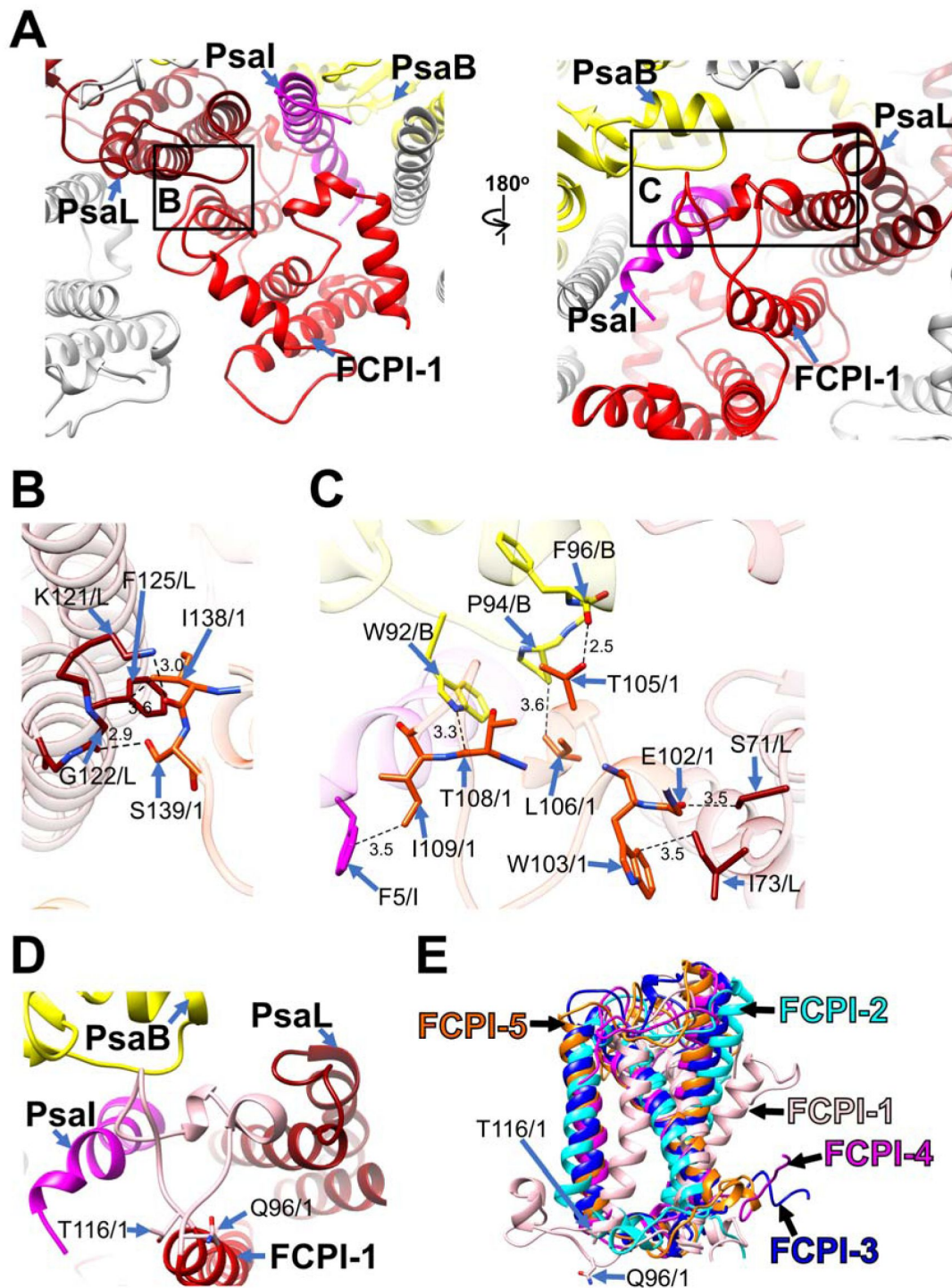


Fig. 3

Structural characteristics of FCPI-1 (RedCAP). **(A)** Interactions of FCPI-1 with PsaB, PsaI, and PsaL viewed from the stromal (left) and luminal (right) sides. The areas encircled by black squares are enlarged in panels **B** and **C**. Yellow, PsaB; magenta, PsaI; dark red, PsaL; red, FCPI-1. **(B, C)** Protein-protein interactions of FCPI-1 with PsaL **(B)** and with PsaB/PsaI/PsaL **(C)**. Interactions are indicated by dashed lines, and the numbers are distances in Å. Amino acid residues involved in the interactions are labeled; for example, S139/1 indicates Ser139 of FCPI-1. **(D)** Characteristic loop structure from Q96 to T116 in FCPI-1, viewed from the luminal side. Q96 and T116 are labeled with sticks, and the Q96-T116 loop is colored pink. **(E)** Superpositions of FCPI-1 with FCPI-2, FCPI-3, FCPI-4, and FCPI-5. Only proteins are depicted. Q96 and T116 in the Q96-T116 loop of FCPI-1 are shown with sticks.

corresponding to I63–Y69, F116–F120, and W134–D140 in Lhcr3 are not conserved in the Lhcr subfamily, comprising Lhcr1, Lhcr4, Lhcr7, Lhcr11, Lhcr12, Lhcr14, Lhcr17, Lhcr18, Lhcr19, and Lhcr20, as reported by Kumazawa et al. (30) (Fig. S9).

FCPI-3 (Lhcq10) is positioned near Psal, with protein-protein interactions at distances of 2.3–4.2 Å at the stromal side (Fig. 5A, B). The amino acid residues L126/I130/L142/Y146/W147/V148/W155 of FCPI-3 are associated with F4/K6/P20/S25/L26/L30 of Psal (Fig. 5B). Given the homology between TpLhcq10 and CgLhcr9 (30), we compared the amino acid sequence of Lhcq10 with the Lhcq and Lhcr subfamilies in *T. pseudonana* (Fig. S10A, B). The sequence L126–W155 of Lhcq10 is not conserved in the Lhcq subfamily, comprising Lhcq1, Lhcq2, Lhcq3, Lhcq4, Lhcq5, Lhcq6, Lhcq7, Lhcq8, and Lhcq9 (Fig. S10A), nor in the Lhcr subfamily, comprising Lhcr1, Lhcr3, Lhcr4, Lhcr7, Lhcr11, Lhcr12, Lhcr14, Lhcr17, Lhcr18, Lhcr19, and Lhcr20, as reported by Kumazawa et al. (30) (Fig. S10B).

FCPI-4 (Lhcf10) is positioned near FCPI-5 through protein-protein interactions with it at distances of 2.6–3.6 Å at the luminal side (Fig. 5A, C). The amino acid residues Y196/P198/F199 of FCPI-4 interact with F82/F86/G87 of FCPI-5 (Fig. 5C). The amino acid sequence Y196–F199 of Lhcf10 is not conserved in the Lhcf subfamily, comprising Lhcf1, Lhcf2, Lhcf3, Lhcf4, Lhcf5, Lhcf6, Lhcf7, Lhcf8, Lhcf9, Lhcf11, and Lhcf12, as reported by Kumazawa et al. (30) (Fig. S11).

FCPI-5 (Lhcq8) is positioned near Psal and FCPI-4 through protein-protein interactions at distances of 2.6–4.1 Å at both the stromal and luminal sides (Fig. 5A, C, D). The amino acid residues P108/Q109/A112/I115 of FCPI-5 interact with F134/I137/S141 of Psal at the luminal side (Fig. 5D). The interactions between FCPI-5 and FCPI-4 are shown in Fig. 5C. The amino acid sequences F82–G87 and P107–I115 of Lhcq8 are not conserved in the Lhcq subfamily, comprising Lhcq1, Lhcq2, Lhcq3, Lhcq4, Lhcq5, Lhcq6, Lhcq7, Lhcq9, and Lhcq10, as reported by Kumazawa et al. (30) (Fig. S12A, B).

Molecular insights into the assembly of FCPIs in diatom PSI-FCPI supercomplexes

To evaluate the molecular assembly of FCPI subunits in the *T. pseudonana* PSI-FCPI structure, we focused on protein-protein interactions based on their close proximities (Fig. 3–5) and the amino acid residues in non-conserved regions among 44 FCPs (Fig. S9–S12). This approach is based on the premise that selective associations of FCPIs with PSI require specific amino acid residues unique to each FCPI. Protein-protein interactions among FCPI subunits, as well as between FCPI and PSI subunits, occur at both the stromal and luminal sides (Fig. 3–5), and are likely recognized by unique amino acid residues of FCPIs that are not conserved in each LHC subfamily (Fig. S9–S12). Thus, the binding and assembly of each FCPI subunit to PSI are likely determined by the amino acid sequences within the loop regions of the 44 FCPs in *T. pseudonana*.

The diatom *C. gracilis* exhibits two distinct PSI-FCPI structures: one with 16 FCPI subunits (15) and the other with 24 FCPI subunits (16). These structural variations arise from changes in the antenna sizes of FCPIs within the *C. gracilis* PSI-FCPI supercomplexes, in response to varying growth conditions, especially CO₂ concentrations and temperatures (39). Notably, the *C.*

gracilis PSI-FCPI structure contains five FCPI subunits located at the same binding sites as FCPI-1 to 5 in the *T. pseudonana* PSI-FCPI structure (Fig. 6A). A summary of the relationship between the *Lhc* genes encoding FCPs, the distinct gene *RedCAP*, and the binding positions of FCPI-1 to 5 in *T. pseudonana* and *C. gracilis* is shown in Fig. 6B. The gene nomenclature for the *C. gracilis* FCPIs follows the conventions established by Kumazawa et al. (30), as discussed in our recent study (14).

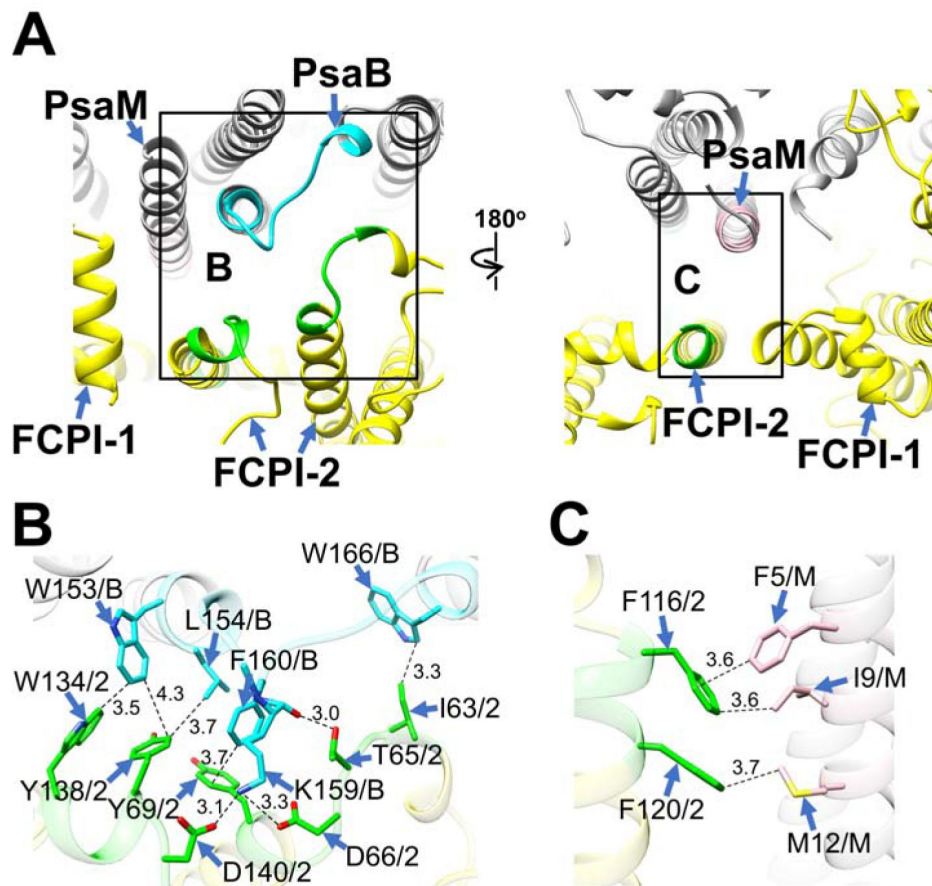


Fig. 4

Structural characteristics of FCPI-2. **(A)** Interactions of FCPI-2 with PsaB and PsaM viewed from the stromal (left) and luminal (right) sides. The areas encircled by black squares are enlarged in panels **B** and **C**. PSI subunits are colored grey, and FCPI subunits are colored yellow. Protein-protein interactions are shown in different colors: green, FCPI-2; cyan, PsaB; pink, PsaM. **(B, C)** Protein-protein interactions of FCPI-2 with PsaB (**B**) and PsaM (**C**). Interactions are indicated by dashed lines, and the numbers represent distances in Å. Amino acid residues involved in the interactions are labeled; for example, Y138/2 indicates Tyr138 of FCPI-2. B, PsaB; M, PsaM; 2, FCPI-2.

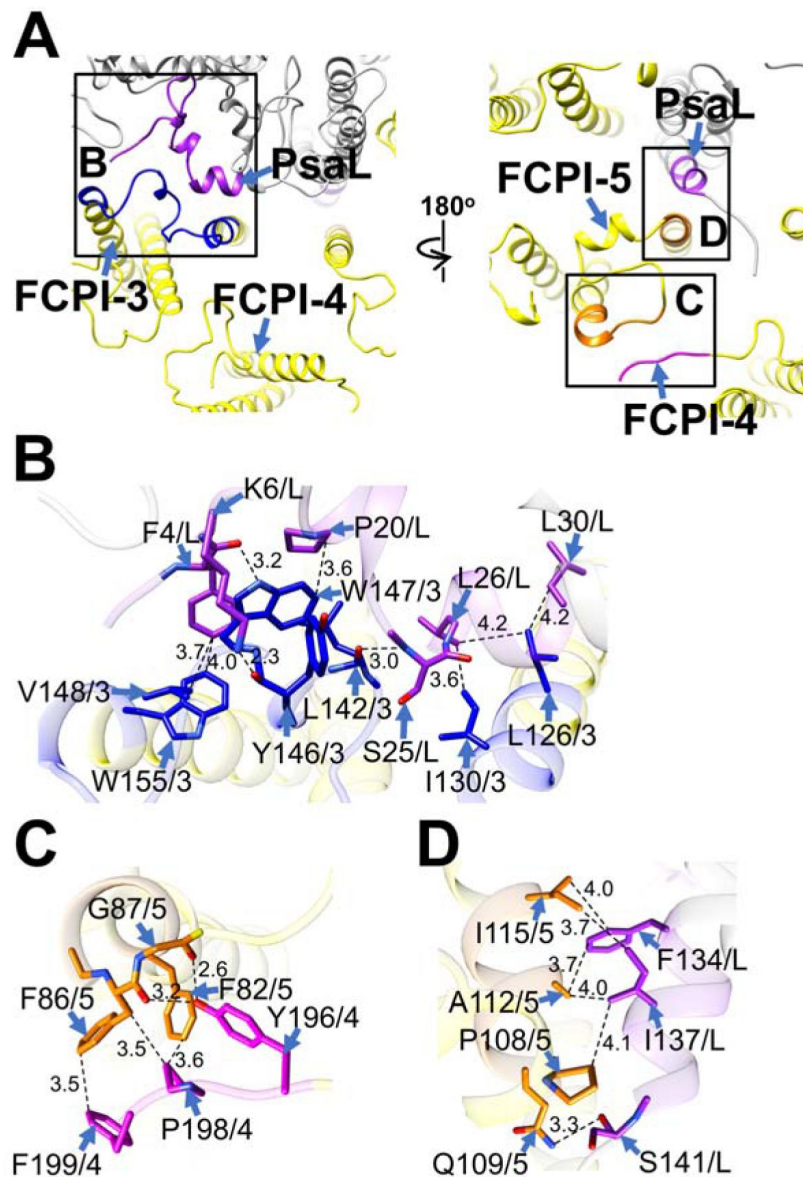


Fig. 5

Structural characteristics of FCPI-3, 4, and 5. **(A)** Interactions among FCPIs and between FCPIs and PsaL, viewed from the stromal (left) and luminal (right) sides. The areas encircled by black squares are enlarged in panels **B-D**. PSI subunits are colored grey, and FCPI subunits are colored yellow. Protein-protein interactions are shown in different colors: blue, FCPI-3; magenta, FCPI-4; orange, FCPI-5; purple, PsaL. **(B-D)** Protein-protein interactions between FCPI-3 and PsaL **(B)**, between FCPI-4 and FCPI-5 **(C)**, and between FCPI-5 and PsaL **(D)**. Interactions are indicated by dashed lines, and the numbers represent distances in Å. Amino acid residues involved in the interactions are labeled; for example, L126/3 indicates Leu126 of FCPI-3. L, PsaL; 3, FCPI-3; 4, FCPI-4; 5, FCPI-5.

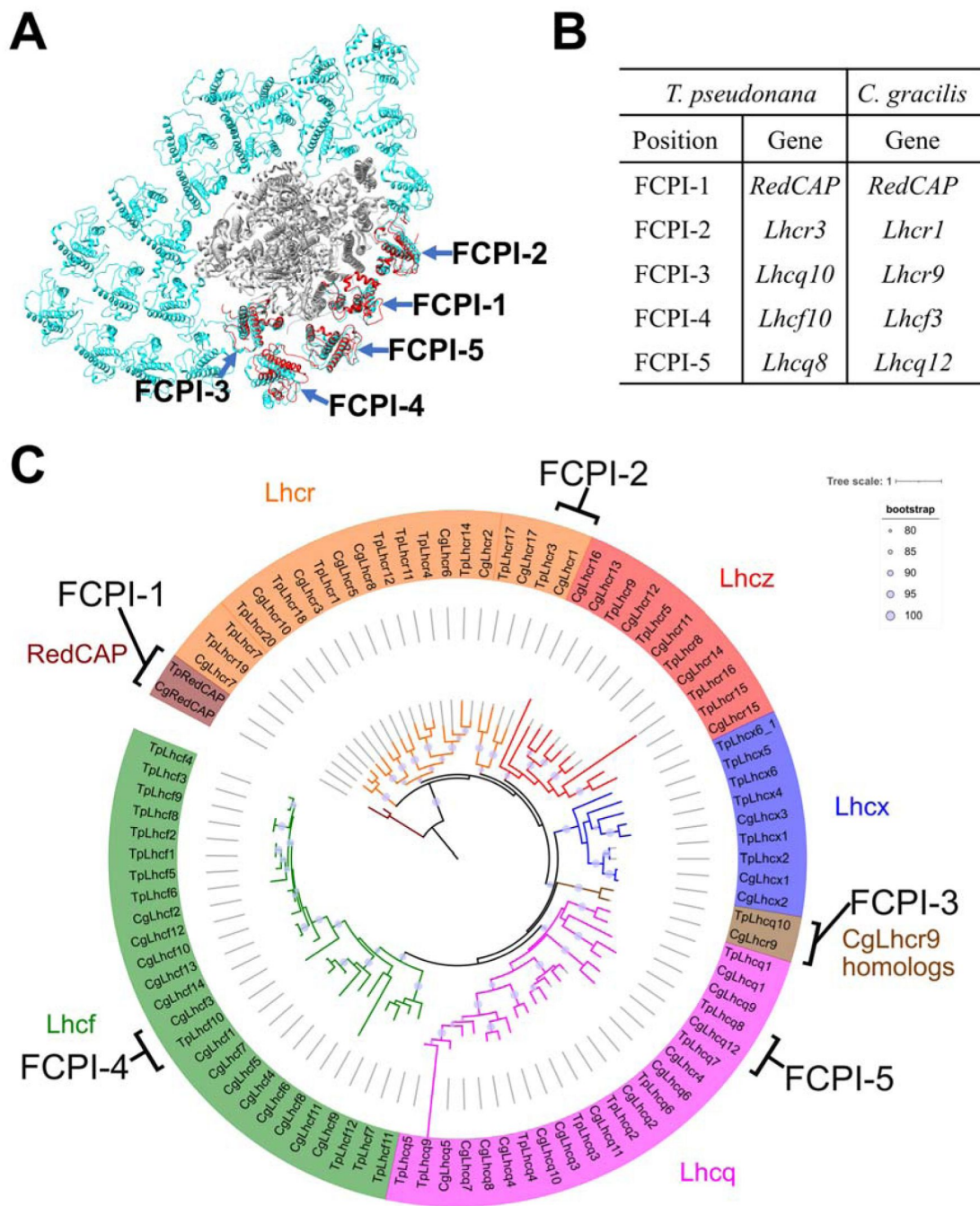


Fig. 6.

Comparisons of structures and sequences of FCPs in the PSI-FCPI structures between *T. pseudonana* and *C. gracilis*. (**A**) Superposition of the PSI-FCPI structures between *T. pseudonana* and *C. gracilis* (PDB: 6LY5). FCPI subunits from *T. pseudonana* and *C. gracilis* are colored red and cyan, respectively. The structures are viewed from the stromal side. The FCPI-1 to 5 sites are labeled. (**B**) Correlation of the names of FCPs in the structures with their corresponding genes between *T. pseudonana* and *C. gracilis*. The FCPI genes are derived from Kumazawa et al. (30) and Kato et al. (14) for *C. gracilis*. (**C**) Phylogenetic analysis of FCPs and RedCAPs from *T. pseudonana* (Tp) and *C. gracilis* (Cg). In addition to the RedCAP family, 44 TpFCPs and 46 CgFCPs are grouped into five Lhc subfamilies and CgLhcr9 homologs. Maroon, RedCAP family; magenta, Lhcq subfamily; red, Lhcz subfamily; orange, Lhcr subfamily; brown, CgLhcr9 homologs; green, Lhcf subfamily; blue, Lhcx subfamily. The FCPs and RedCAPs located at the FCPI-1 to 5 sites are labeled. The tree was inferred using IQ-TREE 2 (59) with the Q.pfam + R4 model selected by ModelFinder (60). The light purple circular symbols on the tree represent bootstrap support (%).

Phylogenetic analysis clearly showed that at the FCPI-1, 2, 3, and 5 sites in the *T. pseudonana* PSI-FCPI structure, TpRedCAP, TpLhcr3, TpLhcq10, and TpLhcq8 are orthologous to CgRedCAP, CgLhcr1, CgLhcr9, and CgLhcq12, respectively (**Fig. 6C** [↗](#)). The characteristic protein loops of TpRedCAP and CpRedCAP likely participate in interactions with PSI at the FCPI-1 site, as noted above (Fig. S8). At the FCPI-2 site, comparative analyses revealed that the amino acid residues facilitating interactions between TpLhcr3 and TpPsaB/TpPsaM closely parallel those observed in the CgLhcr1-CgPsaB and CgLhcr1-CgPsaM pairs (Fig. S13). Similarly, a high degree of similarity characterized the residues involved in the interaction pairs of TpLhcq10-TpPsaL/CgLhcr9-CgPsaL at the FCPI-3 site (Fig. S14A, B), as well as TpLhcq8-TpPsaL/CgLhcq12-CgPsaL at the FCPI-5 site (Fig. S14C, D), underscoring the conserved nature of these interactions. However, TpLhcf10 was not homologous to CgLhcf3 (**Fig. 6C** [↗](#)), despite both being located at the FCPI-4 site in their respective PSI-FCPI structures (**Fig. 6A** [↗](#)). These findings suggest that the two diatoms possess both a conserved mechanism of protein-protein interactions across characteristic protein motifs between FCPI and PSI subunits, and a different mechanism of interactions among FCPIs.

It is notable that the *C. gracilis* PSI-FCPI structure binds remarkably more FCPI subunits than that of *T. pseudonana*, e.g., 16 or 24 subunits in *C. gracilis* as reported in the previous studies ([15](#) [↗](#), [16](#) [↗](#)), versus 5 subunits in *T. pseudonana* in the present study. The reason for this difference remains unclear. One possibility is that some FCPI subunits are released during detergent solubilization in *T. pseudonana*, while they are retained in *C. gracilis*. Alternatively, the number of FCPI subunits may be inherently lower in *T. pseudonana*, which may reflect adaptations to different living environments. Further studies are needed to resolve this question.

Extension to molecular assembly of PSI-LHCI supercomplexes

The mechanisms of protein-protein interactions in diatom PSI-FCPI supercomplexes are likely developed by the specific binding of FCPs selected from 44 TpFCPs and 46 CgFCPs in addition to RedCAPs. Like a lock-and-key mechanism, one FCP cannot be substituted by another in forming the PSI-FCPI supercomplexes in the two diatoms; for example, TpLhcq10 binds specifically at the FCPI-3 site but not at the other sites such as FCPI-2. This selective binding mechanism of FCPIs may dictate the molecular assembly of PSI-FCPI. Importantly, the selective binding of FCPIs was identified for the first time by comparing the structures of PSI-FCPI supercomplexes and the amino acid sequences of FCPIs between the two diatom species. This approach can be extended to the LHC protein superfamily in the green and red lineages, enabling comparisons of protein structures and sequences of PSI-LHCI supercomplexes among closely related species. This, in turn, lays the foundation for elucidating the underlying mechanism of PSI-LHCI supercomplex assembly. Thus, this study will shed light on answering the evolutionary question of how LHCI recognize their binding sites at PSI in photosynthetic organisms.

Methods

Cell growth and preparation of thylakoid membranes

The marine centric diatom *T. pseudonana* CCMP1335 was grown in artificial seawater supplemented with sodium metasilicate and KW21 ([40](#) [↗](#)) at 20°C under a photosynthetic photon flux density of 30 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ provided by white LED, with bubbling of air containing 3% (v/v) CO₂. The cells were harvested by centrifugation, disrupted by agitation with glass beads ([41](#) [↗](#)), and the thylakoid membranes were pelleted by further centrifugation. The resulting thylakoid membranes were suspended in 50 mM Mes-NaOH (pH 6.5) buffer containing 1 M betaine and 1 mM EDTA.

Purification of the PSI-FCPI supercomplex

Thylakoid membranes were solubilized with 1% (w/v) *n*-dodecyl- β -D-maltoside (β -DDM) at a Chl concentration of 0.5 mg mL^{-1} for 20 min on ice in the dark with gentle stirring. After centrifugation at $162,000 \times g$ for 20 min at 4°C , the supernatant was loaded onto a Q-Sepharose anion-exchange column (1.6 cm inner diameter, 25 cm length) equilibrated with 20 mM Mes-NaOH (pH 6.5) buffer containing 0.2 M trehalose, 5 mM CaCl_2 , and 10 mM MgCl_2 , and 0.03% β -DDM (buffer A). The column was washed with buffer A until the eluate became colorless. Elution was performed at a flow rate of 1.0 mL min^{-1} using a linear gradient of buffer A and buffer B (buffer A plus 500 mM NaCl) with the following time and gradient: 0–600 min, 0–60% buffer B; 600–800 min, 60–100% buffer B; 800–900 min, 100% buffer B. The PSI-FCPI-enriched fraction was eluted at 194–247 mM NaCl, then collected and subsequently loaded onto a linear gradient containing 10–40% (w/v) trehalose in 20 mM Mes-NaOH (pH 6.5) buffer containing 5 mM CaCl_2 , 10 mM MgCl_2 , 100 mM NaCl, and 0.03% β -DDM. After centrifugation at $154,000 \times g$ for 18 h at 4°C (P40ST rotor; Hitachi), a green fraction (Fig. S1A) was collected and concentrated using a 150 kDa cut-off filter (Apollo; Orbital Biosciences) at $4,000 \times g$. The concentrated samples were stored in liquid nitrogen until use.

Biochemical and spectroscopic analyses of the PSI-FCPI supercomplex

The polypeptide bands of PSI-FCPI were analyzed by SDS-PAGE with 16% (w/v) acrylamide and 7.5 M urea, following the method of Ikeuchi and Inoue (42) (Fig. S1B). The PSI-FCPI supercomplexes (4 μg of Chl) were solubilized in 3% lithium lauryl sulfate and 75 mM dithiothreitol for 10 min at 60°C , and then loaded onto the gel. A standard molecular weight marker (SP-0110; APRO Science) was used. The absorption spectrum of PSI-FCPI was measured at room temperature using a UV-Vis spectrophotometer (UV-2450; Shimadzu) (Fig. S1C), and the fluorescence emission spectrum of PSI-FCPI was measured at 77 K upon excitation at 430 nm using a spectrofluorometer (RF-5300PC; Shimadzu) (Fig. S1D). The pigment composition of PSI-FCPI was analyzed by HPLC following the method of Nagao et al. (43), and the elution profile was monitored at 440 nm (Fig. S1E).

Cryo-EM data collection

A 3- μL aliquot of the *T. pseudonana* PSI-FCPI supercomplex ($3.0 \text{ mg of Chl mL}^{-1}$) in 20 mM MES-NaOH (pH 6.5) buffer containing 0.5 M betaine, 5 mM CaCl_2 , 10 mM MgCl_2 , and 0.03% β -DDM was applied to Quantifoil R1.2/1.3 Cu 300 mesh grids in the chamber of FEI Vitrobot Mark IV (Thermo Fisher Scientific). The grid was then blotted with filter paper for 4 sec at 4°C under 100% humidity and plunged into liquid ethane cooled by liquid nitrogen. The frozen grid was transferred to a CRYO ARM 300 electron microscope (JEOL) equipped with a cold-field emission gun operated at 300 kV. All image stacks were collected from 5×5 holes per stage adjustment to the central hole and image shifts were applied to the surrounding holes while maintaining an axial coma-free condition. The images were recorded using an in-column energy filter with a slit width of 20 eV at a nominal magnification of $\times 60,000$ on a direct electron detector (Gatan K3, AMETEK). The nominal defocus range was -1.8 to $-1.2 \mu\text{m}$, and the physical pixel size corresponded to $0.752 \text{ \AA}$. Each image stack was exposed at a dose rate of $21.46 \text{ e}^- \text{ \AA}^{-2} \text{ sec}^{-1}$ for 2.33 sec in CDS mode, with dose-fractionated 50 movie frames. A total of 8,950 image stacks were collected.

Cryo-EM image processing

The resultant movie frames were aligned and summed using MotionCor2 (44) to produce dose-weighted images. The contrast transfer function (CTF) estimation was performed using CTFFIND4 (45). All subsequent processes were carried out using RELION-4.0 (46). A total of 2,733,572 particles were automatically picked and subjected to reference-free 2D classification. From these, 1,132,721 particles were selected from well-defined 2D classes and further processed for 3D

classification without imposing any symmetry. An initial model for the first 3D classification was generated *de novo* from the 2D classification. A 240-Å spherical mask was used during the 3D classification and refinement processes. As illustrated in Fig. S2C, the final PSI-FCPI structure was reconstructed from 75,667 particles. The overall resolution of the cryo-EM map was determined to be 2.30 Å, based on the gold-standard FSC curve with a cut-off value of 0.143 (Fig. S3A) (47). Local resolutions were calculated using RELION (Fig. S3C).

Model building and refinement

Two types of the cryo-EM maps were employed for the model building of the PSI-FCPI supercomplex: a postprocessed map and a denoised map generated using Topaz version 0.2.4 (48). The postprocessed map was denoised using a trained model over 100 epochs using two half-maps. Initial models of each subunit in the PSI-FCPI supercomplex were generated by ModelAngelo (38) and subsequently inspected and manually adjusted against the maps with Coot (49). Each model was built based on interpretable features from the density maps at a contour level of 2.5 σ in both the denoised and postprocessed maps. For the assignment of Chls, Chls *a* and *c* were distinguished by inspecting the density map corresponding to the phytol chain at the least level not to link the map of Chls with that of noise. All Chls *c* were assigned as Chl *c*1 due to the inability to distinguish between Chl *c*1 and Chl *c*2 at the present resolution. For the assignment of Cars, Fx and Ddx were distinguished based on the density surrounding the head groups of Cars with the above threshold. The PSI-FCPI structure was refined using phenix.real_space_refine (50) and Servalcat (51), incorporating geometric restraints for protein-cofactor coordination. The final model was validated with MolProbity (52), EMRinger (53), and Q-score (54). The statistics for all data collection and structure refinement are summarized in Table S1, S2. All structural figures were prepared using PyMOL (55), UCSF Chimera (56), and UCSF ChimeraX (57). Since the numbering of Chls, Cars, and other cofactors in this paper differs from those in the PDB data, the corresponding relationships are provided in Table S6–S8.

Phylogenetic analysis

Amino acid sequences were aligned using MAFFT L-INS-i v7.490 or MAFFT E-INS-i v7.520 (58). The alignment was trimmed using ClipKit v1.4.1 with the smart-gap mode. Phylogenetic trees were inferred using IQ-TREE 2 (59) with the model selected by ModelFinder (60). The trees were visualized using iTOL v6 (61). Ultrafast bootstrap approximation was performed with 1000 replicates (62).

Data availability

Atomic coordinates, cryo-EM maps, and raw image data for the reported structure have been deposited in the Protein Data Bank under an accession code 8XLS [https://www.rcsb.org/structure/8XLS], in the Electron Microscopy Data Bank under an accession code EMD-38457 [https://www.ebi.ac.uk/emdb/EMD-38457], and in the Electron Microscopy Public Image Archive under an accession code EMPIAR-12142 [https://doi.org/10.6019/EMPIAR-12142], respectively.

Acknowledgements

We thank Kumiyo Kato and Satoko Kakiuchi for their assistance in this study. The cells of *T. pseudonana* CCMP1335 were given by Prof. Yusuke Matsuda, Kwansei Gakuin University, Japan. Cryo-EM data was obtained using EM01CT and EM02CT of SPring-8 with the approval of the Japan Synchrotron Radiation Research Institute (JASRI Proposal No. 2022B2728 (J.-R.S.) and No. 2023A2715 (Y.N.)). This work was supported by JSPS KAKENHI grant Nos. JP22KJ2017 (M.K.), JP23K14211 (Y.N.), JP22H04916 (J.-R.S.), JP23H02347 (K.I.), and JP23H02423 (R.N.), Takeda Science

Foundation (R.N., K.K.), and Research Support Project for Life Science and Drug Discovery (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under support No. 4176 (J.-R.S.).

Author Contributions

R.N. conceived the project; H.O. and R.N. prepared the PSI-FCPI supercomplexes and analyzed their biochemical characteristics; J.X., M.K., and K.I. performed phylogenetic and sequence analyses; Y.N. collected cryo-EM images; K.K. processed the cryo-EM data and reconstructed the final cryo-EM map; K.K. built the structural model and refined the final model; K.K., Y.N., M.K., K.I., and R.N. drafted the original manuscript; J.-R.S. modified the manuscript; and R.N. wrote the final manuscript, and all authors joined the discussion of the results.

Declaration of competing interest

The authors declare no conflict of interest.

References

1. Blankenship R. E. (2021) **Molecular Mechanisms of Photosynthesis**
2. Golbeck J. H. (1992) **Structure and function of photosystem I** *Ann. Rev. Plant Physiol. Plant Mol. Biol* **43**:293–324
3. Brettel K., Leibl W. (2001) **Electron transfer in photosystem I** *Biochim. Biophys. Acta Bioenerg* **1507**:100–114
4. Shen J.-R. (2015) **The structure of photosystem II and the mechanism of water oxidation in photosynthesis** *Annu. Rev. Plant Biol* **66**:23–48
5. Shevela D., Kern J. F., Govindjee G., Messinger J. (2023) **Solar energy conversion by photosystem II: principles and structures** *Photosynth. Res* **156**:279–307
6. Engelken J., Brinkmann H., Adamska I. (2010) **Taxonomic distribution and origins of the extended LHC (light-harvesting complex) antenna protein superfamily** *BMC Evol. Biol* **10**
7. Sturm S., et al. (2013) **A novel type of light-harvesting antenna protein of red algal origin in algae with secondary plastids** *BMC Evol. Biol* **13**
8. Falkowski P. G., et al. (2004) **The evolution of modern eukaryotic phytoplankton** *Science* **305**:354–360
9. Hippler M., Nelson N. (2021) **The plasticity of photosystem I** *Plant Cell Physiol* **62**:1073–1081
10. Shen J.-R., Harris J. R., Marles-Wright J. (2022) **“Structure, function, and variations of the photosystem I-antenna supercomplex from different photosynthetic organisms”** *Macromolecular Protein Complexes IV. Subcellular Biochemistry*
11. Pi X., et al. (2018) **Unique organization of photosystem I-light-harvesting supercomplex revealed by cryo-EM from a red alga** *Proc. Natl. Acad. Sci. U. S. A* **115**:4423–4428
12. Antoshvili M., Caspy I., Hippler M., Nelson N. (2019) **Structure and function of photosystem I in Cyanidioschyzon merolae** *Photosynth. Res* **139**:499–508
13. You X., et al. (2023) **In situ structure of the red algal phycobilisome-PSII-PSI-LHC megacomplex** *Nature* **616**:199–206
14. Kato K., et al. (2024) **The structure of PSI-LHCI from Cyanidium caldarium provides evolutionary insights into conservation and diversity of red-lineage LHCs** *Proc. Natl. Acad. Sci. U. S. A* **121**
15. Nagao R., et al. (2020) **Structural basis for assembly and function of a diatom photosystem I-light-harvesting supercomplex** *Nat. Commun* **11**
16. Xu C., et al. (2020) **Structural basis for energy transfer in a huge diatom PSI-FCPI supercomplex** *Nat. Commun* **11**

17. Zhao L.-S., et al. (2023) **Structural basis and evolution of the photosystem I-light-harvesting supercomplex of cryptophyte algae** *Plant Cell* **35**:2449–2463
18. Li X., et al. (2024) **Structures and organizations of PSI-AcpPCI supercomplexes from red tidal and coral symbiotic photosynthetic dinoflagellates** *Proc. Natl. Acad. Sci. U. S. A* **121**
19. Zhao L.-S., et al. (2024) **Architecture of symbiotic dinoflagellate photosystem I-light-harvesting supercomplex in Symbiodinium** *Nat. Commun* **15**
20. Field C. B., Behrenfeld M. J., Randerson J. T., Falkowski P. (1998) **Primary production of the biosphere: Integrating terrestrial and oceanic components** *Science* **281**:237–240
21. Green B. R., Durnford D. G. (1996) **The chlorophyll-carotenoid proteins of oxygenic photosynthesis** *Annu. Rev. Plant Physiol. Plant Mol. Biol* **47**:685–714
22. Büchel C. (2020) **Light harvesting complexes in chlorophyll c-containing algae** *Biochim. Biophys. Acta Bioenerg* **1861**
23. Wang W., Shen J.-R., Shen J.-R., Satoh K., Allakhverdiev S. I. (2021) **“Structure, organization and function of light-harvesting complexes associated with photosystem II”** *Photosynthesis: molecular approaches to solar energy conversion*
24. Ikeda Y., et al. (2008) **Photosystem I complexes associated with fucoxanthin-chlorophyll-binding proteins from a marine centric diatom, Chaetoceros gracilis**. *Biochim. Biophys. Acta Bioenerg* **1777**:351–361
25. Nagao R., Kagatani K., Ueno Y., Shen J.-R., Akimoto S. (2019) **Ultrafast excitation energy dynamics in a diatom photosystem I-antenna complex: a femtosecond fluorescence upconversion study** *J. Phys. Chem. B* **123**:2673–2678
26. Nagao R., et al. (2019) **Biochemical characterization of photosystem I complexes having different subunit compositions of fucoxanthin chlorophyll a/c-binding proteins in the diatom Chaetoceros gracilis** *Photosynth. Res* **140**:141–149
27. Nagao R., Yokono M., Ueno Y., Shen J.-R., Akimoto S. (2019) **Low-energy chlorophylls in fucoxanthin chlorophyll a/c-binding protein conduct excitation energy transfer to photosystem I in diatoms** *J. Phys. Chem. B* **123**:66–70
28. Nagao R., Yokono M., Ueno Y., Shen J.-R., Akimoto S. (2019) **pH-sensing machinery of excitation energy transfer in diatom PSI-FCPI complexes** *J. Phys. Chem. Lett* **10**:3531–3535
29. Nagao R., Yokono M., Ueno Y., Shen J.-R., Akimoto S. (2020) **Excitation-energy transfer and quenching in diatom PSI-FCPI upon P700 cation formation** *J. Phys. Chem. B* **124**:1481–1486
30. Kumazawa M., et al. (2022) **Molecular phylogeny of fucoxanthin-chlorophyll a/c proteins from Chaetoceros gracilis and Lhcq/Lhcf diversity** *Physiol. Plant* **174**
31. Ikeda Y., et al. (2013) **Two types of fucoxanthin-chlorophyll-binding proteins I tightly bound to the photosystem I core complex in marine centric diatoms** *Biochim. Biophys. Acta* **1827**:529–539
32. Kashino Y., et al. (2002) **Proteomic analysis of a highly active photosystem II preparation from the cyanobacterium Synechocystis sp. PCC 6803 reveals the presence of novel polypeptides** *Biochemistry* **41**

33. Inoue H., et al. (2004) **Unique constitution of photosystem I with a novel subunit in the cyanobacterium *Gloeobacter violaceus* PCC 7421** *FEBS Lett* **578**:275–279
34. Kato K., et al. (2022) **Structural basis for the absence of low-energy chlorophylls in a photosystem I trimer from *Gloeobacter violaceus*** *eLife* **11**
35. Tomo T., et al. (2008) **Characterization of highly purified photosystem I complexes from the chlorophyll d-dominated cyanobacterium *Acaryochloris marina* MBIC 11017** *J. Biol. Chem* **283**:18198–18209
36. Hamaguchi T., et al. (2021) **Structure of the far-red light utilizing photosystem I of *Acaryochloris marina*** *Nat. Commun* **12**
37. Xu C., et al. (2021) **A unique photosystem I reaction center from a chlorophyll d-containing cyanobacterium *Acaryochloris marina*** *J. Integr. Plant Biol* **63**:1740–1752
38. Jamali K., et al. (2024) **Automated model building and protein identification in cryo-EM maps** *Nature* <https://doi.org/10.1038/s41586-024-07215-4>
39. Nagao R., Ueno Y., Akimoto S., Shen J.-R. (2020) **Effects of CO₂ and temperature on photosynthetic performance in the diatom *Chaetoceros gracilis*** *Photosynth. Res* **146**:189–195
40. Nagao R., et al. (2007) **Isolation and characterization of oxygen-evolving thylakoid membranes and Photosystem II particles from a marine diatom *Chaetoceros gracilis***. *Biochim. Biophys. Acta Bioenerg* **1767**:1353–1362
41. Nagao R., Yamaguchi M., Nakamura S., Ueoka-Nakanishi H., Noguchi T. (2017) **Genetically introduced hydrogen bond interactions reveal an asymmetric charge distribution on the radical cation of the special-pair chlorophyll P680** *J. Biol. Chem* **292**:7474–7486
42. Ikeuchi M., Inoue Y. (1988) **A new photosystem II reaction center component (4.8 kDa protein) encoded by chloroplast genome** *FEBS Lett* **241**:99–104
43. Nagao R., Yokono M., Akimoto S., Tomo T. (2013) **High excitation energy quenching in fucoxanthin chlorophyll a/c-binding protein complexes from the diatom *Chaetoceros gracilis*** *J. Phys. Chem. B* **117**:6888–6895
44. Zheng S. Q., et al. (2017) **MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy** *Nat. Methods* **14**:331–332
45. Mindell J. A., Grigorieff N. (2003) **Accurate determination of local defocus and specimen tilt in electron microscopy** *J. Struct. Biol* **142**:334–347
46. Kimanius D., Dong L., Sharov G., Nakane T., Scheres S. H. W. (2021) **New tools for automated cryo-EM single-particle analysis in RELION-4.0** *Biochem. J* **478**:4169–4185
47. Grigorieff N., Harrison S. C. (2011) **Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy** *Curr. Opin. Struc. Biol* **21**:265–273
48. Bepler T., Kelley K., Noble A. J., Berger B. (2020) **Topaz-Denoise: general deep denoising models for cryoEM and cryoET** *Nat. Commun* **11**

49. Emsley P., Lohkamp B., Scott W. G., Cowtan K. (2010) **Features and development of Coot** *Acta Crystallogr. D Biol. Crystallogr* **66**:486–501
50. Adams P. D., et al. (2010) **PHENIX: a comprehensive Python-based system for macromolecular structure solution** *Acta Crystallogr. D Biol. Crystallogr* **66**:213–221
51. Yamashita K., Palmer C. M., Burnley T., Murshudov G. N. (2021) **Cryo-EM single-particle structure refinement and map calculation using Servalcat** *Acta Crystallogr. D Struct. Biol* **77**:1282–1291
52. Chen V. B., et al. (2010) **MolProbity: all-atom structure validation for macromolecular crystallography** *Acta Crystallogr. D Biol. Crystallogr* **66**:12–21
53. Barad B. A., et al. (2015) **EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy** *Nat. Methods* **12**:943–946
54. Pintilie G., et al. (2020) **Measurement of atom resolvability in cryo-EM maps with Q-scores** *Nat. Methods* **17**:328–334
55. Schrödinger L. L. C. (2021) **Schrödinger L. L. C., The PyMOL Molecular Graphics System. Version 2.5.0** Schrödinger L. L. C. (2021).
56. Pettersen E. F., et al. (2004) **UCSF Chimera - A visualization system for exploratory research and analysis** *J. Comput. Chem* **25**:1605–1612
57. Pettersen E. F., et al. (2021) **UCSF ChimeraX: structure visualization for researchers, educators, and developers** *Protein Sci* **30**:70–82
58. Katoh K., Standley D. M. (2013) **MAFFT multiple sequence alignment software version 7: improvements in performance and usability** *Mol. Biol. Evol* **30**:772–780
59. Minh B. Q., et al. (2020) **IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era** *Mol. Biol. Evol* **37**:1530–1534
60. Kalyaanamoorthy S., Minh B. Q., Wong T. K. F., von Haeseler A., Jermini L. S. (2017) **ModelFinder: fast model selection for accurate phylogenetic estimates** *Nat. Methods* **14**:587–589
61. Letunic I., Bork P. (2021) **Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation** *Nucleic Acids Res* **49**:W293–W296
62. Hoang D. T., Chernomor O., von Haeseler A., Minh B. Q., Vinh L. S. (2018) **UFBoot2: improving the ultrafast bootstrap approximation** *Mol. Biol. Evol* **35**:518–522

Editors

Reviewing Editor

Maria Maldonado

University of California Davis, Davis, United States of America

Senior Editor

Amy Andreotti

Iowa State University, Ames, United States of America

Reviewer #1 (Public review):

The authors present the cryo-EM structure of PSI-fucoanthin chlorophyll a/c-binding proteins (FCPs) supercomplex from the diatom *Thalassiosira pseudonana* CCMP1335 at a global resolution of 2.3 Å. This exceptional resolution allows the authors to construct a near-atomic model of the entire supercomplex and elucidate the molecular details of FCPs arrangement. The high-resolution structure reveals subunits not previously identified in earlier reconstructions and models, as well as sequence analysis of PSI-FCPIs from other diatoms and red algae. Additionally, the authors use their model in conjunction with a phylogenetic analysis to compare and contrast the structural features of the *T. pseudonana* supercomplex with those of *Chaetoceros gracilis*, uncovering key structural features that contribute to the efficiency of light energy conversion in diatoms.

The study employs the advanced technique of single particle cryo-electron microscopy to visualize the complex architecture of the PSI supercomplex at near-atomic resolution and analyze the specific roles of FCPs in enhancing photosynthetic performance in diatoms.

Overall, the approach and data are both compelling and of high quality. The paper is well written and will be of wide interest for comprehending the molecular mechanisms of photosynthesis in diatoms. This work provides valuable insights for applications in bioenergy, environmental conservation, plant physiology, and membrane protein structural biology.

<https://doi.org/10.7554/eLife.99858.2.sa1>

Author response:

The following is the authors' response to the original reviews.

Public Reviews:

Reviewer #1 (Public Review):

The authors present the cryo-EM structure of of PSI-fucoanthin chlorophyll a/c-binding proteins (FCPs) supercomplex from the diatom Thalassiosira pseudonana CCMP1335 at a global resolution of 2.3 Å. This exceptional resolution allows the authors to construct a near-atomic model of the entire supercomplex and elucidate the molecular details of FCPs arrangement. The high-resolution structure reveals subunits not previously identified in earlier reconstructions and models, as well as sequence analysis of PSI-FCPIs from other diatoms and red algae. Additionally, the authors use their model in conjunction with a phylogenetic analysis to compare and contrast the structural features of the T. pseudonana supercomplex with those of Chaetoceros gracilis, uncovering key structural features that contribute to the efficiency of light energy conversion in diatoms.

The study employs the advanced technique of single particle cryo-electron microscopy to visualize the complex architecture of the PSI supercomplex at near-atomic resolution and analyze the specific roles of FCPs in enhancing photosynthetic performance in diatoms.

Overall, the approach and data are both compelling and of high quality. The paper is well written and will be of wide interest for comprehending the molecular mechanisms of photosynthesis in diatoms. This work provides valuable insights for applications in

bioenergy, environmental conservation, plant physiology, and membrane protein structural biology.

We thank you very much for your highly positive evaluation and comments on our manuscript.

Reviewer #2 (Public Review):

Summary:

*This manuscript elucidated the cryo-electron microscopic structure of a PSI supercomplex incorporating fucoxanthin chlorophyll a/c-binding proteins (FCPs), designated as PSI-FCPI, isolated from the diatom *Thalassiosira pseudonana* CCMP1335. Combining structural, sequence, and phylogenetic analyses, the authors provided solid evidence to reveal the evolutionary conservation of protein motifs crucial for the selective binding of individual FCPI subunits and provided valuable information about the molecular mechanisms governing the assembly and selective binding of FCPIs in diatoms.*

Strengths:

The manuscript is well-written and presented clearly as well as consistently. The supplemental figures are also of high quality.

Weaknesses:

Only minor comments (provided in recommendations for authors) to help improve the manuscript.

We thank you very much for your highly positive evaluation and comments on our manuscript.

Reviewer #3 (Public Review):

Summary:

*Understanding the structure and function of the photosynthetic machinery is crucial for grasping its mode of action. Photosystem I (PSI) plays a vital role in light-driven electron transfer, which is essential for generating cellular reducing power. A primary strategy to mitigate light and environmental stresses involves incorporating peripheral light-harvesting proteins. Among various lineages, the number of LHCI and their protein and pigment compositions differ significantly in PSI-LHCI structures. However, it is still unclear how LHCI recognize their specific binding sites in the PSI core. This study aims to address this question by obtaining a high-resolution structure of the PSI supercomplex, including fucoxanthin chlorophyll a/c-binding proteins (FCPs), referred to as PSI-FCPI, isolated from the diatom *Thalassiosira pseudonana*. Through structural and sequence analyses, distinct protein-protein interactions are identified at the interfaces between FCPI and PSI subunits, as well as among FCPI subunits themselves.*

Strengths:

The primary strength of this work lies in its superb isolation and structural determination, followed by clear discussion and conclusions. However, the interactions among the protein complexes and their relevance in formulating general rules are not definitively established. While efficiency is a crucial aspect, preventing damage is equally important, and currently, we cannot infer this from the provided structures.

Weaknesses:

The interactions among the protein complexes and their relevance in formulating general rules are not definitively established. While efficiency is a crucial aspect, preventing damage is equally important, and currently, we cannot infer this from the provided structures.

We thank you very much for your highly positive evaluation and comments on our manuscript. This study is aimed to decipher the interactions among different protein subunits within the PSI-FCPI supercomplex, from which we wish to draw their relevance in formulating general rules. While we agree that damage is equally important, it is unclear to us what kind of damage you are mentioning, and we consider that this may need to be treated in another publication, as we cannot elucidate everything in one paper.

Recommendations for the authors:

Reviewer #1 (Recommendations For The Authors):

(1) Line 69: "Diatoms are one of the most important phytoplankton in aquatic environments and contribute to the primary production in the ocean remarkably." Check the sentence, something is missing.

We modified the sentence as follow:

"Diatoms are among the most essential phytoplankton in aquatic environments, playing a crucial role in the global carbon cycle, supporting marine food webs, and contributing significantly to nutrient cycling, thus ensuring the health and sustainability of marine ecosystems"

(2) Supplementary Figure 1B: The SDS-PAGE gel shows multiple bands. Do the authors know the identity of these proteins, or have they considered analyzing the bands using mass spectrometry? The band at ~17 kDa is particularly intense. Could you comment on this? Have you tried running a Native-PAGE gel?

We did not identify protein bands by MS analysis. The protein bands in the PSI-FCPI supercomplex of this diatom have been identified by Ikeda et al. 2013. The protein bands of our sample were similar to those of Ikeda et al. 2013. To explain this, we modified the sentences and cited Ikeda et al. 2013 in the revised manuscript (lines 89-91).

"The PSI-FCPI supercomplexes were purified from the diatom *T. pseudonana* CCMP1335 and analyzed by biochemical and spectroscopic techniques (Fig. S1). Notably, the protein bands of PSI-FCPI closely resembled those reported in a previous study (31)."

The ~17 kDa protein band appears to be FCPIs, which was identified in Ikeda et al. 2013. We did not perform BN-PAGE of this sample; however, we performed trehalose density gradient centrifugation (Fig. S1A).

(3) Can the authors comment on the position of the FCPI subunits in the PSI supercomplex in diatoms compared to the arrangement of LHCI in complex with PSI in cyanobacteria, green algae, and angiosperms? This information would be useful to incorporate into the text.

We previously compared the PSI-FCPI structures of the diatom *C. gracilis* to the PSI-LHCI structures of land plant, green alga, and red alga (Nagao et al., 2020). Also, Xu et al. 2020 compared the *C. gracilis* PSI-FCPI structure to the PSI-LHCI structures of land plant, green

alga, and red alga. The binding sites between FCPs and LHCl are conserved to some extent. However, our recent study revealed that no orthologous relationship exists among LHCl bound to PSI between primitive red algae and diatoms (Kato et al., 2024). Consequently, we found that the information obtained from structural comparisons alone is extremely limited. To avoid misinterpretation, this study focused on comparing the structures and amino acid sequences of FCPs between *T. pseudonana* and *C. gracilis*.

(4) Line 104: Despite achieving high resolution, the authors modeled only six lipid densities (the PDB model contains actually 9 lipids, you should correct it in the text). Do you believe this is due to the detergent used for purification? Can you comment on the position, identity, and potential role of the lipids within your model?

There are 6 lipids associated with the PSI core and 3 with FCP, giving rise to a total of 9 lipids. We have described it in our original text (lines 102-104 in the modified manuscript). Additionally, our structure reveals unidentified densities which likely represent lipids; they are modeled as 88 unknown lipids (UNLs). Thus, there are more lipids in the supercomplex. However, we also observed 4 β -DDM molecules (LMT) in the structure, which are used as detergents. Thus, it is possible that some lipids have dissociated and replaced by detergents. Many of the observed lipids are located between subunits, likely contributing to the stabilization of the complex.

(5) Line 111: The global resolution is very high. Why does the unknown protein have such low resolution that it was impossible to model it properly and perform de novo identification from the density map? Is it due to a lower abundance of particles with this subunit bound? Have you tried improving this with 3D classification/ focus refinement /density modification?

The Unknown subunit (UNK) is located peripherally, and its density is significantly lower compared to the neighboring subunits, which may suggest a low abundance. We applied density modification using Topaz for 3D map denoising, but the effect was minimal. As the low abundance of UNK may be the cause, 3D classification and focus refinement also had limited impact.

(6) Figure 2A: It would be useful to show the density map for the subunit together with the model, especially to demonstrate visualization of the long loop.

We added the model and map of Psa29 to Figure S4C in the revised manuscript.

(7) Given the proximity of Psa29 to PsaC, is the protein involved in electron shuttling? If so, could you comment on this? In line 131, you state that Psa29 was not found in other organisms. Can the authors speculate on the potential role of this protein in diatoms?

We have no idea about the function of Psa29 at present. However, Psa29 does not contain any cofactors, indicating no contribution of it to electron transfer reactions. To understand the function of Psa29, a deletion mutant of this gene is required for examining its functional and physiological roles in diatom photosynthesis. To explain this, we added the following sentences to the revised manuscript (lines 129-133):

"However, the functional and physiological roles of Psa29 remain unclear at present. It is evident that Psa29 does not have any pigments, quinones, or metal complexes, suggesting no contribution of Psa29 to electron transfer reactions within PSI. Further mutagenesis studies will be necessary to investigate the role of Psa29 in diatom photosynthesis."

(8) Line 163: "Among the FCPI subunits, only FCPI-1 has BCRs in addition to Fxs and Ddxs (Figure S6A). FCPI-1 is a RedCAP, which belongs to the LHC protein superfamily but is distinct from the LHC protein family (6, 7)." It would be useful if the authors could add the carotenoid model embedded in the cryoEM density map to the figure to show the features that led to modeling BCR instead of other carotenoids. Additionally, it would be helpful to include in the text why RedCAPs differ from LHCI and their proposed role.

We added the model and map of two BCRs in FCPI-1 (RedCAP) to Figure S4F in the revised manuscript.

Phylogenetic analysis showed that RedCAPs are distinct from the LHC protein family. This has been explained in lines 163-164. Also, the functional and physiological roles of RedCAP remain unclear. To explain this, we added the sentence "; however, the functional and physiological roles of RedCAP remain unclear" to the revised manuscript (lines 164-165).

(9) Line 185: "However, it is unknown (i) whether CgRedCAP is indeed bound to the *C. gracilis* PSI-FCPI supercomplex and (ii) if a loop structure corresponding to the Q96-T116 loop of TpRedCAP exists in CgRedCAP." Have the authors attempted to model the protein using AlphaFold? If so, are there significant differences? Could you speculate on the absence of RedCAP in *C. gracilis*? Do you believe it is due to using a different detergent or related to environmental factors?

We did not model CgRedCAP using AlphaFold. Our recent study "Kato et al. 2024" proposed that CgRedCAP binds to the LHCI-1 site in the PSI-FCPI structure based on sequence comparison. There are two types of PSI-FCPI supercomplexes, one having 16 FCPIs and the other having 24 FCPIs, from *C. gracilis*. The different antenna sizes may depend on the growth conditions of *C. gracilis* (Nagao et al. 2020). These explanations were already described in the manuscript (lines 243-246).

(10) Line 193: Figure 8 is mentioned before Figures 4-7.

We are sorry for the mistake of Figure number. Figure 8 is Supplementary Figure 8, so that we modified Fig. S8B in the revised manuscript.

(11) Line 223: FCPI-4 interacts only with FCPI-5, primarily through the interaction of Y196/4 with the FCPI-5 backbone. Is this interaction facilitated by other factors such as lipids, carotenoids, or other ligands? Also, FCPI-4 occupies a peculiar position compared to other LHCI proteins (it is peripheral to FCPI-4 and FCPI-5). Do you believe this could be due to a transient interaction with the complex? Could the presence of this protein be related to the growth conditions experienced by the plant? Are there any literature reports on environmental conditions influencing FCPI arrangements? Including this information in the text would be interesting.

Y196/4 interacts with only backbones by hydrogen-bond interactions; therefore, other cofactors do not contribute to the interactions.

We do not believe that the interaction of FCPI-4 is transient; rather, this binding appears to be stable within the complex. Given that the PSI-FCPI supercomplexes were isolated by anion exchange chromatography, FCPI-4 and FCPI-5 are tightly associated within this complex. However, it is important to note that the expression of diatom FCPI proteins can indeed vary depending on growth conditions, as highlighted in our previous study (Nagao et al., 2020). While the peculiar position of FCPI-4 may not be directly related to transient interactions, environmental conditions could still influence the overall arrangement and expression levels of FCPIs. This information has already been described in the manuscript (lines 243-246).

(12) Given the high resolution of your map, the overall model quality does not seem to match the map quality. Specifically, the clash score (10) and sidechain outliers (3%) are elevated. Could you comment on this? Do you believe it is related to the high number of ligands?

Our structure contains a total of 295 ligands, including cofactors, detergents, and unknown lipids. We believe the high clash score and number of sidechain outliers are due to the large number of ligands present.

(13) Supplementary Figure 2: You should show the 3D classes that were discarded.

According to your comment, we added the 3D classes that were discarded and the sentence "Red boxes highlight selected particles from each 3D classification." to Figure S2 and its legend in the revised manuscript.

(14) Which masks were used for refinement? How were they generated, and which parameters were chosen? This information should be added to the Materials and Methods section. You should show the masks used during classification, for example.

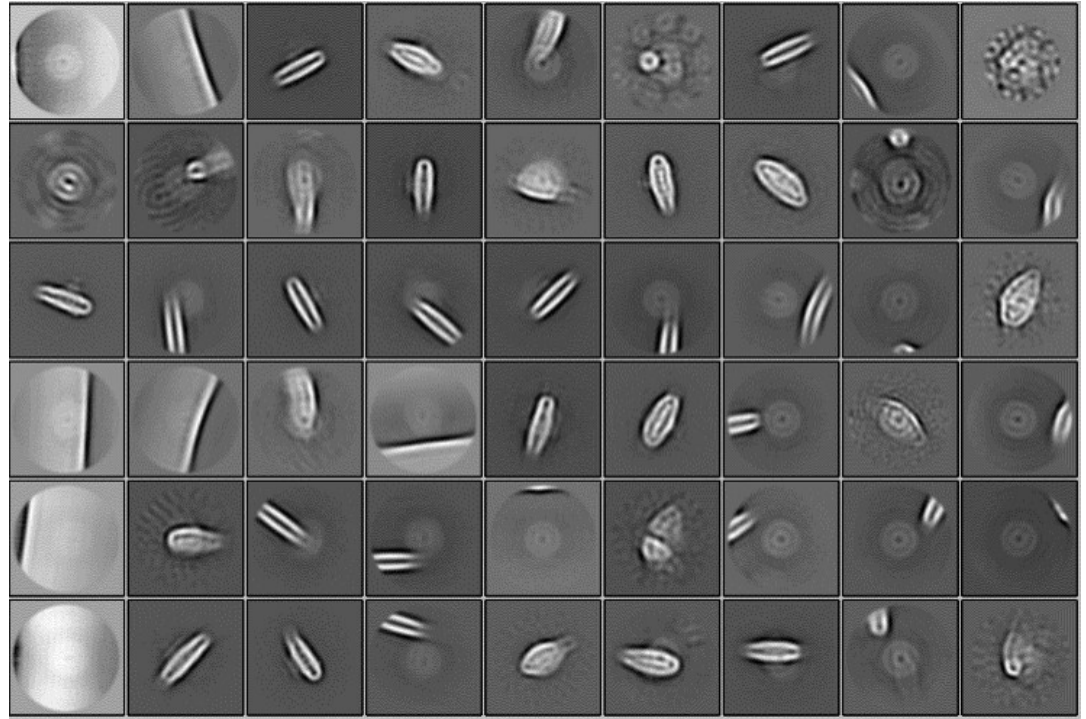
We used a 240 Å spherical mask for refinement and classification, without applying any reference mask as input. To explain this, we added the corresponding sentence to Methods in the revised manuscript (lines 347-348) as follow:

"A 240-Å spherical mask was used during the 3D classification and refinement processes."

(15) Were any extra proteins detected in the early stages of the cryoEM analysis (i.e., 2D classification) that were discarded? Could you visualize the superior oligomeric states of the supercomplex?

In the single-particle analysis, no larger particles than the analyzed complex were detected. The results of 2D classification using a sufficiently large spherical mask with a diameter of 320 Å are shown below.

Author response image 1.



(16) Have you tried using cryoSPARC for data analysis? If so, could you comment on that?

We did not use cryoSPARC for data analysis.

Reviewer #2 (Recommendations For The Authors):

I have some minor comments below to help improve the manuscript. The line numbers below refer to those in the Word version of the manuscript.

(1) Figure 1 legend, line 559, "membrane normal"? Panel A and B, structures with the same colors, do they refer to the closely related or interacted parts? For example, the red color for FCP1-1 in A and PsaA in B. If not, the authors may want to clarify it.

The term 'membrane normal' refers to the direction perpendicular to the surface of a membrane. It is a concept frequently used in physics and biology to describe the orientation relative to the membrane's plane.

We do not refer to either the closely related or interacted parts used in Figure 1. According to your comments, the colors of subunits were revised in the revised manuscript.

(2) Line 109-117. "Psa28 is a novel subunit found in the C. gracilis PSI-FCPI structure, and its name follows the nomenclature as suggested previously (31)... After psaZ, the newly identified genes should be named psa27, psa28, etc., and the corresponding proteins are called Psa27, Psa28, etc... Psa28 was also named PsaR in the PSI-FCPI structure of C. gracilis (16)". It is confusing. Was Psa28 named twice, PsaR and Psa28? It would be helpful to add a simple explanation here.

According to your comment, we modified the sentence as follow (lines 117-118):

" However, Xu et al. named the subunit as PsaR in the PSI-FCPI structure of *C. gracilis* "

(3) Line 134, "One of the Car molecules in PsaJ was identified as ZXT103 in the *T. pseudonana* PSI-FCPI structure but it is BCR112 in the *C. gracilis* PSI-FCPI structure (15)". Figure S4D mentioned BCR863 but did not mention BCR112. Figure S4C, D, it may need better explanations of the colors and labels, and indicate which parts are from *T. pseudonana* or *C. gracilis*.

BCR112 was misnumbered; the correct number is BCR103. In response to your comments, we revised Figure S4C and D by labeling the characteristic pigments in the revised manuscript.

(4) Figure S7, although mentioned in the legend, it would be helpful to label interaction pairs on the figure directly with corresponding colours.

According to your comments, we modified the Figure and legends in the revised manuscript.

(5) Figure 3E, it is better to avoid red/green colours in one figure as some readers may be colour-blind. It would also be helpful to label each FCPI with the same colour as its structure on the figure directly.

According to your comments, we modified Figure 3E in the revised manuscript.

(6) Line 185, "structures similar to the Q96-T116 loop in TpRedCAP found in the present study (Figure 8B)". The authors refer to Figure S8B? I have the same comment for line 186, Figure 8C.

We are sorry for the mistake of Figure number. Figure 8 is Supplementary Figure 8, so we modified it as Fig. S8B in the revised manuscript.

(7) Line 270, "TpLhcq10 cannot bind at the FCPI-2 site". Why not use FCPI-3 for TpLhcq10?

This means that the gene product of TpLhcq10 binds at the FCPI-3 site but not at the other sites such as FCPI-2. To avoid misreading, we modified the sentence as follows:

"TpLhcq10 binds specifically at the FCPI-3 site but not at the other sites such as FCPI-2" (lines 278-279)

Reviewer #3 (Recommendations For The Authors):

I have no technical or conceptual suggestions at the current stage.

Thank you.

<https://doi.org/10.7554/eLife.99858.2.sa0>