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

Koji Kato[1,†], Yoshiki Nakajima[1,†], Jian Xing[2,†], Minoru Kumazawa[2], Haruya Ogawa[1], Jian-Ren Shen[1,*], Kentaro Ifuku[2,*], Ryo Nagao[3,*]

[1]Research Institute for Interdisciplinary Science and Graduate School of Environmental, Life, Natural Science and Technology, Okayama University, Okayama 700-8530, Japan

[2]Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan

[3]Faculty of Agriculture, Shizuoka University, Shizuoka 422-8529, Japan

[†]These authors contributed equally to this work.

*Corresponding Authors:

Jian-Ren Shen, E-mail: shen@cc.okayama-u.ac.jp

Kentaro Ifuku, E-mail: ifuku.kentaro.2m@kyoto-u.ac.jp

Ryo Nagao, E-mail: nagry@shizuoka.ac.jp

Abstract

Light-harvesting complexes (LHCs) in photosynthetic organisms are attached to photosystem I (PSI), forming a PSI-LHCI supercomplex; however, it is unknown how LHCIs recognize their specific binding sites in the PSI core. Here we determine a cryo-electron microscopic structure of a PSI supercomplex containing fucoxanthin chlorophyll *a*/*c*-binding proteins (FCPs), namely PSI-FCPI, isolated from the diatom *Thalassiosira pseudonana* CCMP1335. The PSI-FCPI structure shows five FCPI subunits together with a PSI monomer. The five FCPIs are identified as RedCAP, Lhcr3, Lhcq10, Lhcf10, and Lhcq8. Structural and sequence analyses reveal specific protein-protein interactions at the binding sites between FCPI and PSI subunits and among FCPI subunits. Structural comparisons of PSI-FCPI supercomplexes and phylogenetic analysis of FCPs between *T. pseudonana* and the diatom *Chaetoceros gracilis* unveil the conservation of protein motifs involved in selective bindings of individual FCPI subunits. These findings provide molecular insights into the assembly and selective binding mechanisms of FCPIs.

34 Oxygenic photosynthesis in cyanobacteria, algae, and land plants converts solar energy into chemical
35 energy concomitant with the evolution of oxygen molecules (*1*). The light-energy conversion
36 reactions occur within two multi-subunit membrane protein complexes known as photosystem I and
37 photosystem II (PSI and PSII, respectively), which perform light harvesting, charge separation, and
38 electron transfer reactions (*2-5*). To harvest light energy efficiently, a number of light-harvesting
39 antenna subunits are attached to the periphery of PSI and PSII core complexes and transfer excitation
40 energy to the two types of photosystem cores (*1*). Light-harvesting antennae have a wide variety
41 among photosynthetic organisms in relation to protein sequences and pigment compositions and can
42 be categorized into two major groups: membrane proteins and water-soluble proteins (*1*).

43 The membrane protein category is composed mainly by the light-harvesting complex (LHC)
44 protein superfamily (*6, 7*), which absorbs light energy using chlorophylls (Chls) and carotenoids
45 (Cars). The number and types of Chls and Cars are highly diversified among LHCs, resulting in color
46 variations in photosynthetic organisms, which can be classified into green and red lineages (*8*). The
47 green lineage organisms include green algae and land plants, whereas the red lineage organisms
48 include red algae, diatoms, haptophytes, cryptophytes, and dinoflagellates (*8*). LHCs that are specific
49 to PSI (LHCIs) interact with a eukaryotic PSI monomer, thereby forming a PSI-LHCI supercomplex
50 (*9, 10*), whose structures have been revealed by cryo-electron microscopy (cryo-EM) in a number of
51 eukaryotes (*9, 10*). In the red lineage, the number of LHCIs and their protein and pigment
52 compositions have been found to differ greatly among the PSI-LHCI structures of red algae (*11-14*),
53 a diatom (*15, 16*), a cryptophyte (*17*), and dinoflagellates (*18, 19*).

54 We have recently shown the conservation and diversity of LHCIs among red-lineage algae
55 based on structural and phylogenetic analyses of PSI-LHCI supercomplexes (*14*). This work clearly
56 showed that the binding sites of LHCIs to PSI were conserved to some extent among red-lineage
57 algae but their evolutionary relationship was poor. It is known that LHCIs have similar protein
58 structures, especially their three-transmembrane helices, among photosynthetic organisms
59 irrespective of the green and red lineages (*9, 10*); nevertheless, individual LHCIs have changed their
60 sequences and structures to adapt their binding sites to the PSI cores during assembly of the PSI-

61 LHCI supercomplexes. These observations raise a critical question as to how LHCIs recognize their
62 binding sites in the PSI core.

63 Diatoms are one of the most important phytoplankton in aquatic environments and contribute
64 to the primary production in the ocean remarkably (*20*). Diatoms have unique LHCs called
65 fucoxanthin Chl *a*/*c*-binding proteins (FCPs), whose pigment compositions and amino acid sequences
66 are different from those of LHCs in land plants (*21-23*). Isolation and characterization of PSI-FCPI
67 supercomplexes isolated from the diatom *Chaetoceros gracilis* and their structures have been reported
68 (*15, 16, 24-29*). Kumazawa et al. showed the high diversity of FCPs between two types of diatoms,
69 *C. gracilis* and *Thalassiosira pseudonana* (*30*). *C. gracilis* and *T. pseudonana* have 46 and 44 FCPs,
70 respectively, which were categorized into multiple, closely related subgroups (*30*), and the amino
71 acid sequences of FCPs are not completely identical between the two diatoms. Therefore, comparing
72 FCPIs including their amino acid residues and protein structures at similar binding sites in PSI-FCPIs
73 between the two diatoms may lead to molecular insights into how FCPIs are bound to PSI. However,
74 an overall structure of the *T. pseudonana* PSI-FCPI has not been solved yet.

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

80

81 **Results and discussion**

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

83 The PSI-FCPI supercomplexes were purified from the diatom *T. pseudonana* CCMP1335 as
84 described in the Methods section. Biochemical and spectroscopic analyses showed that this complex
85 is intact with several bands corresponding to FCPI (Fig. S1). Cryo-EM images of the PSI-FCPI
86 supercomplex were obtained by a JEOL CRYO ARM 300 electron microscope operated at 300 kV.

87 The final cryo-EM map was determined at a resolution of 2.30 Å with a C1 symmetry (Fig. S2, S3,
88 Table S1), based on the "gold standard" Fourier shell correlation (FSC) = 0.143 criterion (Fig. S3A).
89 The atomic model of PSI-FCPI was built based on the resultant cryo-EM map (see Methods;
90 Fig. S3, Table S1–S3). The structure reveals a monomeric PSI core associated with five FCPI subunits
91 (Fig. 1A, B). The five FCPI subunits were named FCPI-1 to 5 (Fig. 1A) following the naming of
92 LHCI subunits in the PSI-LHCI structure of *Cyanidium caldarium* RK-1 (NIES-2137) (*14*); in
93 particular, the sites of FCPI-1 and FCPI-2 in the *T. pseudonana* PSI-FCPI structure (Fig. 1A)
94 corresponded to those of LHCI-1 and LHCI-2 in the *C. caldarium* PSI-LHCI structure. The PSI core
95 contains 94 Chls *a*, 18 *β*-carotenes (BCRs), 1 zeaxanthin (ZXT), 3 [4Fe-4S] clusters, 2
96 phylloquinones, and 6 lipid molecules, whereas the five FCPI subunits contain 45 Chls *a*, 7 Chls *c*, 2
97 BCRs, 15 fucoxanthins (Fxs), 7 diadinoxanthins (Ddxs), and 3 lipid molecules (Table S3).
98
99 **Structure of the *T. pseudonana* PSI core**
100 The PSI core contains 12 subunits, 11 of which are identified as PsaA, PsaB, PsaC, PsaD, PsaE, PsaF,
101 PsaI, PsaJ, PsaL, PsaM, and Psa29 (Fig. 1B), whereas the remaining one could not be assigned and
102 was modeled as polyalanines because of insufficient map resolution for identification of this protein
103 (Fig. S4A). The unidentified subunit was named Unknown, which is located at the same site as Psa28
104 in the *C. gracilis* PSI-FCPI (*15*). The protein structure of Unknown is virtually identical to that of
105 Psa28 in the *C. gracilis* PSI-FCPI (Fig. S4B). Psa28 is a novel subunit found in the *C. gracilis* PSI-
106 FCPI structure (*15*), and its name follows the nomenclature as suggested previously (*31*). It is known
107 that the genes encoding various PSI proteins have been designated as *psaA*, *psaB*, etc. PsaZ has been
108 identified in PSI of *Gloeobacter violaceus* PCC 7421 (*32, 33*). After *psaZ*, the newly identified genes
109 should be named *psa27*, *psa28*, etc., and the corresponding proteins are called Psa27, Psa28, etc.
110 Psa27 has been identified in PSI of *Acaryochloris marina* MBIC11017 (*34-36*). Thus, we named the
111 novel subunit as Psa28 (*15*). In contrast, Psa28 was also named PsaR in the PSI-FCPI structure of *C.*
112 *gracilis* (*16*) and the structure of a PSI supercomplex containing alloxanthin Chl *a*/*c*-binding protein
113 (PSI-APCI) from *Chroomonas placoidea* (*17*).

114 Psa29 is newly identified in the *T. pseudonana* PSI-FCPI structure using ModelAngelo (*37*)
115 and the NCBI database (https://www.ncbi.nlm.nih.gov/) (Fig. 2). The subunit corresponding to Psa29
116 was found in the *C. gracilis* PSI-FCPI structures (*15, 16*), which was modeled as polyalanines and
117 named as either Unknown1 (*15*) or PsaS (*16*), respectively. Psa29 shows a unique structure distinct
118 from the other PSI subunits in the *T. pseudonana* PSI-FCPI (Fig. 2A), and engages in multiple
119 interactions with PsaB, PsaC, PsaD, and PsaL at distances of 2.5–3.2 Å (Fig. 2B–G). Based on
120 sequence analyses, Psa29 can exhibit evolutionary divergence between Bacillariophyceae (diatoms)
121 and Bolidophyceae, the latter of which is a sister group of diatoms within Stramenopiles (Fig. 2H);
122 however, it was not found from other organisms. The arrangement of each PSI subunit in the *T.*
123 *pseudonana* PSI-FCPI is virtually identical to that in the *C. gracilis* PSI-FCPI structures already
124 reported (*15, 16*).
125 The number and arrangement of Chls and Cars within the PSI core in the *T. pseudonana* PSI-
126 FCPI structure (Fig. S4C, D) are largely similar to those in the *C. gracilis* PSI-FCPI structure (*15*).
127 However, Chl a102 of PsaI was found in the *T. pseudonana* PSI-FCPI structure but not in the *C.*
128 *gracilis* PSI-FCPI structure (*15*), whereas a845 of PsaA and BCR863 of PsaB were identified in the
129 *C. gracilis* PSI-FCPI structure (*15*) but not in the *T. pseudonana* PSI-FCPI structure. One of the Car
130 molecules in PsaJ was identified as ZXT103 in the *T. pseudonana* PSI-FCPI structure but it is
131 BCR112 in the *C. gracilis* PSI-FCPI structure (*15*).
132
133 **Structure of the *T. pseudonana* FCPIs**
134 Kumazawa et al. classified 44 *Lhc* genes in *T. pseudonana* with their characteristic names of *Lhcf*,
135 *Lhcq*, *Lhcr*, *Lhcx*, and *Lhcz* (*30*). Based on this report, the five FCPI subunits in the PSI-FCPI structure
136 were identified using five genes, namely, *RedCAP*, *Lhcr3*, *Lhcq10*, *Lhcf10*, and *Lhcq8*, at the sites of
137 FCPI-1 to 5, respectively (Fig. 1A). It should be noted that RedCAP is not included in the 44 *Lhc*
138 genes (*30*) but grouped into the LHC protein superfamily (*6, 7*). For the assignments of each FCPI
139 subunit, we focused on characteristic amino acid residues based on their cryo-EM map, especially
140 S61/V62/Q63 in FCPI-1; A70/R71/W72 in FCPI-2; Y64/R65/E66 in FCPI-3; M63/R64/Y65 in FCPI-

141 4; and A62/R63/R64 in FCPI-5 (Fig. S5). The root mean square deviations (RMSDs) of the structures
142 between FCPI-4 and the other four FCPIs range from 1.91 to 3.73 Å (Table S4).
143 Each FCPI subunit binds several Chl and Car molecules, namely, 7 Chls *a*/1 Chl *c*/2 Fxs/3
144 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
145 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. S6, Table
146 S3). The axial ligands of the central Mg atoms of Chls within each FCPI are provided mainly by the
147 main and side chains of amino acid residues (Table S5). Possible excitation-energy-transfer pathways
148 can be proposed based on close physical interactions among Chls between FCPI-3 and PsaA, between
149 FCPI-3 and PsaL, between FCPI-1 and PsaI, and between FCPI-2 and PsaB (Fig. S7).
150
151 **Structural characteristics of RedCAP and its evolutionary implications**
152 Among the FCPI subunits, only FCPI-1 has BCRs in addition to Fxs and Ddxs (Fig. S6A). This is a
153 first report for the binding of BCRs to FCPIs in diatoms. FCPI-1 is a RedCAP, which belongs to the
154 LHC protein superfamily but is distinct from the LHC protein family (*6, 7*). FCPI-1 is located near
155 PsaB, PsaI, and PsaL through protein-protein interactions with these subunits at the stromal and
156 lumenal sides (Fig. 3A). I138 and S139 of FCPI-1 interact with K121, G122, and F125 of PsaL at the
157 stromal side (Fig. 3B), whereas at the lumenal side, multiple interactions are found between I109 of
158 FCPI-1 and F5 of PsaI, between T105/L106/T108 of FCPI-1 and W92/P94/F96 of PsaB, and between
159 E102/W103 of FCPI-1 and S71/I73 of PsaL (Fig. 3C). The protein-protein interactions at the lumenal
160 side (Fig. 3C) appear to be caused by a loop structure of FCPI-1 from Q96 to T116 (pink in Fig. 3D),
161 which is characteristic of FCPI-1 but not present in other four FCPI subunits (pink in Fig. 3E). This
162 loop structure is inserted into a cavity formed by PsaB, PsaI, and PsaL (Fig. 3C, D). These findings
163 indicate that the Q96–T116 loop recognizes the cavity for its binding at the specific FCPI-1 site.
164 RedCAP of *C. gracilis* (CgRedCAP) was not identified in the *C. gracilis* PSI-FCPI structures
165 (*15, 16*). As discussed previously (*14*), we identified CgRedCAP by sequence analysis and suggested
166 the binding of CgRedCAP to the *C. gracilis* PSI core at the site similar to LHCI-1 of the red alga *C.*
167 *caldarium* PSI-LHCI, whose site corresponds to the FCPI-1 site in the PSI-FCPI of *T. pseudonana* in

168 this study. A sequence alignment between RedCAP of *T. pseudonana* (TpRedCAP) and CgRedCAP
169 is shown in Fig. S8A, which shows that CgRedCAP has a sequence similarity of 72% to TpRedCAP
170 and has a protein motif of Q106–I113 (QWGTLATI) corresponding to E102–I109 (EWGTLATI) of
171 TpRedCAP (Fig. 3C). This suggests the possible binding of CgRedCAP to PSI of *C. gracilis* at the
172 position similar to FCPI-1 in the *T. pseudonana* PSI-FCPI structure. However, it is unknown (i)
173 whether CgRedCAP is indeed bound to the *C. gracilis* PSI-FCPI supercomplex and (ii) if a loop
174 structure corresponding the Q96–T116 loop of TpRedCAP exists in CgRedCAP. Further structural
175 study of the *C. gracilis* PSI-FCPI will pave the way for elucidating the mechanism of molecular
176 assembly of diatom RedCAPs.

177 RedCAPs have been found in the structures of PSI-LHCI of the red alga *Porphyridium*
178 *purpureum* (*13*) and a PSI supercomplex with alloxanthin Chl *a*/*c*-binding proteins (PSI-ACPI) of the
179 cryptophyte *Chroomonas placoidea* (*17*), which have been summarized in our previous study (*14*).
180 Both of the RedCAPs of *P. purpureum* (PpRedCAP) and *C. placoidea* (CpRedCAP) display loop
181 structures similar to the Q96–T116 loop in TpRedCAP found in the present study (Fig. 8B). Multiple
182 sequence alignments of TpRedCAP with PpRedCAP and CpRedCAP are shown in Fig. 8C.
183 PpRedCAP and CpRedCAP show sequence similarities of 39% and 60%, respectively, to TpRedCAP.
184 PpRedCAP has a protein motif of V105–L112 (VWGPLAQL), while CpRedCAP has a protein motif
185 of Q117–A124 (QWGPLASA). These two motifs correspond to E102–I109 (EWGTLATI) of
186 TpRedCAP; however, the conservation of sequences is lower between TpRedCAP and
187 PpRedCAP/CpRedCAP than between TpRedCAP and CgRedCAP. Among the four RedCAPs, the
188 four amino acids of Trp, Gly, Leu, and Ala are conserved in the protein motifs (xWGxLAxx),
189 implying that the characteristic loop structure including the conserved protein motifs (xWGxLAxx)
190 contributes to the binding of RedCAP to PSI among the red-lineage algae.

191

192 **Protein-protein interactions of other FCPI subunits**

193 FCPI-2 (Lhcr3) is located near PsaB and PsaM through protein-protein interactions with these
194 subunits at distances of 3.0–4.3 Å at the stromal and lumenal sides (Fig. 4). The amino acid residues

195 of I63/T65/D66/Y69/W134/Y138/D140 in FCPI-2 are associated with
196 W153/L154/K159/F160/W166 in PsaB at the stromal side (Fig. 4B), whereas F116 and F120 of
197 FCPI-2 interact with F5/I9/M12 of PsaM at the lumenal side (Fig. 4C). The amino acid sequences of
198 I63–Y69, F116–F120, and W134–D140 in Lhcr3 are not conserved in the Lhcr subfamily comprising
199 of Lhcr1, Lhcr4, Lhcr7, Lhcr11, Lhcr12, Lhcr14, Lhcr17, Lhcr18, Lhcr19, and Lhcr20 according to
200 Kumazawa et al. (*30*) (Fig. S9).

201 FCPI-3 (Lhcq10) is located near PsaL through protein-protein interactions with it at
202 distances of 2.3–4.2 Å at the stromal side (Fig. 5A, B). The amino acid residues of
203 L126/I130/L142/Y146/W147/V148/W155 in FCPI-3 are associated with F4/K6/P20/S25/L26/L30 in
204 PsaL (Fig. 5B). Because TpLhcq10 is homologous to CgLhcr9 (*30*), we compared the amino acid
205 sequence of Lhcq10 with the Lhcq and Lhcr subfamilies in *T. pseudonana* (Fig. S10A, B). It was
206 shown that the amino acid sequence of L126–W155 in Lhcq10 is not conserved in the Lhcq subfamily
207 consisting of Lhcq1, Lhcq2, Lhcq3, Lhcq4, Lhcq5, Lhcq6, Lhcq7, Lhcq8, and Lhcq9 (Fig. S10A), as
208 well as in the Lhcr subfamily consisting of Lhcr1, Lhcr3, Lhcr4, Lhcr7, Lhcr11, Lhcr12, Lhcr14,
209 Lhcr17, Lhcr18, Lhcr19, and Lhcr20 according to Kumazawa et al. (*30*) (Fig. S10B).

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

215 FCPI-5 (Lhcq8) is located near PsaL and FCPI-4 through protein-protein interactions with
216 them at distances of 2.6–4.1 Å at the stromal and lumenal sides (Fig. 5A, C, D). The amino acid
217 residues of P108/Q109/A112/I115 in FCPI-5 are associated with F134/I137/S141 in PsaL at the
218 lumenal side (Fig. 5D). The interactions of FCPI-5 with FCPI-4 are shown in Fig. 5C. The amino
219 acid sequences of F82–G87 and P107–I115 in Lhcq8 are not conserved in the Lhcq subfamily
220 consisting of Lhcq1, Lhcq2, Lhcq3, Lhcq4, Lhcq5, Lhcq6, Lhcq7, Lhcq9, and Lhcq10 according to
221 Kumazawa et al.(*30*) (Fig. S12A, B).

222

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

224 To evaluate the molecular assembly of FCPI subunits in the *T. pseudonana* PSI-FCPI structure, we
225 focused on both protein-protein interactions based on their close distances (Fig. 3–5) and amino acid
226 residues of non-conserved region among 44 FCPs (Fig. S9–S12). This is because selective
227 associations of FCPIs with PSI require specific amino acid residues of each FCPI. The protein-protein
228 interactions among FCPI subunits and between FCPI and PSI subunits exist at the stromal and
229 lumenal sides (Fig. 3–5) and appear to be recognized by unique amino acid residues of FCPIs, which
230 are not conserved in each LHC subfamily (Fig. S9–S12). Thus, the binding and assembly of each
231 FCPI subunit to PSI is likely determined based on amino acid sequences among 44 FCPs in *T.*
232 *pseudonana*, especially in their loop regions.

233 The diatom *C. gracilis* has shown two different PSI-FCPI structures, one with 16 FCPI
234 subunits (*15*) and the other with 24 FCPI subunits (*16*). This is because of alterations of antenna sizes
235 of FCPIs in the *C. gracilis* PSI-FCPI supercomplexes in response to growth conditions, especially
236 CO_2 concentrations and temperatures (*38*). The *C. gracilis* PSI-FCPI structure has five FCPI subunits
237 at the same binding sites of FCPI-1 to 5 in the *T. pseudonana* PSI-FCPI structure (Fig. 6A). A
238 relationship of the *Lhc* genes encoding FCPs and RedCAP with the binding positions of FCPI-1 to 5
239 in *T. pseudonana* and *C. gracilis* is summarized in Fig. 6B. It should be noted that the gene names of
240 the *C. gracilis* FCPIs are based on Kumazawa et al. (*30*) as discussed in our recent study (*14*).

241 Phylogenetic analysis clearly showed that at the sites of FCPI-1, 2, 3, and 5 in the *T.*
242 *pseudonana* PSI-FCPI structure, TpRedCAP, TpLhcr3, TpLhcq10, and TpLhcq8 are orthologous to
243 CgRedCAP, CgLhcr1, CgLhcr9, and CgLhcq12, respectively (Fig. 6C). The characteristic protein
244 loops of TpRedCAP and CpRedCAP appear to be involved in interactions with PSI at the FCPI-1 site,
245 as described above (Fig. S8) At the FCPI-2 site, comparative analyses revealed that the amino acid
246 residues facilitating interactions between TpLhcr3 and TpPsaB/TpPsaM closely parallel those
247 observed in the CgLhcr1-CgPsaB and CgLhcr1-CgPsaM pairs (Fig. S13). Similarly, a high degree of
248 similarity characterized the residues involved in the interaction pairs of TpLhcq10-TpPsaL/CgLhcr9-

249 CgPsaL at the FCPI-3 site (Fig. S14A, B) and TpLhcq8-TpPsaL/CgLhcq12-CgPsaL at the FCPI-5
250 site (Fig. S14C, D). However, TpLhcf10 was not homologous to CgLhcf3 (Fig. 6C), both of which
251 are located at the FCPI-4 site in each PSI-FCPI structure (Fig. 6A). Thus, the two diatoms appear to
252 possess both a conserved mechanism of protein-protein interactions across characteristic protein
253 motifs between FCPI and PSI subunits and a different interaction mechanism among FCPIs.
254 It is of note that the PSI-FCPI structure of *C. gracilis* binds much more FCPI subunits than
255 that of *T. pseudonana*, namely, 16 or 24 subunits in *C. gracilis* found in the previous studies (*15, 16*)
256 vs. 5 subunits in *T. pseudonana* found in the present study. We do not know the reason for this
257 difference at present. It is possible that some of the FCPI subunits are released during detergent
258 solubilization in *T. pseudonana*, but they are reserved in *C. gracilis*. Alternatively, it is also possible
259 that the number of FCPI subunits is inherently less than that of *C. gracilis* probably due to the
260 differences in their living environments. Further studies will clarify this question.
261
262 **Extension to molecular assembly of PSI-LHCI supercomplexes**
263 The mechanisms of protein-protein interactions for diatom PSI-FCPI supercomplexes are likely
264 developed by specific bindings of FCPs selected from 44 TpFCPs and 46 CgFCPs in addition to
265 RedCAPs. Like a lock-and-key mechanism, one FCP cannot be substituted by the other FCPs to form
266 the PSI-FCPI supercomplexes in the two diatoms, e.g., TpLhcq10 cannot bind at the FCPI-2 site. The
267 selective binding mechanism of FCPIs may dictate the fate of the molecular assembly of PSI-FCPI.
268 It is important to note that the selective bindings of FCPIs are found for the first time by comparing
269 structures of PSI-FCPI supercomplexes and amino acid sequences of FCPIs in the two diatom species.
270 This approach can be extended to the LHC protein superfamily in the green and red lineages, so that
271 comparing protein structures and sequences of PSI-LHCI supercomplexes among closely related
272 species lays the foundation for elucidating the general mechanism of molecular assembly of PSI-
273 LHCI supercomplexes. Thus, this study will shed light on answering the evolutionary question as to
274 how LHCIs recognize their binding sites to PSI in photosynthetic organisms.
275

276 **Methods**

277 **Cell growth and preparation of thylakoid membranes**

278 The marine centric diatom, *T. pseudonana* CCMP1335, was grown in artificial seawater containing
279 sodium metasilicate and KW21 (*39*) at 20 °C at a photosynthetic photon flux density of 30 μmol
280 photons m^{-2} s^{-1} provided by white LED with bubbling of air containing 3% (v/v) CO_2. The cells were
281 harvested by centrifugation, and then disrupted by agitation with glass beads (*40*), followed by
282 centrifugation to pellet the thylakoid membranes. The thylakoid membranes thus prepared were
283 suspended with a 50 mM Mes-NaOH (pH 6.5) buffer containing 1 M betaine and 1 mM EDTA.

284

285 **Purification of the PSI-FCPI supercomplex**

286 Thylakoid membranes were solubilized with 1% (w/v) *n*-dodecyl-*β*-D-maltoside (*β*-DDM) at a Chl
287 concentration of 0.5 mg mL^{-1} for 20 min on ice in the dark with gentle stirring. After centrifugation
288 at 162,000 × g for 20 min at 4°C, the resultant supernatant was loaded onto a Q-Sepharose anion-
289 exchange column (1.6 cm of inner diameter and 25 cm of length) equilibrated with a 20 mM MES-
290 NaOH (pH 6.5) buffer containing 0.2 M trehalose, 5 mM $CaCl_2$, and 10 mM $MgCl_2$, and 0.03% *β*-
291 DDM (buffer A). The column was washed with buffer A until the eluate became colorless. Elution
292 was performed at a flow rate of 1.0 mL min^{-1} using a linear gradient of buffer A and buffer B (buffer
293 A plus 500 mM NaCl): 0–600 min, 0–60% buffer B; 600–800 min, 60–100% buffer B; 800–900 min,
294 100% buffer B. The fraction enriched in PSI-FCPI was eluted at 194–247 mM NaCl, which was
295 collected and loaded onto a linear gradient containing 10–40% (w/v) trehalose in a medium of 20 mM
296 MES-NaOH (pH 6.5), 5 mM $CaCl_2$, 10 mM $MgCl_2$, 100 mM NaCl, and 0.03% *β*-DDM. After
297 centrifugation at 154,000 × g for 18 h at 4°C (P40ST rotor; Hitachi), a green fraction (fig. S1A) was
298 collected and concentrated using a 150 kDa cut-off filter (Apollo; Orbital Biosciences) at 4,000 × g.
299 The concentrated samples were stored in liquid nitrogen until use.

300

301 **Biochemical and spectroscopic analyses of the PSI-FCPI supercomplex**

302 The polypeptide bands of PSI-FCPI were analyzed by SDS-polyacrylamide gel electrophoresis
303 (PAGE) containing 16% (w/v) acrylamide and 7.5 M urea according to Ikeuchi and Inoue (*41*) (Fig.
304 S1B). The PSI-FCPI supercomplexes (4 μg of Chl) were solubilized by 3% lithium lauryl sulfate and
305 75 mM dithiothreitol for 10 min at 60°C, and loaded onto the gel. A standard molecular weight marker
306 (SP-0110; APRO Science) was used. The absorption spectrum of PSI-FCPI was measured at room
307 temperature using a UV-Vis spectrophotometer (UV-2450; Shimadzu) (Fig. S1C), and the
308 fluorescence-emission spectrum of PSI-FCPI was measured at 77 K upon excitation at 430 nm using
309 a spectrofluorometer (RF-5300PC; Shimadzu) (Fig. S1D). The pigment composition of PSI-FCPI
310 was analyzed by HPLC according to Nagao et al. (*42*), and the elution profile was monitored at 440
311 nm (Fig. S1E).
312
313 **Cryo-EM data collection**
314 A 3-μL aliquot of the *T. pseudonana* PSI-FCPI supercomplex (3.0 mg of Chl mL^{-1}) in a 20 mM MES-
315 NaOH (pH 6.5) buffer containing 0.5 M betaine, 5 mM $CaCl_2$, 10 mM $MgCl_2$, and 0.03% *β*-DDM
316 was applied to Quantifoil R1.2/1.3 Cu 300 mesh grids in the chamber of FEI Vitrobot Mark IV
317 (Thermo Fisher Scientific). Then, the grid was blotted with a filter paper for 4 sec at 4°C under 100%
318 humidity and plunged into liquid ethane cooled by liquid nitrogen. The frozen grid was transferred
319 into a CRYO ARM 300 electron microscope (JEOL) equipped with a cold-field emission gun
320 operated at 300 kV. All image stacks were collected from 5 × 5 holes per stage adjustment to the
321 central hole and image shifts were applied to the surrounding holes while maintaining an axial coma-
322 free condition. The images were recorded with an in-column energy filter with a slit width of 20 eV
323 and at a nominal magnification of × 60,000 on a direct electron detector (Gatan K3, AMETEK). The
324 nominal defocus range was −1.8 to −1.2 μm. Physical pixel size corresponded to 0.752 Å. Each image
325 stack was exposed at a dose rate of 21.46 $e^{-}Å^{-2}sec^{-1}$ for 2.33 sec in CDS mode with dose-fractionated
326 50 movie frames. In total 8,950 image stacks were collected.
327
328 **Cryo-EM image processing**

329 The resultant movie frames were aligned and summed using MotionCor2 (*43*) to yield dose-weighted
330 images. Estimation of the contrast transfer function (CTF) was performed using CTFFIND4 (*44*). All
331 of the following processes were performed using RELION-4.0 (*45*). In total 2,733,572 particles were
332 automatically picked up and used for reference-free 2D classification. Then, 1,132,721 particles were
333 selected from good 2D classes and subsequently subjected to 3D classification without any symmetry.
334 An initial model for the first 3D classification was generated *de novo* from 2D classification. As
335 shown in Fig. S2C, the final PSI-FCPI structure was reconstructed from 75,667 particles. The overall
336 resolution of the cryo-EM map was estimated to be 2.30 Å by the gold-standard FSC curve with a
337 cut-off value of 0.143 (Fig. S3A) (*46*). Local resolutions were calculated using RELION (Fig. S3C).
338

339 **Model building and refinement**

340 Two types of the cryo-EM maps were used for the model building of the PSI-FCPI supercomplex:
341 one was a postprocessed map, and the other was a denoised map using Topaz version 0.2.4 (*47*). The
342 postprocessed map was denoised using the trained model in 100 epochs with two half-maps. Initial
343 models of each subunit in the PSI-FCPI supercomplex were built by ModelAngelo (*37*), and then
344 their structures were inspected and manually adjusted against the maps with Coot (*48*). Each model
345 was built based on interpretable features from the density maps at a contour level of 2.5 σ in the
346 denoised and postprocessed maps. For the assignment of Chls, Chls *a* and *c* were distinguished by
347 inspection of the density map corresponding to the phytol chain with the above thresholds, which was
348 found to be the least level not to link the map of Chls with that of noise. All Chls *c* were assigned as
349 Chl *c*1, because of difficulty in distinction between Chl *c*1 and Chl *c*2 at the present resolution. For
350 the assignment of Cars, Fx and Ddx were distinguished based on the density covering the head group
351 of carotenoids with the above threshold. The PSI-FCPI structure was refined with
352 phenix.real_space_refine (*49*) and Servalcat (*50*) with geometric restraints for the protein-cofactor
353 coordination. The final model was validated with MolProbity (*51*), EMRinger (*52*), and *Q*-score (*53*).
354 The statistics for all data collection and structure refinement are summarized in Table S1, S2. All
355 structural figures were made by PyMOL (*54*), UCSF Chimera (*55*), and UCSF ChimeraX (*56*).

Since the numbering of Chls, Cars, and other cofactors in this paper were different from those of the PDB data, we listed the relationship of their numbering in this paper with those in the PDB data 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 (*57*). The alignment was trimmed using ClipKit v1.4.1 with smart-gap mode. The phylogenetic tree was inferred using IQ-TREE 2 (*58*) with the model selected by ModelFinder (*59*). The tree was visualized by iTOL v6 (*60*). Ultrafast bootstrap approximation was performed with 1000 replicates (*61*).

Data availability

Atomic coordinate and cryo-EM maps for the reported structure have been deposited in the Protein Data Bank under an accession code 8XLS [https://www.rcsb.org/structure/8XLS] and in the Electron Microscopy Data Bank under an accession code EMD-38457 [https://www.ebi.ac.uk/emdb/EMD-38457].

References

1. R. E. Blankenship, *Molecular Mechanisms of Photosynthesis*. (Wiley-Blackwell, Oxford, U.K., ed. 3rd, 2021).
2. J. H. Golbeck, Structure and function of photosystem I. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* **43**, 293-324 (1992).
3. K. Brettel, W. Leibl, Electron transfer in photosystem I. *Biochim. Biophys. Acta, Bioenerg.* **1507**, 100-114 (2001).
4. J.-R. Shen, The structure of photosystem II and the mechanism of water oxidation in photosynthesis. *Annu. Rev. Plant Biol.* **66**, 23-48 (2015).
5. D. Shevela, J. F. Kern, G. Govindjee, J. Messinger, Solar energy conversion by photosystem II: principles and structures. *Photosynth. Res.* **156**, 279-307 (2023).
6. J. Engelken, H. Brinkmann, I. Adamska, Taxonomic distribution and origins of the extended LHC (light-harvesting complex) antenna protein superfamily. *BMC Evol. Biol.* **10**, 233 (2010).
7. S. Sturm *et al.*, A novel type of light-harvesting antenna protein of red algal origin in algae with secondary plastids. *BMC Evol. Biol.* **13**, 159 (2013).
8. P. G. Falkowski *et al.*, The evolution of modern eukaryotic phytoplankton. *Science* **305**, 354-360 (2004).
9. M. Hippler, N. Nelson, The plasticity of photosystem I. *Plant Cell Physiol.* **62**, 1073-1081 (2021).
10. J.-R. Shen, in *Macromolecular Protein Complexes IV. Subcellular Biochemistry,* J. R. Harris, J. Marles-Wright, Eds. (Springer, Cham., 2022), vol. 99, pp. 351-377.
11. X. Pi *et al.*, 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 (2018).
12. M. Antoshvili, I. Caspy, M. Hippler, N. Nelson, Structure and function of photosystem I in *Cyanidioschyzon merolae*. *Photosynth. Res.* **139**, 499-508 (2019).
13. X. You *et al.*, In situ structure of the red algal phycobilisome-PSII-PSI-LHC megacomplex. *Nature* **616**, 199-206 (2023).
14. K. Kato *et al.*, 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**, e2319658121 (2024).
15. R. Nagao *et al.*, Structural basis for assembly and function of a diatom photosystem I-light-harvesting supercomplex. *Nat. Commun.* **11**, 2481 (2020).
16. C. Xu *et al.*, Structural basis for energy transfer in a huge diatom PSI-FCPI supercomplex. *Nat. Commun.* **11**, 5081 (2020).
17. L.-S. Zhao *et al.*, Structural basis and evolution of the photosystem I-light-harvesting supercomplex of cryptophyte algae. *Plant Cell* **35**, 2449-2463 (2023).
18. X. Li *et al.*, Structures and organizations of PSI-AcpPCI supercomplexes from red tidal and coral symbiotic photosynthetic dinoflagellates. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2315476121 (2024).
19. L.-S. Zhao *et al.*, Architecture of symbiotic dinoflagellate photosystem I-light-harvesting

supercomplex in *Symbiodinium*. *Nat. Commun.* **15**, 2392 (2024).
20. C. B. Field, M. J. Behrenfeld, J. T. Randerson, P. Falkowski, Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **281**, 237-240 (1998).
21. B. R. Green, D. G. Durnford, The chlorophyll-carotenoid proteins of oxygenic photosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 685-714 (1996).
22. C. Büchel, Light harvesting complexes in chlorophyll *c*-containing algae. *Biochim. Biophys. Acta Bioenerg.* **1861**, 148027 (2020).
23. W. Wang, J.-R. Shen, in *Photosynthesis: molecular approaches to solar energy conversion,* J.-R. Shen, K. Satoh, S. I. Allakhverdiev, Eds. (Springer, Cham., 2021), pp. 163-194.
24. Y. Ikeda *et al.*, Photosystem I complexes associated with fucoxanthin-chlorophyll-binding proteins from a marine centric diatom, *Chaetoceros gracilis*. *Biochim. Biophys. Acta, Bioenerg.* **1777**, 351-361 (2008).
25. R. Nagao, K. Kagatani, Y. Ueno, J.-R. Shen, S. Akimoto, Ultrafast excitation energy dynamics in a diatom photosystem I-antenna complex: a femtosecond fluorescence upconversion study. *J. Phys. Chem. B* **123**, 2673-2678 (2019).
26. R. Nagao *et al.*, 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 (2019).
27. R. Nagao, M. Yokono, Y. Ueno, J.-R. Shen, S. Akimoto, 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 (2019).
28. R. Nagao, M. Yokono, Y. Ueno, J.-R. Shen, S. Akimoto, pH-sensing machinery of excitation energy transfer in diatom PSI-FCPI complexes. *J. Phys. Chem. Lett.* **10**, 3531-3535 (2019).
29. R. Nagao, M. Yokono, Y. Ueno, J.-R. Shen, S. Akimoto, Excitation-energy transfer and quenching in diatom PSI-FCPI upon P700 cation formation. *J. Phys. Chem. B* **124**, 1481-1486 (2020).
30. M. Kumazawa *et al.*, Molecular phylogeny of fucoxanthin-chlorophyll *a*/*c* proteins from *Chaetoceros gracilis* and Lhcq/Lhcf diversity. *Physiol. Plant.* **174**, e13598 (2022).
31. Y. Kashino *et al.*, Proteomic analysis of a highly active photosystem II preparation from the cyanobacterium *Synechocystis* sp. PCC 6803 reveals the presence of novel polypeptides. *Biochemistry* **41**, 8004-8012 (2002).
32. H. Inoue *et al.*, Unique constitution of photosystem I with a novel subunit in the cyanobacterium *Gloeobacter violaceus* PCC 7421. *FEBS Lett.* **578**, 275-279 (2004).
33. K. Kato *et al.*, Structural basis for the absence of low-energy chlorophylls in a photosystem I trimer from *Gloeobacter violaceus*. *eLife* **11**, e73990 (2022).
34. T. Tomo *et al.*, Characterization of highly purified photosystem I complexes from the chlorophyll *d*-dominated cyanobacterium *Acaryochloris marina* MBIC 11017. *J. Biol. Chem.* **283**, 18198-18209 (2008).
35. T. Hamaguchi *et al.*, Structure of the far-red light utilizing photosystem I of *Acaryochloris marina*. *Nat. Commun.* **12**, 2333 (2021).

453 36. C. Xu *et al.*, A unique photosystem I reaction center from a chlorophyll *d*-containing
454 cyanobacterium *Acaryochloris marina*. *J. Integr. Plant Biol.* **63**, 1740-1752 (2021).
455 37. K. Jamali *et al.*, Automated model building and protein identification in cryo-EM maps.
456 *Nature*, (2024).
457 38. R. Nagao, Y. Ueno, S. Akimoto, J.-R. Shen, Effects of CO_2 and temperature on photosynthetic
458 performance in the diatom *Chaetoceros gracilis*. *Photosynth. Res.* **146**, 189-195 (2020).
459 39. R. Nagao *et al.*, Isolation and characterization of oxygen-evolving thylakoid membranes and
460 Photosystem II particles from a marine diatom *Chaetoceros gracilis*. *Biochim. Biophys. Acta,*
461 *Bioenerg.* **1767**, 1353-1362 (2007).
462 40. R. Nagao, M. Yamaguchi, S. Nakamura, H. Ueoka-Nakanishi, T. Noguchi, Genetically
463 introduced hydrogen bond interactions reveal an asymmetric charge distribution on the radical
464 cation of the special-pair chlorophyll P680. *J. Biol. Chem.* **292**, 7474-7486 (2017).
465 41. M. Ikeuchi, Y. Inoue, A new photosystem II reaction center component (4.8 kDa protein)
466 encoded by chloroplast genome. *FEBS Lett.* **241**, 99-104 (1988).
467 42. R. Nagao, M. Yokono, S. Akimoto, T. Tomo, High excitation energy quenching in fucoxanthin
468 chlorophyll *a*/*c*-binding protein complexes from the diatom *Chaetoceros gracilis*. *J. Phys.*
469 *Chem. B* **117**, 6888-6895 (2013).
470 43. S. Q. Zheng *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved
471 cryo-electron microscopy. *Nat. Methods* **14**, 331-332 (2017).
472 44. J. A. Mindell, N. Grigorieff, Accurate determination of local defocus and specimen tilt in
473 electron microscopy. *J. Struct. Biol.* **142**, 334-347 (2003).
474 45. D. Kimanius, L. Dong, G. Sharov, T. Nakane, S. H. W. Scheres, New tools for automated cryo-
475 EM single-particle analysis in RELION-4.0. *Biochem. J.* **478**, 4169-4185 (2021).
476 46. N. Grigorieff, S. C. Harrison, Near-atomic resolution reconstructions of icosahedral viruses
477 from electron cryo-microscopy. *Curr. Opin. Struc. Biol.* **21**, 265-273 (2011).
478 47. T. Bepler, K. Kelley, A. J. Noble, B. Berger, Topaz-Denoise: general deep denoising models
479 for cryoEM and cryoET. *Nat. Commun.* **11**, 5208 (2020).
480 48. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of *Coot*. *Acta*
481 *Crystallogr. D Biol. Crystallogr.* **66**, 486-501 (2010).
482 49. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular
483 structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221 (2010).
484 50. K. Yamashita, C. M. Palmer, T. Burnley, G. N. Murshudov, Cryo-EM single-particle structure
485 refinement and map calculation using Servalcat. *Acta Crystallogr. D Struct. Biol.* **77**, 1282-
486 1291 (2021).
487 51. V. B. Chen *et al.*, MolProbity: all-atom structure validation for macromolecular
488 crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12-21 (2010).
489 52. B. A. Barad *et al.*, EMRinger: side chain-directed model and map validation for 3D cryo-
490 electron microscopy. *Nat. Methods* **12**, 943-946 (2015).
491 53. G. Pintilie *et al.*, Measurement of atom resolvability in cryo-EM maps with *Q*-scores. *Nat.*
492 *Methods* **17**, 328-334 (2020).

54. L. L. C. Schrödinger, The PyMOL Molecular Graphics System. Version 2.5.0., (2021).
55. E. F. Pettersen *et al.*, UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612 (2004).
56. E. F. Pettersen *et al.*, UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70-82 (2021).
57. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780 (2013).
58. B. Q. Minh *et al.*, IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530-1534 (2020).
59. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587-589 (2017).
60. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293-W296 (2021).
61. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518-522 (2018).

510 **Acknowledgements**

511 We thank Kumiyo Kato and Satoko Kakiuchi for their assistance in this study. The cells of *T.*
512 *pseudonana* CCMP1335 were given by Prof. Yusuke Matsuda, Kwansei Gakuin University, Japan.
513 Cryo-EM data was obtained using EM01CT and EM02CT of SPring-8 with the approval of the Japan
514 Synchrotron Radiation Research Institute (JASRI Proposal No. 2022B2728 (J.-R.S.) and No.
515 2023A2715 (Y.N.)). This work was supported by JSPS KAKENHI grant Nos. JP22KJ2017 (M.K.),
516 JP23K14211 (Y.N.), JP22H04916 (J.-R.S.), JP23H02347 (K.I.), and JP23H02423 (R.N.), Takeda
517 Science Foundation (R.N., K.K.), and Research Support Project for Life Science and Drug Discovery
518 (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED
519 under support No. 4176 (J.-R.S.).

520

521 **Author Contributions**

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

528

529 **Declaration of competing interest**

530 The authors declare no conflict of interest.

531

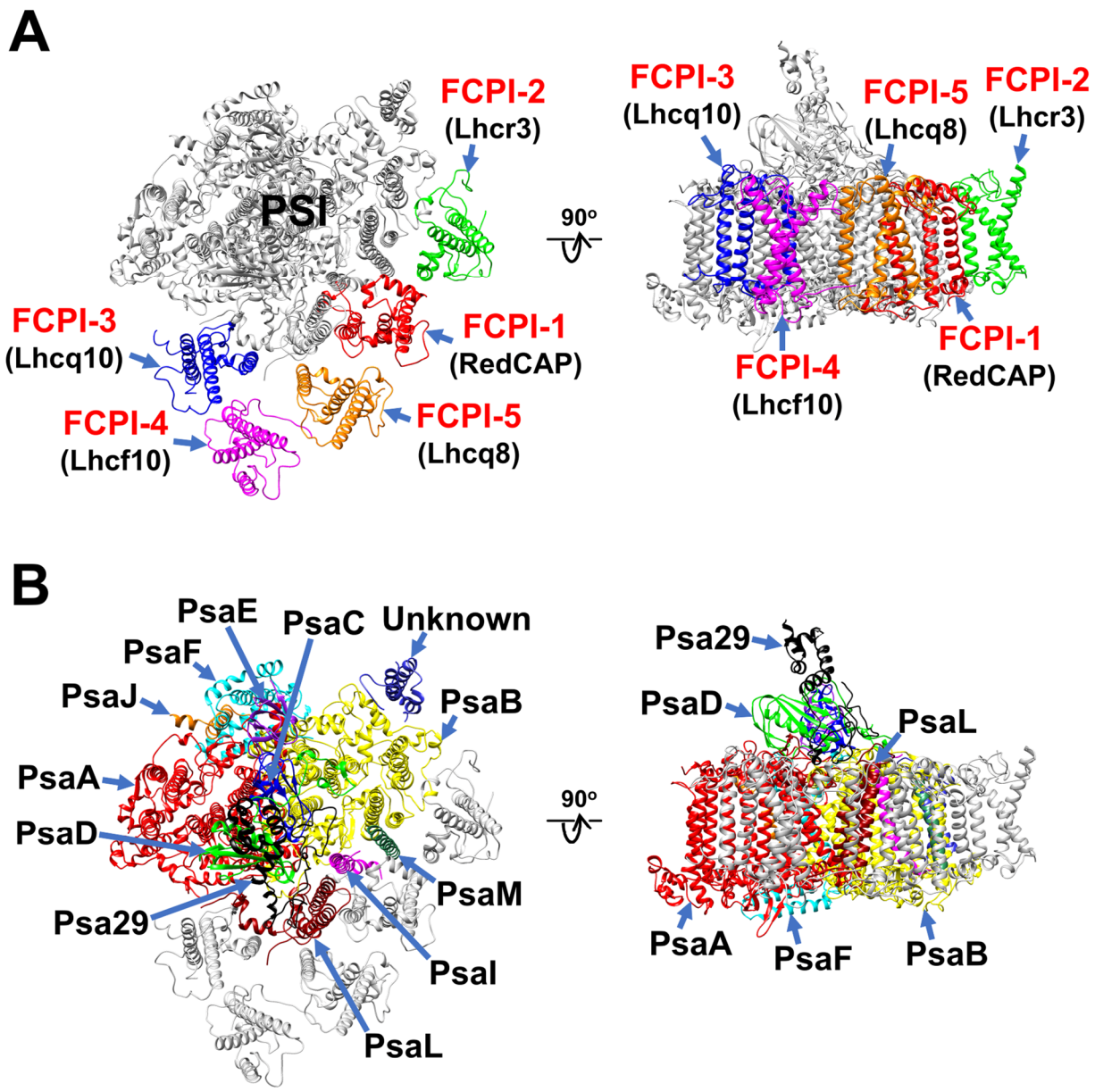


Fig. 1 | Overall structure of the PSI-FCPI supercomplex from *T. pseudonana*. Structures are viewed from the stromal side (left panels) and the direction perpendicular to the membrane normal (right panels). Only protein structures are shown, and cofactors are omitted for clarity. The FCPI (**A**) and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

538

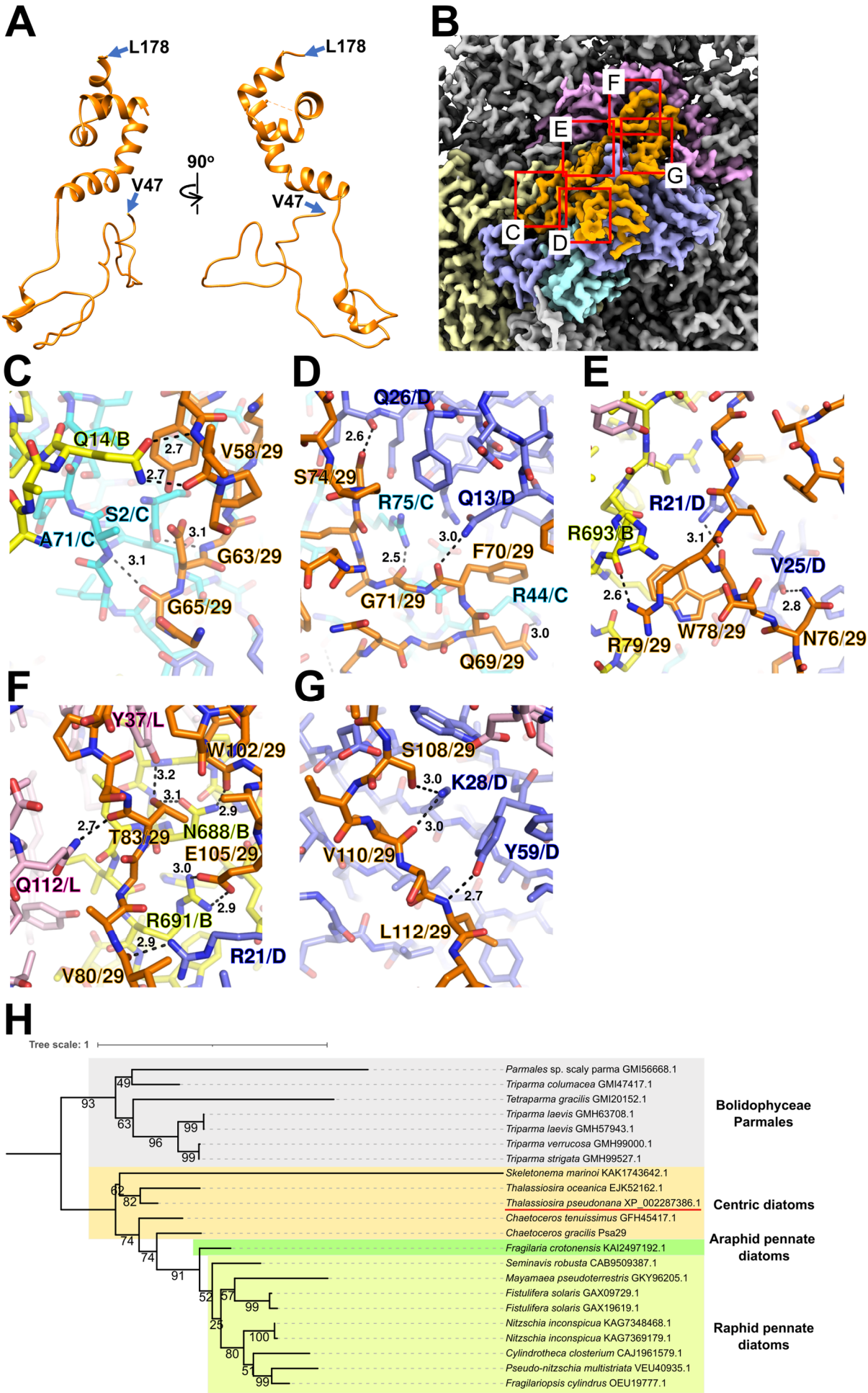

A
L178
V47
90°
L178
V47
B
F
E
G
C
D
C
Q14/B
2.7
V58/29
2.7
S2/C
A71/C
3.1
3.1
G63/29
G65/29
D
Q26/D
2.6
S74/29
R75/C
Q13/D
3.0
2.5
F70/29
G71/29
R44/C
3.0
Q69/29
E
R21/D
R693/B
3.1
V25/D
2.6
2.8
R79/29
W78/29
N76/29
F
Y37/L
W102/29
3.2
3.1
2.9
2.7
T83/29
N688/B
E105/29
3.0
Q112/L
2.9
R691/B
2.9
R21/D
V80/29
G
S108/29
3.0
K28/D
3.0
V110/29
Y59/D
2.7
L112/29
H
Tree scale: 1
Parmales sp. scaly parma GMI56668.1
Triparma columacea GMI47417.1
Tetraparma gracilis GMI20152.1
Triparma laevis GMH63708.1
Triparma laevis GMH57943.1
Triparma verrucosa GMH99000.1
Triparma strigata GMH99527.1
Skeletonema marinoi KAK1743642.1
Thalassiosira oceanica EJK52162.1
Thalassiosira pseudonana XP_002287386.1
Chaetoceros tenuissimus GFH45417.1
Chaetoceros gracilis Psa29
Fragilaria crotonensis KAI2497192.1
Seminavis robusta CAB9509387.1
Mayamaea pseudoterrestris GKY96205.1
Fistulifera solaris GAX09729.1
Fistulifera solaris GAX19619.1
Nitzschia inconspicua KAG7348468.1
Nitzschia inconspicua KAG7369179.1
Cylindrotheca closterium CAJ1961579.1
Pseudo-nitzschia multistriata VEU40935.1
Fragilariopsis cylindrus OEU19777.1
Bolidophyceae Parmales
Centric diatoms
Araphid pennate diatoms
Raphid pennate diatoms
93
49
63
96
99
99
62
82
74
74
91
52
57
99
25
100
80
51
99

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 surrounded environments viewed from the stromal side. The red squared areas are enlarged in panels **C–G**. Yellow, PsaB; cyan, PsaC; blue, PsaD; magenta, PsaL; orange, Psa29. (**C–G**) Protein-protein interactions of Psa29 with PsaB/PsaC (**C**), PsaC/PsaD (**D**), PsaB/PsaD (**E**), PsaB/PsaD/PsaL (**F**), and PsaD (**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 means Ala71 of PsaC. B, PsaB; C, PsaC; D, PsaD, L, PsaL; 29, Psa29. (**H**) Phylogenetic analysis of Psa29 in photosynthetic organisms. A maximum-likelihood tree of the Psa29 proteins was inferred by IQ-TREE v2.2.2.7 using WAG+F+G4 model and trimmed alignment of 22 sequences with 245 amino-acid residues. Numbers at nodes are 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 line.

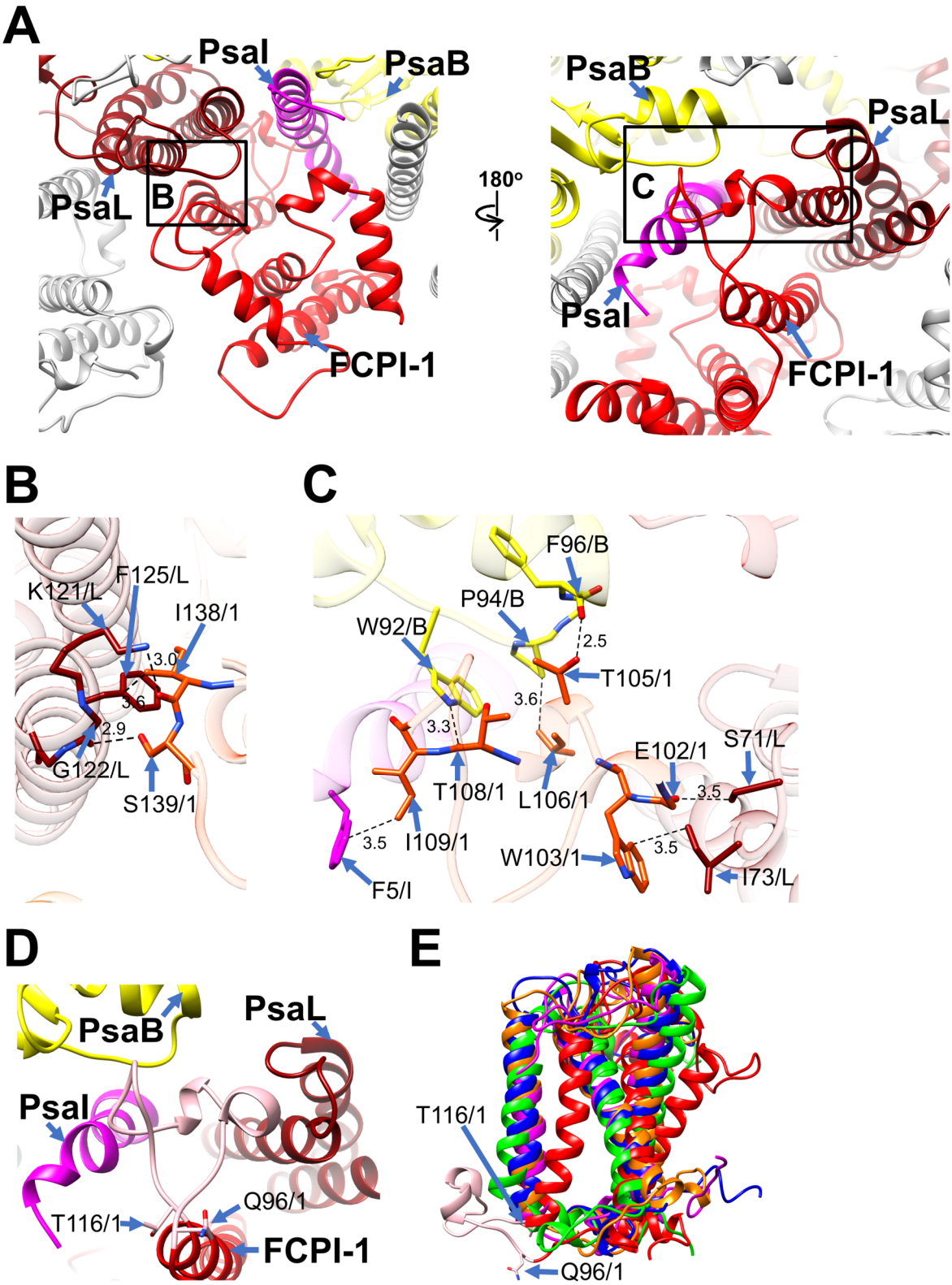


Fig. 3 | Structural characteristics of FCPI-1 (RedCAP).

(**A**) Interactions of FCPI-1 with PsaB, PsaI, and PsaL viewed from the stromal (left) and lumenal (right) sides. The areas encircled by black squares are enlarged in panels **B**, **C**. Yellow, PsaB; magenta, PsaI; dark red, PsaL; red, FCPI-1. (**B**, **C**) Protein-protein interactions of FCPI-1 with PsaL (**B**) and PsaB/PsaI/PsaL (**C**), respectively. Interactions are indicated by dashed lines, and the numbers are distances in Å. Amino acid residues participating in the interactions are labeled; for example, S139/1 means Ser139 of FCPI-1. B, PsaB; I, PsaI; L, PsaL; 1, FCPI-1. (**D**) Characteristic loop structure from

Q96 to T116 in FCPI-1 viewed from the lumenal side. Q96 and T116 are labeled with sticks and the Q96–T116 loop is colored pink. (**E**) Superpositions of FCPI-1 (red) with FCPI-2 (green), FCPI-3 (blue), FCPI-4 (magenta), and FCPI-5 (orange). Only proteins are depicted. The Q96–T116 loop of FCPI-1 is colored pink and Q96 and T116 are labeled with sticks.

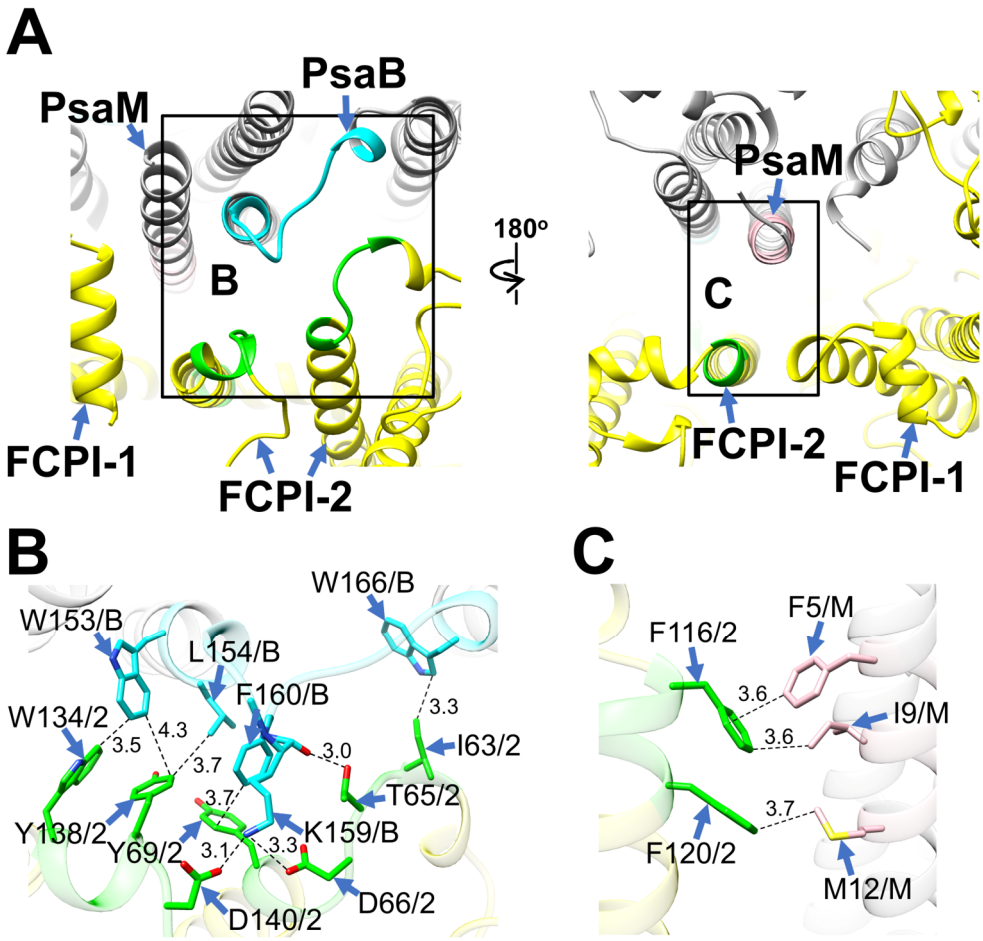


Fig. 4 | Structural characteristics of FCPI-2.

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

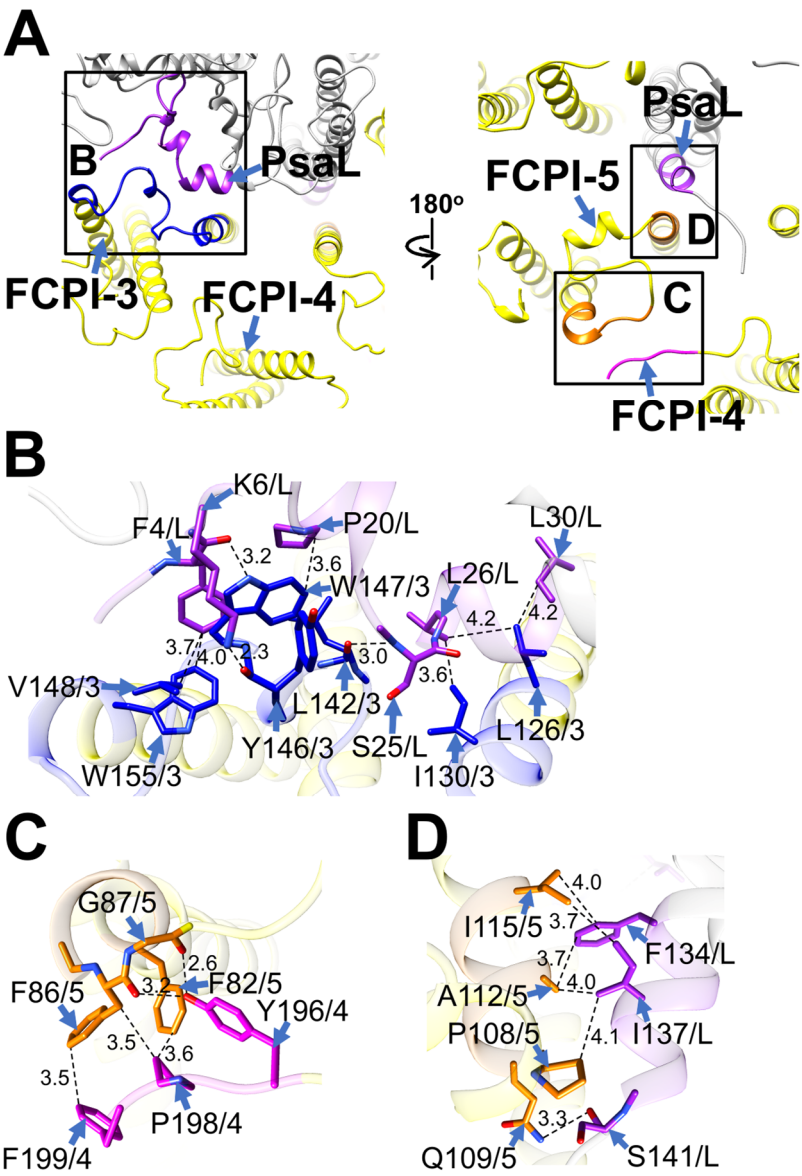
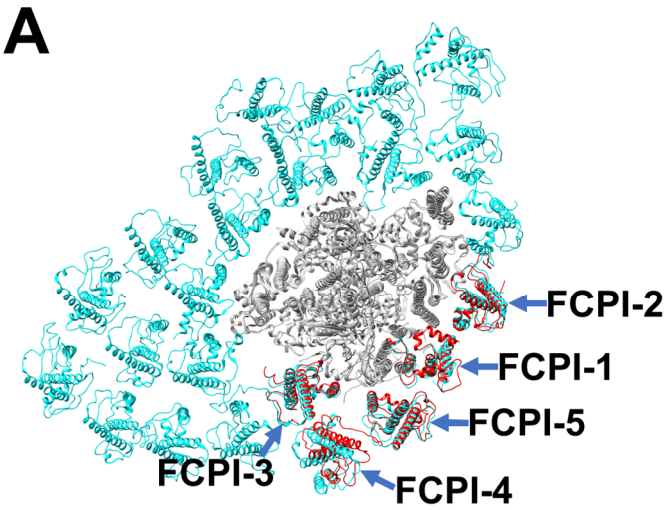


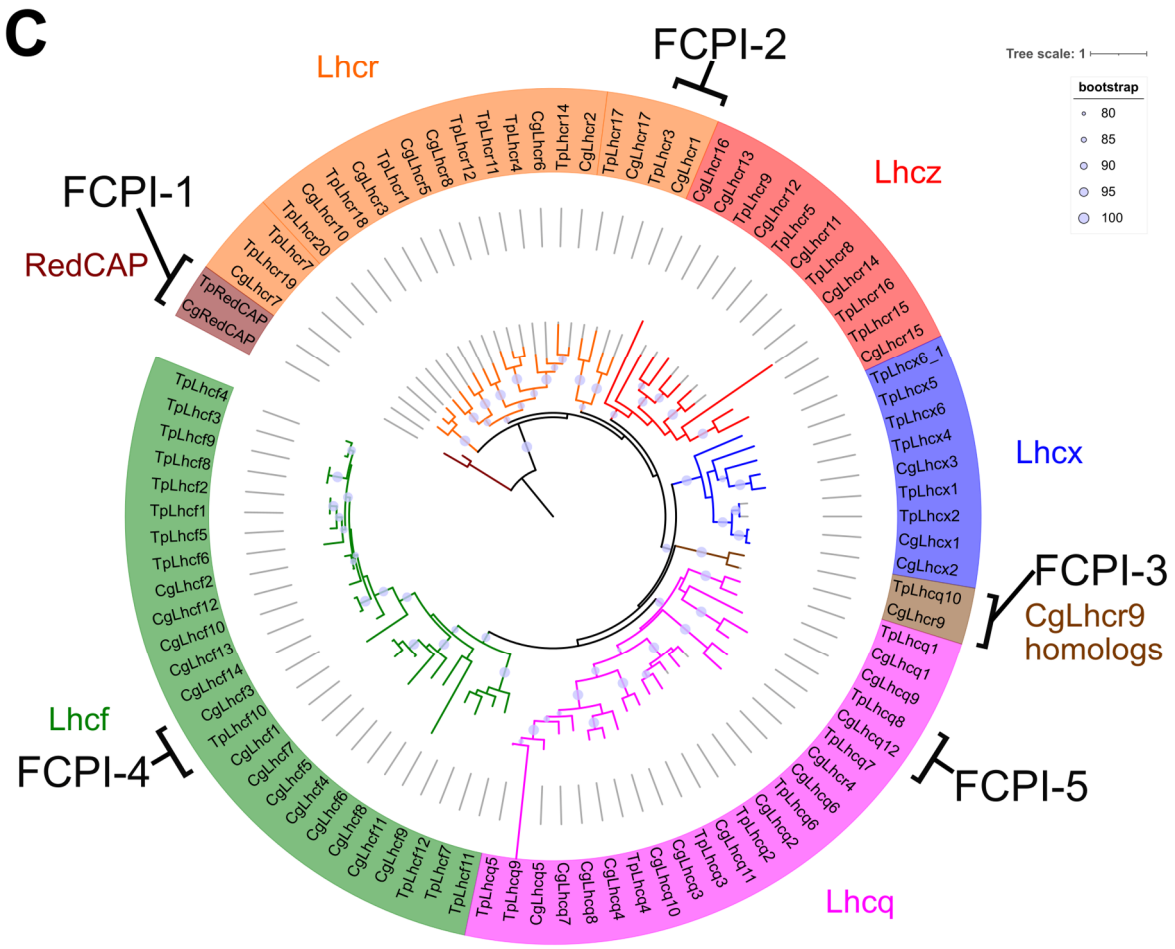
Fig. 5 | Structural characteristics of FCPI-3, 4, and 5.

(**A**) Interactions among FCPIs and between FCPIs and PsaL viewed from the stromal (left) and lumenal (right) sides. The areas encircled by black squares are enlarged in panels **B**–**D**. The PSI and FCPI subunits are colored grey and yellow, respectively. Protein-protein interactions are shown as 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**), respectively. Interactions are indicated by dashed lines, and the numbers are distances in Å. Amino acid residues participating in the interactions are labeled; for example, L126/3 means Leu126 of FCPI-3. 4, FCPI-4; 5, FCPI-5; L, PsaL.



B

T. pseudonana		*C. gracilis*
Position	Gene	Gene
FCPI-1	*RedCAP*	*RedCAP*
FCPI-2	*Lhcr3*	*Lhcr1*
FCPI-3	*Lhcq10*	*Lhcr9*
FCPI-4	*Lhcf10*	*Lhcf3*
FCPI-5	*Lhcq8*	*Lhcq12*



584 **Fig. 6. Comparisons of structures and sequences of FCPIs in the PSI-FCPI structures between**
585 ***T. pseudonana* and *C. gracilis*.**
586 (**A**) Superposition of the PSI-FCPI structures between *T. pseudonana* and *C. gracilis* (PDB: 6LY5).
587 The *T. pseudonana* and *C. gracilis* FCPI subunits are colored red and cyan, respectively. The
588 structures are viewed from the stromal side. The FCPI-1 to 5 sites are labeled. (**B**) Correlation of the
589 names of FCPIs in the structures with their genes between *T. pseudonana* and *C. gracilis*. The genes
590 of FCPIs are derived from Kumazawa et al. (*30*) and Kato et al. (*14*) for *C. gracilis*. (**C**) Phylogenetic
591 analysis of FCPs and RedCAPs from *T. pseudonana* (Tp) and *C. gracilis* (Cg). In addition to the
592 RedCAP subfamily, 44 TpFCPs and 46 CgFCPs are grouped into five Lhc subfamilies and CgLhcr9

homologs. Maroon, RedCAP subfamily; 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 by IQ-TREE 2 (*58*) using the Q.pfam + R4 model selected with ModelFinder (*59*). The light purple circular symbols on the tree represent bootstrap support (%).