

# In situ structure of the red algal phycobilisome–PSII–PSI–LHC megacomplex

<https://doi.org/10.1038/s41586-023-05831-0>

Received: 26 June 2022

Accepted: 13 February 2023

Published online: 15 March 2023



Xin You<sup>1,6</sup>, Xing Zhang<sup>2,6</sup>, Jing Cheng<sup>3,6</sup>, Yanan Xiao<sup>4,6</sup>, Jianfei Ma<sup>1</sup>, Shan Sun<sup>1</sup>, Xinzhen Zhang<sup>3,5</sup>✉, Hong-Wei Wang<sup>2</sup>✉ & Sen-Fang Sui<sup>1,4</sup>✉

In oxygenic photosynthetic organisms, light energy is captured by antenna systems and transferred to photosystem II (PSII) and photosystem I (PSI) to drive photosynthesis<sup>1,2</sup>. The antenna systems of red algae consist of soluble phycobilisomes (PBSs) and transmembrane light-harvesting complexes (LHCs)<sup>3</sup>. Excitation energy transfer pathways from PBS to photosystems remain unclear owing to the lack of structural information. Here we present in situ structures of PBS–PSII–PSI–LHC megacomplexes from the red alga *Porphyridium purpureum* at near-atomic resolution using cryogenic electron tomography and in situ single-particle analysis<sup>4</sup>, providing interaction details between PBS, PSII and PSI. The structures reveal several unidentified and incomplete proteins and their roles in the assembly of the megacomplex, as well as a huge and sophisticated pigment network. This work provides a solid structural basis for unravelling the mechanisms of PBS–PSII–PSI–LHC megacomplex assembly, efficient energy transfer from PBS to the two photosystems, and regulation of energy distribution between PSII and PSI.

Oxygenic photosynthesis is the principal way to convert sunlight into chemical energy on earth<sup>5</sup>. Light is captured by antenna systems, and transferred to PSII and PSI to induce the subsequent charge separation<sup>1,6</sup>. The light-harvesting systems of red algae contain both soluble phycobilisomes (PBSs) and membrane intrinsic light-harvesting complexes (LHCs)<sup>3,7</sup>. The cryogenic electron microscopy (cryo-EM) structures of PBS from red algae and cyanobacteria have been resolved<sup>8–11</sup>. The PBS from the red alga *Porphyridium purpureum* is hemi-ellipsoidal with a two-fold symmetry. A pyramidal allophycocyanin (APC) core is surrounded by 14 peripheral rods that are categorized into two types according to their phycobiliprotein composition: type I rods consist of two phycoerythrin hexamers distal to the core and one phycocyanin (PC) hexamer proximal to the core, whereas type II rods are composed entirely of phycoerythrin<sup>9</sup>. The core-membrane linker protein (L<sub>CM</sub>) and phycobiliprotein allophycocyanin D (ApcD) are considered as two terminal emitters to transfer the excited energy to PSII and PSI<sup>12,13</sup>. PSII catalyses the light-driven electron transfer from water to plastoquinone, and PSI drives electron transfer from plastocyanin (Pc) or c-type cytochromes to ferredoxin<sup>14</sup>. Structures of PSII–LHCII and PSI–LHCI from various organisms<sup>15–18</sup> show that the PSII and PSI cores are highly conserved, whereas the light-harvesting antenna systems show large diversity.

However, no high-resolution structures of PBS–PSII or/and PBS–PSI have been reported so far, although the interactions among them have been investigated by chemical cross-linking coupled with mass spectrometry (CXMS) analysis<sup>19</sup> and some low-resolution structures<sup>20,21</sup>. The CXMS study of cyanobacteria indicated the formation of a megacomplex composed of PBS, PSI and PSII<sup>19</sup>. The negative-staining EM structures of the

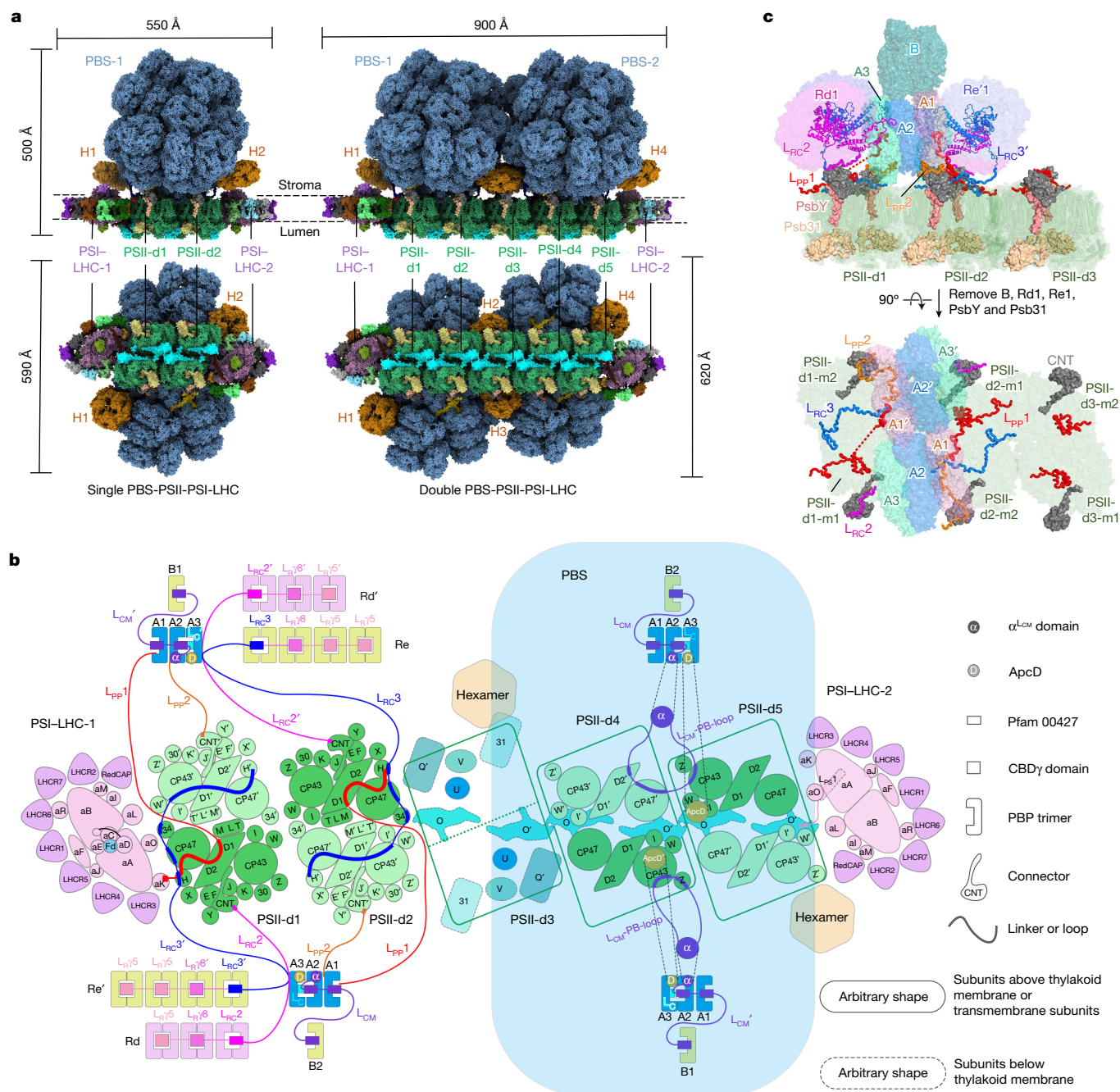
PBS–PSII supercomplex and PBS–CpL–PSI supercomplex purified from *Anabaena* sp. PCC 7120 showed the association of PBS with PSII<sup>21</sup> and PSI tetramer<sup>20</sup>, respectively. The cryo-electron tomography (cryo-ET) structure of the PBS–PSII supercomplex from *P. purpureum* revealed that α<sup>L</sup><sub>CM</sub> and ApcD contact PSII directly<sup>22</sup>. However, none of these studies provided precise subunit interactions and pigment arrangement at the molecular level. Spectral studies have demonstrated that red algae use a spillover mechanism, in which PSII transfers its excess excitation energy to PSI, to balance the light energy distribution between PSI and PSII<sup>23</sup>. However, the pigment network mediating the spillover remains unclear.

Here, we resolved the in situ structures of the PBS–PSII–PSI–LHC megacomplexes at overall resolutions of 3.3 Å for single PBS–PSII–PSI–LHC and 4.3 Å for double PBS–PSII–PSI–LHC. Our results provide a solid structural basis for unravelling the mechanism of PBS–PSII–PSI–LHC megacomplex assembly, the energy transfer routes within it and the spillover mechanism.

## Overall structures of PBS–PSII–PSI–LHC

Cryo-ET and an in situ single-particle analysis (isSPA) method<sup>4</sup> were combined to reconstruct the PBS–PSII–PSI–LHC megacomplex at near-atomic resolution in the native context (Extended Data Figs. 1 and 2). The structure of the in situ PBS is almost identical to the structure of the isolated PBS<sup>9</sup>. The overall structure of the PBS–PSII–PSI–LHC megacomplex shows two types of assembly (named single PBS–PSII–PSI–LHC and double PBS–PSII–PSI–LHC, respectively), showing completely different sizes, masses and PBS:PSII:PSI stoichiometric ratios of

<sup>1</sup>State Key Laboratory of Membrane Biology, Beijing Frontier Research Center for Biological Structures, Beijing Advanced Innovation Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing, China. <sup>2</sup>Ministry of Education Key Laboratory of Protein Sciences, Tsinghua-Peking Joint Center for Life Sciences, Beijing Advanced Innovation Center for Structural Biology, Beijing Frontier Research Center for Biological Structures, School of Life Sciences, Tsinghua University, Beijing, China. <sup>3</sup>National Laboratory of Biomacromolecules, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China. <sup>4</sup>School of Life Sciences, Cryo-EM Center, Southern University of Science and Technology, Shenzhen, China. <sup>5</sup>University of Chinese Academy of Sciences, Beijing, China. <sup>6</sup>These authors contributed equally: Xin You, Xing Zhang, Jing Cheng, Yanan Xiao. ✉e-mail: xzzhang@ibp.ac.cn; hongweiwang@tsinghua.edu.cn; suisf@mail.tsinghua.edu.cn

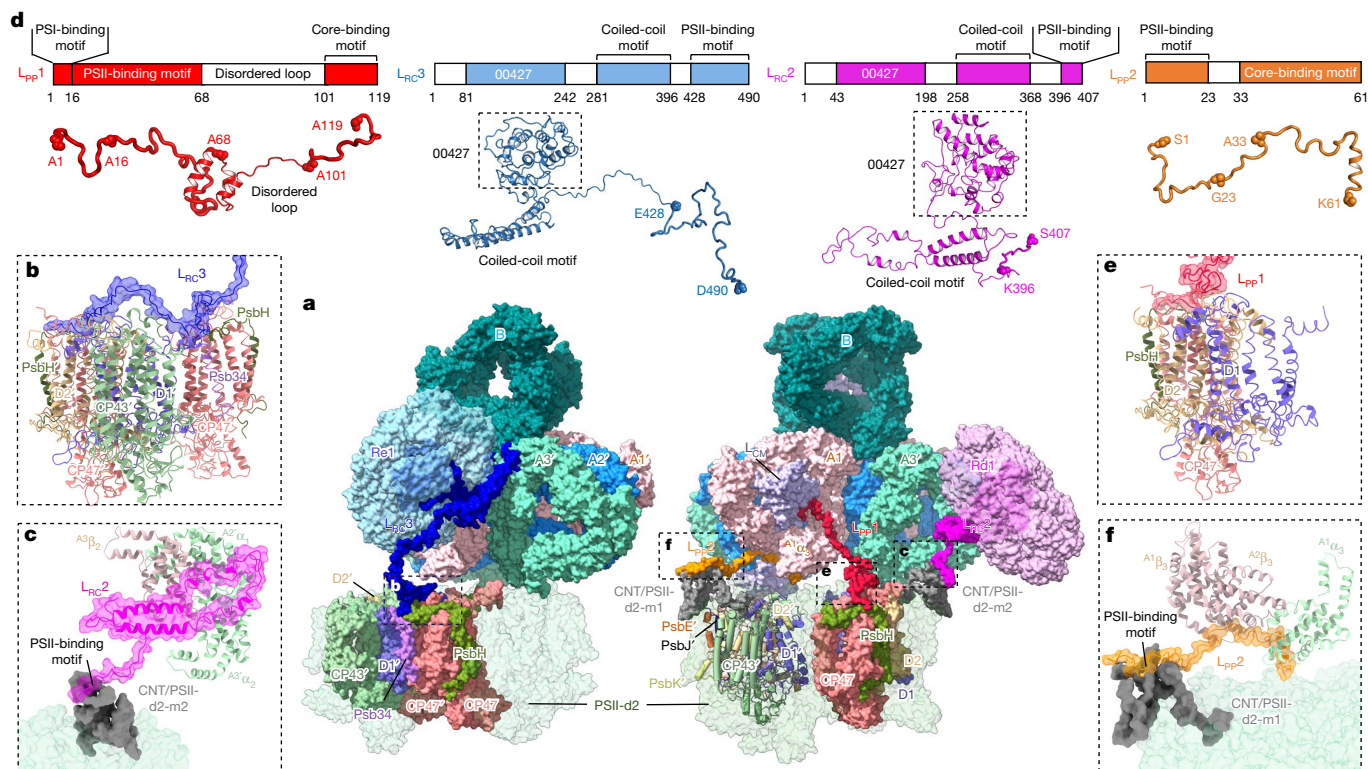


**Fig. 1 | Overall structures of the PBS-PSII-PSI-LHC megacomplex from *P. purpureum*.** **a**, The overall structures of two types of assembly of PBS-PSII-PSI-LHC are shown as a surface representation. PBSs, lateral hexamers, PSII dimers and PSI-LHCs are coloured cornflower blue, dark goldenrod, medium sea green and pink, respectively. The key components of PSII dimers and PSI-LHCs are coded in different colours. The refinements were repeated at least two times with similar results. **b**, Schematic model of the PBS-PSII-PSI-LHC architecture, showing the components and connections of PBS, PSII and PSI-LHC. Transmembrane subunits are shown in PSI-LHC-1 and -2, whereas

subunits on the stromal side and the luminal side are shown in PSI-LHC-1 and PSI-LHC-2, respectively. Five extrinsic PSII subunits (PsbQ, PsbO, PsbV, PsbU and Psb31) are shown in PSII-d3. PSII-d3 and -d4 show the interaction between PSII dimers through PsbO. PSII-d1 and -d2 and PSII-d3 and -d4 show the interactions between PBS and PSII and between the PBS core ( $L_{CM}$ -PB-loop and ApcD) and PSII, respectively. The interacting regions of proteins are shown in bold. PSII-d5 and PSI-LHC-2 show the connection between  $L_{PS1}$  and PsbO. **c**, Organization of PBS core, PSII and linker proteins  $L_{RC3}$ ,  $L_{RC2}$ ,  $L_{PP1}$  and  $L_{PP2}$ .

roughly  $550 \times 590 \times 500 \text{ \AA}^3$ , about 16.7 MDa, 1:4:2 for single PBS-PSII-PSI-LHC and roughly  $900 \times 620 \times 500 \text{ \AA}^3$ , about 33.0 MDa, 2:10:2 for double PBS-PSII-PSI-LHC (Fig. 1a). For the single PBS-PSII-PSI-LHC, a PBS directly anchors to and fully covers two PSII dimers (PSII-d1 and PSII-d2). Two PSI-LHCs are associated with the two PSII dimers in a side-on orientation. Two symmetrical phycoerythrin hexamers (H1 and H2) are located on both sides of PBS. In the double PBS-PSII-PSI-LHC,

two PBSs fully cover five PSII dimers (PSII-d1 to PSII-d5) in which the middle PSII-d3 has no direct interaction with both PBSS, and two PSI-LHCs are connected to two ends of the PSII linear array (Fig. 1a). Hexamers H2 and H3 located between the two PBSs show apparently flatter conformations in comparison to hexamers H1 and H4 located at the ends of PSII (Fig. 1a). By placing the subtomogram maps of single and double PBS-PSII-PSI-LHC back into the corresponding tomograms,



**Fig. 2 | PBS and PSII are linked through four linker proteins. a**, Overview of four linker proteins interacting with both PBS and PSII from two different views. **b, c**, Interactions between  $L_{RC3}$  and PSII (**b**), and  $L_{RC2}$  with the PBS core

and PSII (**c**). **d**, Representative diagrams of structural elements of  $L_{PP1}$ ,  $L_{RC3}$ ,  $L_{RC2}$  and  $L_{PP2}$  are shown above the structures. **e, f**, Interactions between  $L_{PP1}$  and PSII (**e**) and  $L_{PP2}$  with PBS core and PSII (**f**).

we found that the single and double assemblies were packed randomly yet tightly on the thylakoid membrane at a ratio of around 3:1 (Extended Data Fig. 3). The overall resolution for single PBS–PSII–PSI–LHC is 3.3 Å (C1 symmetry) and for double PBS–PSII–PSI–LHC is 4.3 Å (C2 symmetry). After applying classifications and refinements with individual local masks, the resolutions of the subcomplexes were improved to 3.2 Å for PSII-d1 and -d2, 3.4 Å for PSII-d3, 3.6 Å for PSI–LHC and 6.3 Å for the lateral hexamer (Extended Data Fig. 2 and Extended Data Table 1). We also identified 1,434 phycoerythrobilin, 48 phycourobilin and 120 phycocyanobilin for each PBS, 70 chlorophylls (Chls), four haem molecules, 20 carotenoids and 54 lipids for each PSII dimer and 187 Chls, 21 carotenoids, two zeaxanthins and four lipids for each PSI–LHC. All pigments are shown in Supplementary Table 1.

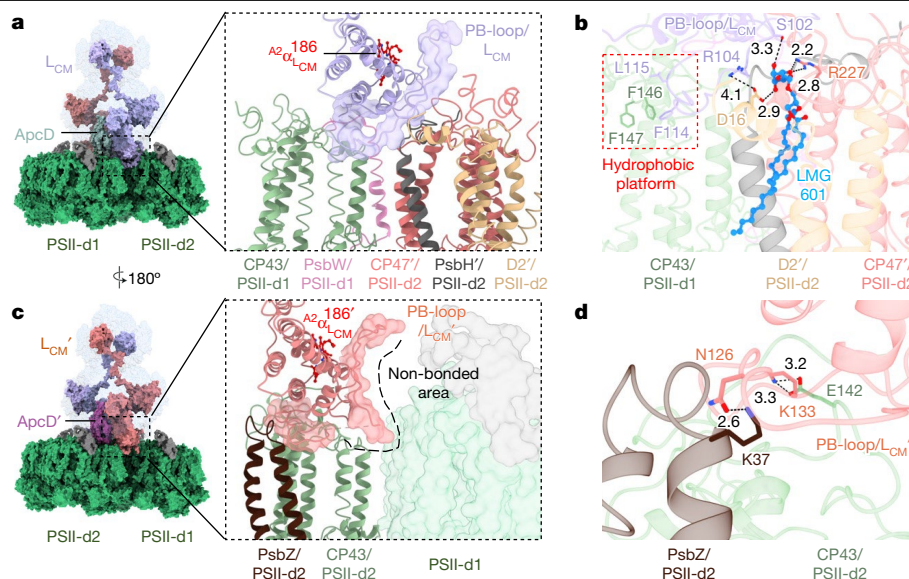
### New components in in situ PSII structure

The structures of PSII from different species have been solved through X-ray and cryo-EM<sup>15,17,24</sup>. Nevertheless, some subunits, especially those in the peripheral regions, could be lost during purification. The in situ structure of the PSII dimer presented here contains more complete PSII core subunits than before and reveals several unprecedented details. Each PSII monomer consists of 19 core subunits (D1, CP43, CP47, D2, PsbE, PsbF, PsbH, PsbI, PsbJ, PsbK, PsbL, PsbM, PsbN, PsbO, PsbP, PsbQ, PsbR, PsbS, PsbT, PsbU, PsbV, PsbW, PsbX, PsbY, PsbZ and Psb31) and five extrinsic subunits (PsbQ', PsbO, PsbU, PsbV and Psb31) (Fig. 1b and Supplementary Table 1). Among them, PsbY and Psb31 are found in the red algal PSII structure (Fig. 1c). Three extra densities defined as 'connectors' in the published cryo-ET PBS–PSII structure<sup>22</sup> are identified as the same molecule (still named connectors and abbreviated to CNT) according to the high-resolution map in this work (Fig. 1c). Unfortunately, the sequence of CNT could not be identified, so they are modelled as poly-Ala residues. Overall, CNT resembles the shape of a question mark roughly 50 Å in height

and 55 Å in width and consists of two distinct domains. The N-terminal domain, the tail of the question mark, contains a long loop followed by a short  $\alpha$ -helix. The C-terminal domain is composed of eight  $\beta$ -strands that form the head of the question mark (Extended Data Fig. 4a,b). The CNT sits stably on the stromal side of each PSII monomer, and the N-terminal domain partly inserts into a large hydrophobic cavity formed by the transmembrane helices (TMHs) of Cyt b559 (PsbE and PsbF), PsbJ, PsbK, D1, D2 and CP43, suggesting that CNT may be an inherent component of PSII (Fig. 2a and Extended Data Fig. 4c). This cavity has been considered to be an exchange channel between plastoquinone  $Q_B$  and the PQ pool<sup>25</sup>, suggesting a possible role of CNT in regulating the exchange of  $Q_B$ .

PsbY is a peripheral TMH of PSII and has been observed only in cyanobacteria previously<sup>26</sup>. The in situ PSII structure shows that PsbY is located in close proximity to Cyt b559 and CNT (Fig. 1b,c and Extended Data Fig. 4d). Similar to the cyanobacterial PsbY<sup>26</sup>, the red algal PsbY are vital to stabilize the conformation of PsbE and PsbF through polar or hydrophobic interactions (Extended Data Fig. 4d,e). Sequence alignment also shows that the interacting residues are highly conserved as charged or hydrophobic amino acids among different species, indicating their important roles in the localization and function of PsbY (Extended Data Fig. 4f). Furthermore, the C terminus of PsbY stabilizes the N terminus of PsbF that extends into the gap between the N- and C-terminal domains of CNT to consolidate the binding of CNT with PSII (Extended Data Fig. 4g).

Psb31, the fifth extrinsic protein, has been previously found only in the diatom PSII<sup>17</sup>. Evolutionarily, diatom Psb31 may have arisen by gene divergence of red algal PsbQ' after a secondary endosymbiosis event<sup>27,28</sup>. However, the red algal Psb31 (*Pp* Psb31) shares high structural conservation with the diatom Psb31, although the sequence identity between them is low (33.3%), suggesting that Psb31 has diversified in red algae and diatom Psb31 has inherited the gene encoding red algal



**Fig. 3 | Interaction between  $L_{CM}$  and PSII tetramers.** **a**, Interaction between  $L_{CM}$  and PSII. Right, an enlarged view showing that the PB-loop/ $L_{CM}$  binds to both PSII-d1 and PSII-d2. **b**, Details of the interaction between  $L_{CM}$  and PSII subunits. The residues are shown as a stick representation, and the lipid LMG is shown as

a ball and stick representation. **c**, Interaction between  $L'_{CM}$  and PSII. Right, an enlarged view showing that the PB-loop/ $L'_{CM}$  binds only to PSII-d2. **d**, Details of the interaction between  $L'_{CM}$  and PSII subunits. The residues are shown as stick representations.

Psb31 (Extended Data Fig. 4h–j). Compared to diatom Psb31, *Pp*-Psb31 has two extra loop regions, extending the four conserved helices (Extended Data Fig. 4i, j). Several cationic residues of *Pp*-Psb31 (K69, R71 and R73) form a positively charged area binding to a negatively charged area formed by the CP47 and D2 subunits (Extended Data Fig. 5k), as described in diatoms previously<sup>17</sup>.

#### Four PBS–PSII linker proteins

The structures of the in situ PBS–PSII–PSI–LHC megacomplexes showed the coupling mode between PBS and the two photosystems, showing that PBS has direct interactions with PSII but not with PSI (Fig. 1). Four linker proteins,  $L_{RC3}$ ,  $L_{RC2}$ ,  $L_{pp1}$  (linker of PBS–PSII 1) and  $L_{pp2}$  (linker of PBS–PSII 2), are essential to stabilize the coupling of PBS and PSII (Figs. 1c and 2). Furthermore,  $L_{pp1}$  is also involved in the interaction between PSII and PSI (described later).

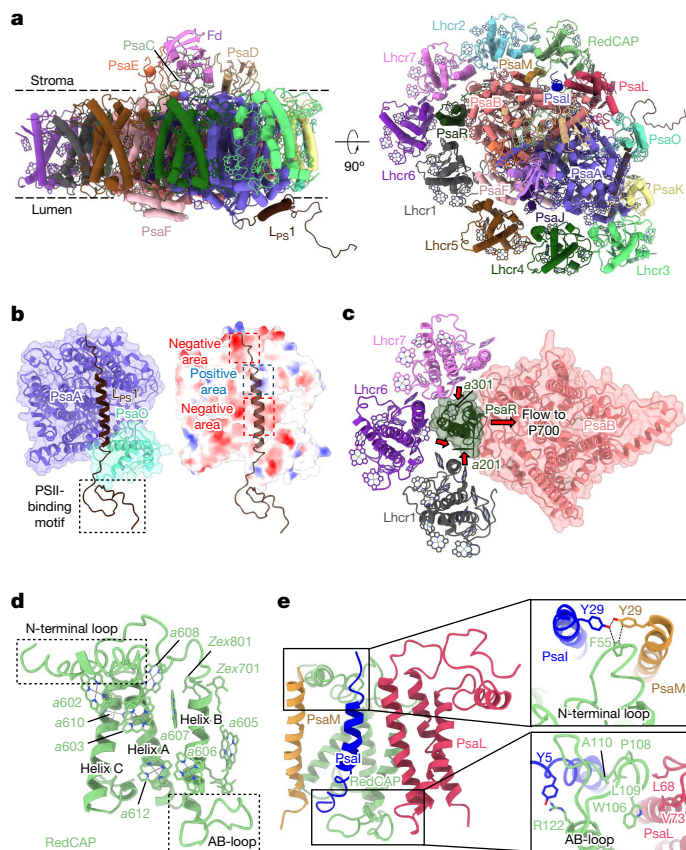
In the PBS structures published previously,  $L_{RC3}$  contains an N-terminal Pfam00427 domain inserted into the central cavity of the phycoerythrin hexamers of Rod e and an  $\alpha$ -helix motif contacting with the  $\alpha$ -subunit of the PBS core, and  $L_{RC2}$  uses a similar method to attach Rod d to the PBS core<sup>8,9</sup>. However, the residues at their very C termini are absent. These missing regions are clearly resolved in the in situ structures. The C-terminal region of  $L_{RC3}$  (E428 to D490) spans the PSII surface and binds to PsbH, CP47, Psb34, D1', CP43', CP47', D2' and PsbH' through polar and electrostatic interactions (Fig. 2a, b and Extended Data Fig. 5a). In particular, successive charged residues (R460/ $L_{RC3}$  to E466/ $L_{RC3}$ ) interact with R461/CP43' and E462/CP43', and E434/ $L_{RC3}$  forms a salt bridge with R4/PsbH (Extended Data Fig. 5b). In addition, a short C-terminal loop of  $L_{RC3}$  (K443 to N451) interacts with the C terminus of Psb34, resulting in an obvious structural shift of the latter compared to its symmetric counterpart Psb34' that does not contact  $L_{RC3}$  (Extended Data Fig. 5c). In the case of  $L_{RC2}$ , a short C-terminal loop extends into the groove of a CNT of PSII dimer (Fig. 2c). Thus, in addition to connecting the rods to the core, both  $L_{RC3}$  and  $L_{RC2}$  use their C-terminal regions to bind to the stromal surface of PSII.

$L_{pp1}$  and  $L_{pp2}$  are two new subunits that are directly involved in linking PBS to PSII (Fig. 2).  $L_{pp1}$ , which is modelled with poly-Ala residues, contains N-terminal PSI- and PSII-binding motifs and a C-terminal

PBS-core-binding motif (Fig. 2d). Through electrostatic interactions, the PSII-binding motif (A16 to A68) attaches to the PSII surface by binding with D1, D2, CP47 and PsbH, and the PBS-core-binding motif (A101 to A119) is nestled into the groove formed by  $L_{CM}$  and  $A^1\alpha_3$  of the core layer A1 (Fig. 2e and Extended Data Fig. 5d). Furthermore,  $L_{pp1}$  has different conformations in PSII-d1, PSII-d2 and PSII-d3 (Fig. 1c). Compared to  $L_{pp1}$ /PSII-d2, the disordered loop at the middle of  $L_{pp1}$ /PSII-d1 is not resolved. In fact, owing to a roughly 14° rotation and 30 Å shift of the PBS core relative to the PSII dimer pair<sup>22</sup>, the distance between the binding sites of  $L_{pp1}$ /PSII-d1 with PSII-d1 and  $L_{pp1}$ /PSII-d1 with the PBS core is greater than that between the binding sites of  $L_{pp1}$ /PSII-d2 with PSII-d2 and  $L_{pp1}$ /PSII-d2 with the PBS core, which may lead to the longer middle loop in a random conformation. Different from the above two  $L_{pp1}$  molecules, only the PSII-binding motif is resolved in  $L_{pp1}$ /PSII-d3 because its C terminus lacks interactions with the PBS core and therefore was instable (Fig. 1c). Notably, in PSII-d3, two  $L_{pp1}$ s bind to two PSII monomers by their N-terminal domains, whereas in PSII-d1 and PSII-d2 only one PSII monomer is associated with  $L_{pp1}$  and the other one is associated with  $L_{RC3}$  (Fig. 1c). Indeed, the PSII-binding motifs of  $L_{pp1}$  and  $L_{RC3}$  share a highly similar structural feature in their interactions with PSII and bind to almost the same sites of PSII (Fig. 2b, e).  $L_{pp2}$ , a homologue of ApcG recently identified in the cyanobacterial PBS<sup>11</sup>, is another new PBS–PSII linker protein (Fig. 2d). Different from ApcG, which was reported to bind only to the membrane-facing side of PBS<sup>11</sup> probably owing to the lack of PSII in the isolated PBS,  $L_{pp2}$  binds to both PBS and PSII in the in situ structure. The N terminus of  $L_{pp2}$  (S1 to G23) is the PSII-binding motif and interacts with a groove of CNT (Fig. 2f), and its C terminus (A33 to K61) is the PBS core-binding motif that extends into the gap between core A1 and A2 layer by hydrophobic interactions. Therefore, one PSII monomer in the PSII dimer associates with the PBS core through  $L_{RC2}$  and  $L_{pp1}$  and the other one through  $L_{RC3}$  and  $L_{pp2}$  (Fig. 1c).

#### Interactions of $L_{CM}$ and ApcD with PSII

$L_{CM}$  and ApcD are essential for energy migration from PBS to photosystems<sup>19</sup>. In our structures, the PB domain of  $L_{CM}$  (PB-loop), which was absent in the published maps, is clearly resolved (Fig. 3). The



**Fig. 4 | Structural features of the PSI-LHC supercomplex.** **a**, Overall structure of the PSI-LHC supercomplex viewed parallel (left) and normal (right) from the stromal side to the membrane plane. **b**, Interaction of  $L_{ps1}$  with PsaA and PsaO. Electrostatic surface representation of PsaA and PsaO are shown in the right panel. Positively and negatively charged areas are boxed in blue and red rectangles, respectively. **c**, PsaR is located at the periphery of the PSI core and surrounded by three antenna subunits. The red arrows indicate the direction of the potential energy transfer. **d**, Structure and arrangement of pigments in RedCAP. **e**, Interactions between RedCAP and PsaM, PsaL and PsaL.

interaction interfaces of  $L_{CM}$  and  $L_{CM}'$  with PSII dimers are different owing to the rotation between PBS and PSII (as mentioned above). For  $L_{CM}$ , the PB-loop/ $L_{CM}$  interacts with both PSII-d1 and PSII-d2. F114/ $L_{CM}$  and L115/ $L_{CM}$  are stacked against F146/CP43 and F147/CP43 of PSII-d1, forming a hydrophobic binding platform to stabilize PB-loop/ $L_{CM}$  and PSII-d1. Furthermore, one lipid (LMG601) between PB-loop/ $L_{CM}$  and D2' of PSII-d2 forms a hydrogen bond (H-bond) network by binding to S102/ $L_{CM}$ , D16/PSII-d2<sub>D2'</sub> and R227/PSII-d2<sub>CP47'</sub> (Fig. 3a,b). Thus, LMG601 could be involved in the stabilization of PB-loop/ $L_{CM}$ , D2' and CP47' of PSII-d2. R104/ $L_{CM}$  forms a salt bridge with D16/PSII-d2<sub>D2'</sub>, stabilizing the PB-loop conformation (Fig. 3b). However, for  $L_{CM}'$ , its PB-loop interacts with PSII-d2 only. K133/ $L_{CM}'$  and N126/ $L_{CM}'$  form a salt bridge and a H-bond with E142/PSII-d2<sub>CP43</sub> and K37/PSII-d2<sub>PsbZ</sub>, respectively (Fig. 3c,d). Thus, compared to PB-loop/ $L_{CM}$ , PB-loop/ $L_{CM}'$  is more flexible and could swing randomly (Extended Data Fig. 5e). Sequence alignment of the PB-loop shows that although the sequence of the PB-loop is not conserved well, three key residues, F114, L115 and K133, involved in the PBS-PSII interaction maintain the corresponding properties in different species, suggesting that the PB-loop may mediate the interaction between PBS and PSII through some key residues in both red algae and cyanobacteria (Extended Data Fig. 5f). Similarly, ApcD and ApcD' contact PsbW and PsbI of PSII-d1 and CP43 of PSII-d2, respectively (Extended Data Fig. 5g). Compared to the symmetrical CP43/PSII-d1, the N terminus of CP43/PSII-d2 swings out owing to the steric hindrance from ApcD'

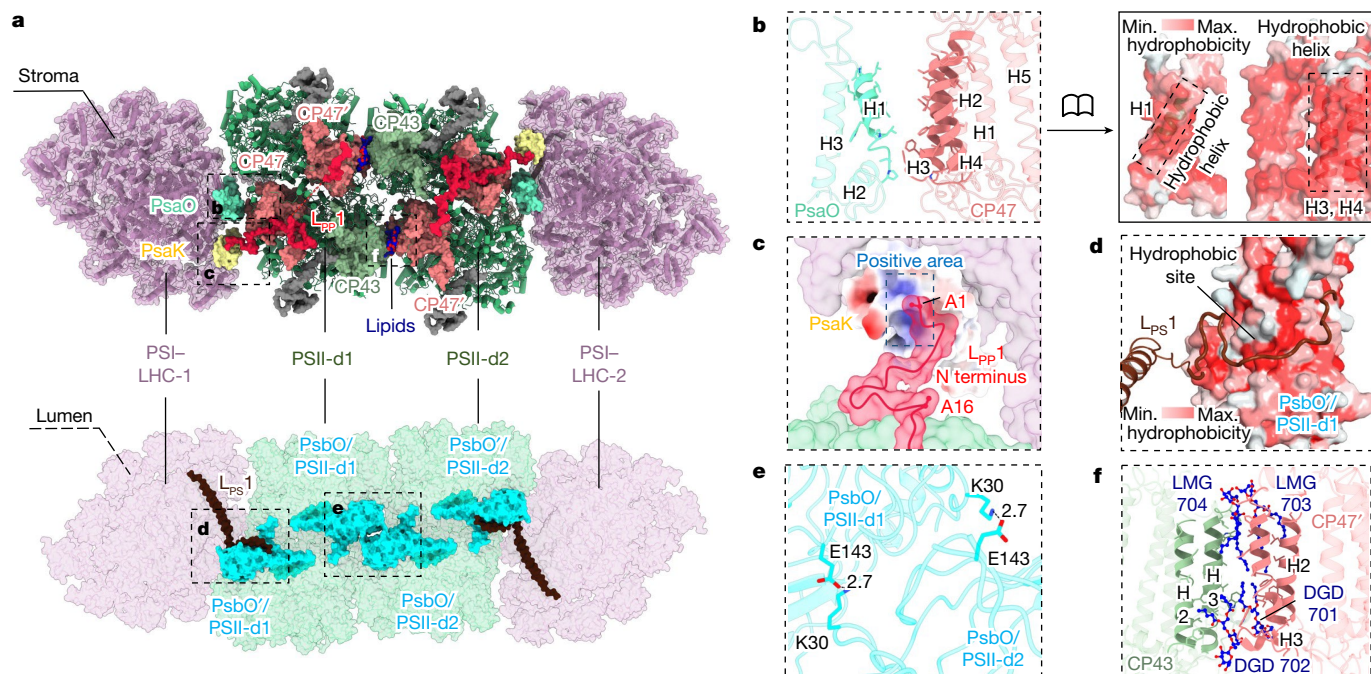
(Extended Data Fig. 5h,i). Previous CXMS analysis suggested several possible binding sites of  $L_{CM}$  with PSII and ApcD with PSI in cyanobacteria *in vivo*<sup>19</sup>, which are different from those in our structures, especially with respect to the ApcD. This may indicate a different coupling and energy transfer mechanism between PBS and photosystems in red algae and cyanobacteria.

## The in situ structure of PSI-LHC

PSI-LHC consist of 13 core subunits (PsaA, PsaB, PsaC, PsaD, PsaE, PsaF, PsaI, PsaJ, PsaK, PsaL, PsaM, PsaO and PsaR) with the electron carrier ferredoxin (Fd) and eight antenna subunits (seven Lhcrs and one RedCAP, Chl *a/b*-binding-like protein) (Fig. 4a and Supplementary Table 1). One subunit located on the luminal surface could not be assigned. It contains a C-terminal helix interacting with three charged regions formed by PsaA and PsaO of PSI through electrostatic contacts, and its N terminus extends into the adjacent PSII subunit PsbO. Thus, it may mediate the interaction between PSII and PSI (discussed later), and is denoted as a photosystem linker protein ( $L_{ps1}$ ) (Fig. 4b). The PSI core subunits are mostly consistent with those of the PSI-LHCR supercomplex from the red alga *Cyanidioschyzon merolae*<sup>29</sup>, except for PsaR, which was considered to exist only in diatoms<sup>30</sup>. Thus, diatom PsaR was probably inherited from red algae after a secondary endosymbiotic event. Compared to PsaR from a diatom (Cg\_PsaR, Protein Data Bank (PDB) entry 6L4U), the red algal PsaR (Pp\_PsaR) contains an extra Chl *a* molecule ( $\alpha 301$ ) in addition to  $\alpha 201$ . Structure and sequence alignments of Pp\_PsaR with Cg\_PsaR show that a loop of Pp\_PsaR has disappeared and the side chain of Q95/Pp\_PsaR is oriented towards the central atom Mg of  $\alpha 301$ , which seems to have a stronger attraction to  $\alpha 301$  than Y95/Cg\_PsaR (Extended Data Fig. 6a). Pp\_PsaR is located at the periphery of the PSI core and surrounded by Lhcr1, Lhcr6 and Lhcr7. This suggests that it may function as an 'intermediator' to facilitate the association of the PSI core with antennas, and mediate the energy transfer from Lhcrs to the PSI core (Fig. 4c).

Fd associates with PSI for intermolecular electron transfer<sup>31</sup>. In our structure, Fd contacts PsaA, PsaC, PsaD and PsaE through electronic interactions, as shown previously<sup>32</sup> (Extended Data Fig. 7a-c). There is a shorter distance between the [2Fe-2S] cluster of Fd and the [4Fe-4S] cluster of  $F_B$  in PsaC (7.9 Å) compared with the published PSI-Fd supercomplex, which may improve the efficiency of electron transfer<sup>32,33</sup> (Extended Data Fig. 7d).

Eight antennas are arranged in a notched oval-like shape around the PSI with the opening on the PsaL-PsaO-PsaK side (Fig. 4a). The phylogenetic tree shows that RedCAPs cluster together and form a monophyletic clade distinct from Lhcrs (Extended Data Fig. 6b), indicating that RedCAPs possibly fulfil functions other than light harvesting. Although the position of RedCAP is similar to that of Lhcr1\* from *C. merolae*, it is notably shifted to the PsaL-PsaI-PsaM side, forming tighter interactions with the PSI core than Lhcr1\* (Fig. 4d,e and Extended Data Fig. 7e-g). The N-terminal loop of RedCAP swings towards the PsaI-PsaM side, and F55/RedCAP forms two T-shaped  $\pi$ - $\pi$  interactions with Y29/PsaI and Y29/PsaM (Fig. 4e and Extended Data Fig. 7g). Moreover, the AB-loop of RedCAP also wriggles to the luminal side of PsaI-PsaL. The completely conserved R122/RedCAP forms one cation- $\pi$  interaction with Y5/PsaI (Fig. 4e and Extended Data Fig. 6c). A conserved hydrophobic region (W106-A110) of RedCAP provides a hydrophobic platform to bind with L68/PsaL and V73/PsaL (Fig. 4e and Extended Data Fig. 6c). Sequence alignments of PsaI, PsaM and PsaL from various species showed that the residues of PsaI, PsaM and PsaL in the binding sites are basically conserved, except for Y29/PsaI and Y29/PsaM, which are replaced by two alanine residues in *C. merolae*, resulting in the disappearance of the two T-shaped  $\pi$ - $\pi$  interactions (Extended Data Fig. 6d). This could be responsible for the inability of the PSI from *C. merolae* to bind RedCAP<sup>34</sup>. Furthermore, structural superposition of the *P. purpureum* PSI-LHC with the *C. merolae* PSI-LHCRs using the PSI core as a



**Fig. 5 | Interaction patterns between PSII and PSI.** **a**, The interaction patterns of PSII and PSI at the stromal side (upper) and the lumen side (lower). **b**, Hydrophobic interaction between CP47 and PsaO. On the left, proteins and interacting residues are shown as cartoon and stick representations, respectively. On the right, the proteins are shown as a surface representation and coloured on the basis of amino acid hydrophobicity. **c**, Electrostatic

interaction between PsaK and the N terminus of L<sub>pp1</sub>. **d**, Hydrophobic interaction between the N terminus of L<sub>ps1</sub> and the groove of PsbO'. PsbO' is coloured on the basis of amino acid hydrophobicity. **e**, Close-up view of the interaction between PsbO molecules from adjacent PSII dimers. **f**, Lipid-mediated hydrophobic interactions between CP43 and CP47' of adjacent PSII dimers. Min., minimum; max., maximum.

reference revealed some differences in the antennas. Three extra Lhcrs (Lhcr1, Lhcr6 and Lhcr7) are directly associated with PsaR (absent in *C. merolae*) through several polar interactions, which may mediate energy transfer towards the PSI core (Extended Data Fig. 7e,h). In the diatom PSI-FCPI supercomplex, similar antennas were also found at the same places, suggesting that PsaR may be involved in the assembly of the antennas<sup>16</sup>.

### Interaction patterns of two photosystems

The megacomplex structures demonstrate that PSII dimers and PSI-LHC are tightly packed on the thylakoid membrane. Detailed analysis revealed PSII-PSI interconnections in three aspects (Fig. 5). First, a plentiful hydrophobic residues cover  $\alpha$ -helices H3 and H4 of CP47/PSII-d1 and H1 of PsaO, forming a large hydrophobic interface within the thylakoid membrane (Fig. 5b). Second, the N terminus of L<sub>pp1</sub> (A1 to A16) of the outmost PSII dimers extends to the stromal side of PSI and binds to PsaK through electrostatic interactions (Fig. 5c). Thus, linker protein L<sub>pp1</sub> participates in the associations of PBS, PSII and PSI. Third, L<sub>ps1</sub> uses its N terminus to tightly bind to the PsbO subunit of PSII through hydrophobic interactions on the lumen side (Fig. 5d). In sum, the connections between two photosystems occur on both the stromal side and the luminal side, as well as within the membrane.

PsbO on the luminal side also mediates the connection between two PSII dimers. For each PSII dimer, one PSII monomer provides a PsbO connecting with PSI through L<sub>ps1</sub> and the other monomer contributes a PsbO to interact with an adjacent PSII dimer through two salt bridges between K30 and E143 (Fig. 5a,e). Moreover, two groups of lipids (DGD701, DGD702, LMG703 and LMG704) are inserted into the gap between CP43 and CP47' of PSII-d1 and -d2, forming a hydrophobic network with  $\alpha$ -helices H3 and H4 of CP43 and CP47' (Fig. 5f).

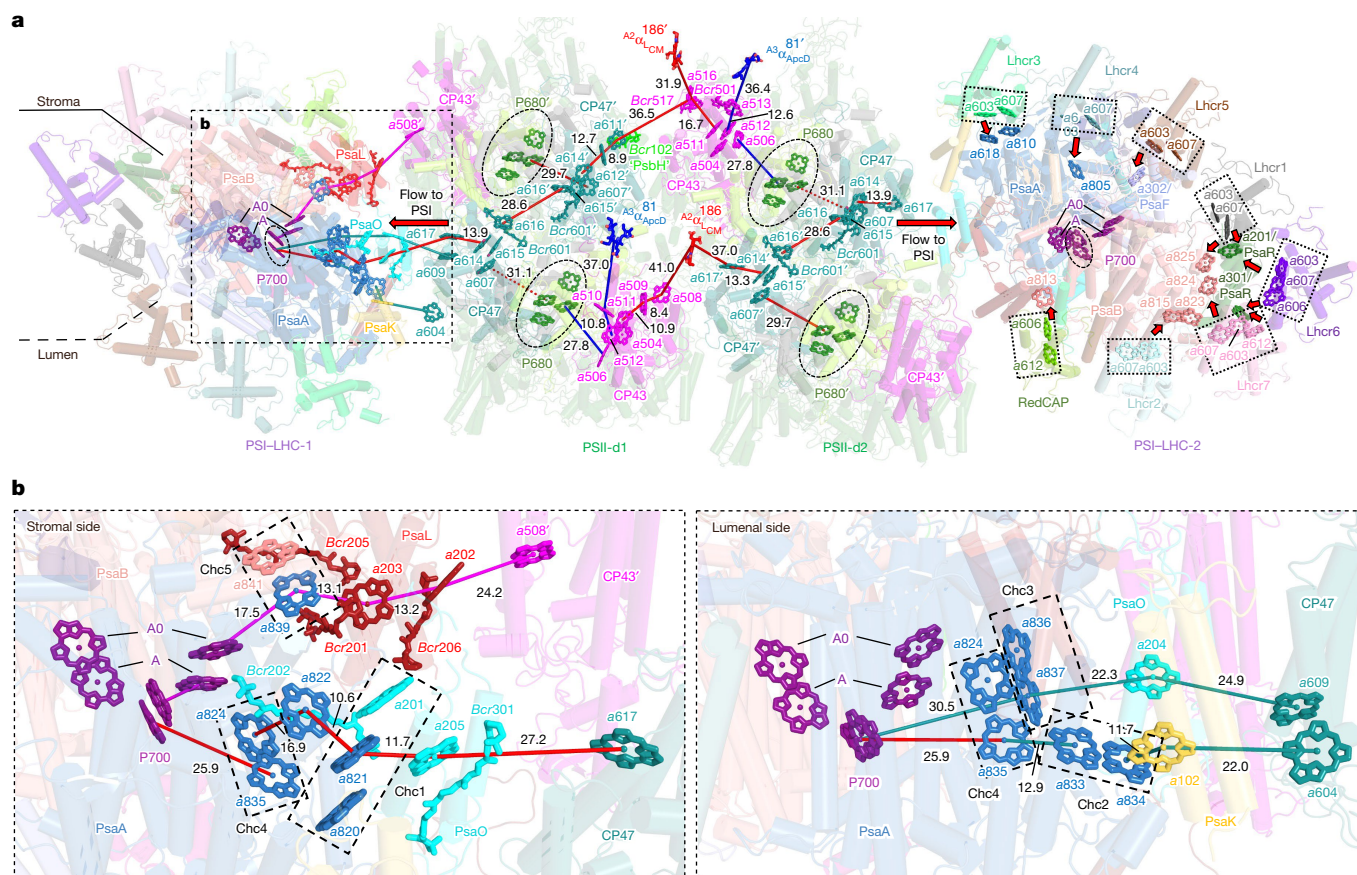
Thus, the link between PSII dimers seems through both PsbO mediated electrostatic contacts and lipid-mediated hydrophobic interactions between adjacent CP43 and CP47' subunits.

### Insights into energy transfer pathways

We use the single PBS-PSII-PSI-LHC megacomplex to analyse possible energy transfer pathways, mainly according to the distances and relative positions between chromophores. It should be noted that other routes may be possible if both distance and orientation are taken into account.

Two pairs of energy terminal emitters ( $A^3\alpha_{\text{ApcD}}$ ,  $A^3\alpha_{\text{ApcD}}'$ ,  $A^2\alpha_{\text{Lcm}}$  and  $A^2\alpha_{\text{Lcm}}'$ ) of PBS are directly involved in the energy migration from PBS to PSII<sup>19,35</sup>. The four terminal emitters are located above PSII-d1 and PSII-d2 with various distances to CP43 and CP47' owing to the rotation and shift of PBS relative to PSII<sup>22</sup> (Fig. 6a and Extended Data Fig. 8a). The energy captured by  $A^2\alpha_{\text{Lcm}}/A^2\alpha_{\text{Lcm}}'$  and  $A^3\alpha_{\text{ApcD}}/A^3\alpha_{\text{ApcD}}'$  could be delivered through some lower energy-state Chls clusters, a504/506/511/512 of CP43 or CP43' and a607/614/615/616 of CP47 or CP47', and finally to P680/PSII (Fig. 6a and Extended Data Fig. 8b-e). These Chls clusters can form excitonic coupling state, and thus could be subjected to downgrade their energy level<sup>18</sup> (Extended Data Fig. 8a-e). In addition, some  $\beta$ -carotene (Bcr) molecules are closely neighboured with Chls  $\alpha$ , which are known to mitigate the extent of photodamage during energy transfer by quenching the Chl triplets, singlet oxygen ( $^1\text{O}_2$ ) and reactive oxygen species produced by high light levels<sup>18,36,37</sup> (Fig. 6a). For the double PBS-PSII-PSI-LHC megacomplex, PSII-d3 has no direct interaction with both PBSs (Fig. 1a). Thus, PSII-d3 could not directly receive the energy from PBS and only transfer energy along the PSII arrays.

Energy transfer from PBS to PSI was thought to occur directly (PBS  $\rightarrow$  PSI) and indirectly (PBS  $\rightarrow$  PSII  $\rightarrow$  PSI)<sup>38,39</sup>. In the indirect PBS  $\rightarrow$  PSII  $\rightarrow$  PSI pathway, energy distribution between PSII and PSI



**Fig. 6 | Key pigment arrangements and possible energy transfer pathways.** **a**, Distribution of key pigments and possible energy transfer pathways in PBS–PSII–PSI–LHC megacomplexes with top views normal to the membrane plane. The key pigments are shown as bold-stick and the  $\pi$ – $\pi$  distances (Å) for the adjacent pigments are labelled in black. P680 and P700 are boxed in ovals. The

low-energy-state Chl pairs of PSI antennas are boxed in rectangles. **b**, Magnified views of the possible energy transfer pathways from PSII to PSI at the stromal (left) and lumenal sides (right). The Chl clusters in the low-energy state are boxed in rectangles.

was proposed to be controlled by the spillover mechanism, in which PSII diverts the excess excitation energy to PSI<sup>40,41</sup>. However, the pigment network mediating the spillover is still unclear. In the PBS–PSII–PSI–LHC structure, Chl a617/CP47 is located at the periphery of PSII and adjacent to PSI, suggesting that it could receive the energy from PBS by the two Chls clusters of CP47/CP47 (mentioned above) and further deliver to PSI (Fig. 6 and Extended Data Fig. 8d,e). Thus, the Chls clusters might act as two protective ‘energy mediators’ to allocate the excess energy from PSII to PSI. The fluorescence emission spectra with phycoerythrobilin excitation showed that most energy from PBS is transferred to PSII (685 and 695 nm) (Extended Data Fig. 9a). Notably, there is also a PSI emission peak (715 nm), indicating that PSII transfers the energy captured by PBS to PSI, which is consistent with our predicted energy transfer pathway of PBS → PSII → PSI based on the structures. Previous studies have demonstrated that energy migration from PBS to PSI occurs directly through ApcD in cyanobacteria<sup>39,42</sup>, and the obvious diffusion of the PBS observed by fluorescence recovery after photobleaching provides an exclusive ‘mobile PBS’ mechanism for energy transfer in cyanobacteria<sup>43</sup>. However, our structure and the time-resolved fluorescence spectra of *P. purpureum*<sup>44</sup> all support the idea that red algae use spillover to share the sequential transfer through PSII (PBS → PSII → PSI) and ApcD is only involved in energy transfer to PSII in red algae. In addition, the position of CP43’ and CP47 of PSII-d1 is similar to two LHClI monomers in the green algal PSI–LHCl–LHClI supercomplex, in which the LHClI is considered to redistribute the energy between PSII and PSI<sup>45</sup>.

Thus, CP43’ and CP47 might also be involved in state transitions. Indeed, the energy from CP43’ and CP47 could be transferred to five Chls clusters (Chc1 to Chc5), which form a ‘low-energy belt’ between PSII and P700/PSI, mediating the efficient energy transfer (Fig. 6b and Extended Data Fig. 8f).

Furthermore, the excitation energy from eight peripheral antennas (Lhcr1–7 and RedCAP) could be transferred to P700, as illustrated in Extended Data Fig. 8g. One parallel-displaced Chl *a* dimer of each antenna is found in the boundary between antennas with PSI, showing similar structural features to the red forms of the Chl pair in plants<sup>18</sup>, which is vital for energy migration to the PSI core (Fig. 6a). The 440 nm excitation (absorbed by Chl *a*) showed a major PSI peak (715 nm) and a lesser PSII peak (685 nm) (Extended Data Fig. 9b). It is reasonable because the eight LHCs provide an extra 108 Chl *a* molecules compared with PSII, which results in 80% of Chl *a* binding to PSI in *P. purpureum*<sup>46</sup>.

In conclusion, the in situ high-resolution structures of red algal PBS–PSII–PSI–LHC megacomplexes reveal the positions of PBS, PSII and PSI in one module and their interaction patterns, as well as several incomplete proteins previously unidentified and their roles in the assembly. The whole complex shows a sophisticated pigment network, ensuring efficient excitation energy transfer within it and regulation of the energy distribution between PSII and PSI. These structures demonstrate that the in vivo arrangement of the supercomplexes is much more complex than those resolved for each of the complexes in their isolated state. Theoretical and spectroscopic studies can be designed on the basis of this structure for a detailed examination of energy transfer in vivo.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-023-05831-0>.

1. Glazer, A. N. Light harvesting by phycobilisomes. *Annu. Rev. Biophys. Biophys. Chem.* **14**, 47–77 (1985).
2. Croce, R. & van Amerongen, H. Natural strategies for photosynthetic light harvesting. *Nat. Chem. Biol.* **10**, 492–501 (2014).
3. Green, B. R. & Durnford, D. G. The chlorophyll-carotenoid proteins of oxygenic photosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 685–714 (1996).
4. Cheng, J., Li, B., Si, L. & Zhang, X. Determining structures in a native environment using single-particle cryoelectron microscopy images. *Innovation* **2**, 100166 (2021).
5. Nelson, N. & Yocum, C. F. Structure and function of photosystems I and II. *Annu. Rev. Plant Biol.* **57**, 521–565 (2006).
6. Grossman, A. R., Bhaya, D., Apt, K. E. & Kehoe, D. M. Light-harvesting complexes in oxygenic photosynthesis: diversity, control, and evolution. *Annu. Rev. Genet.* **29**, 231–288 (1995).
7. Wolfe, G. R., Cunningham, F. X., Durnford, D., Green, B. R. & Gantt, E. Evidence for a common origin of chloroplasts with light-harvesting complexes of different pigmentation. *Nature* **367**, 566–568 (1994).
8. Zhang, J. et al. Structure of phycobilisome from the red alga *Griffithsia pacifica*. *Nature* **551**, 57–63 (2017).
9. Ma, J. et al. Structural basis of energy transfer in *Porphyridium purpureum* phycobilisome. *Nature* **579**, 146–151 (2020).
10. Zheng, L. et al. Structural insight into the mechanism of energy transfer in cyanobacterial phycobilisomes. *Nat. Commun.* **12**, 5497 (2021).
11. Dominguez-Martin, M. A. et al. Structures of a phycobilisome in light-harvesting and photoprotected states. *Nature* **609**, 835–845 (2022).
12. Gindt, Y. M., Zhou, J., Bryant, D. A. & Sauer, K. Spectroscopic studies of phycobilisome subcore preparations lacking key core chromophores: assignment of excited state energies to the Lcm, beta 18 and alpha AP-B chromophores. *Biochim. Biophys. Acta* **1186**, 153–162 (1994).
13. Tang, K. et al. The terminal phycobilisome emitter, LCM: a light-harvesting pigment with a phytochrome chromophore. *Proc. Natl Acad. Sci. USA* **112**, 15880–15885 (2015).
14. Lundell, D. J., Yamanaka, G. & Glazer, A. N. A terminal energy acceptor of the phycobilisome: the 75,000-dalton polypeptide of *Synechococcus* 6301 phycobilisomes—a new biliprotein. *J. Cell Biol.* **91**, 315–319 (1981).
15. Wei, X. et al. Structure of spinach photosystem II-LHCII supercomplex at 3.2 Å resolution. *Nature* **534**, 69–74 (2016).
16. Xu, C. et al. Structural basis for energy transfer in a huge diatom PSI-FCPI supercomplex. *Nat. Commun.* **11**, 5801 (2020).
17. Pi, X. et al. The pigment-protein network of a diatom photosystem II-light-harvesting antenna supercomplex. *Science* **365**, eaax4406 (2019).
18. Qin, X., Suga, M., Kuang, T. & Shen, J. R. Photosynthesis. Structural basis for energy transfer pathways in the plant PSI-LHCI supercomplex. *Science* **348**, 989–995 (2015).
19. Liu, H. et al. Phycobilisomes supply excitations to both photosystems in a megacomplex in cyanobacteria. *Science* **342**, 1104–1107 (2013).
20. Watanabe, M. et al. Attachment of phycobilisomes in an antenna-photosystem I supercomplex of cyanobacteria. *Proc. Natl Acad. Sci. USA* **111**, 2512–2517 (2014).
21. Chang, L. et al. Structural organization of an intact phycobilisome and its association with photosystem II. *Cell Res.* **25**, 726–737 (2015).
22. Li, M., Ma, J., Li, X. & Sui, S. F. In situ cryo-ET structure of phycobilisome-photosystem II supercomplex from red alga. *eLife* **10**, e69635 (2021).
23. Ley, A. C. & Butler, W. L. Efficiency of energy transfer from photosystem II to photosystem I in *Porphyridium cruentum*. *Proc. Natl Acad. Sci. USA* **73**, 3957–3960 (1976).
24. Umena, Y., Kawakami, K., Shen, J. R. & Kamiya, N. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. *Nature* **473**, 55–60 (2011).
25. Van Eerden, F. J., Melo, M. N., Frederix, P., Periole, X. & Marrink, S. J. Exchange pathways of plastoquinone and plastoquinol in the photosystem II complex. *Nat. Commun.* **8**, 15214 (2017).
26. Koua, F. H., Umena, Y., Kawakami, K. & Shen, J. R. Structure of Sr-substituted photosystem II at 2.1 Å resolution and its implications in the mechanism of water oxidation. *Proc. Natl Acad. Sci. USA* **110**, 3889–3894 (2013).
27. Nagao, R. et al. Crystal structure of Psb31, a novel extrinsic protein of photosystem II from a marine centric diatom and implications for its binding and function. *Biochemistry* **52**, 6646–6652 (2013).
28. Falkowski, P. G. et al. The evolution of modern eukaryotic phytoplankton. *Science* **305**, 354–360 (2004).
29. Pi, X. et al. Unique organization of photosystem I-light-harvesting supercomplex revealed by cryo-EM from a red alga. *Proc. Natl Acad. Sci. USA* **115**, 4423–4428 (2018).
30. Nagao, R. et al. Structural basis for assembly and function of a diatom photosystem I-light-harvesting supercomplex. *Nat. Commun.* **11**, 2481 (2020).
31. Nelson, N. & Junge, W. Structure and energy transfer in photosystems of oxygenic photosynthesis. *Annu. Rev. Biochem.* **84**, 659–683 (2015).
32. Li, J. et al. Structure of cyanobacterial photosystem I complexed with ferredoxin at 1.97 Å resolution. *Commun. Biol.* **5**, 951 (2022).
33. Caspy, I., Borovikova-Sheinker, A., Klaiman, D., Shkolnisky, Y. & Nelson, N. The structure of a triple complex of plant photosystem I with ferredoxin and plastocyanin. *Nat. Plants* **6**, 1300–1305 (2020).
34. Matsuzaki, M. et al. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428**, 653–657 (2004).
35. Mullineaux, C. W. Phycobilisome-reaction centre interaction in cyanobacteria. *Photosynth. Res.* **95**, 175–182 (2008).
36. Ostroumov, E. E., Mulvaney, R. M., Cogdell, R. J. & Scholes, G. D. Broadband 2D electronic spectroscopy reveals a carotenoid dark state in purple bacteria. *Science* **340**, 52–56 (2013).
37. Telfer, A. Too much light? How beta-carotene protects the photosystem II reaction centre. *Photoch. Photobiol. Sci.* **4**, 950–956 (2005).
38. Murata, N. Control of excitation transfer in photosynthesis. I. Light-induced change of chlorophyll a fluorescence in *Porphyridium cruentum*. *Biochim. Biophys. Acta* **172**, 242–251 (1969).
39. Dong, C. et al. ApcD is necessary for efficient energy transfer from phycobilisomes to photosystem I and helps to prevent photoinhibition in the cyanobacterium *Synechococcus* sp. PCC 7002. *Biochim. Biophys. Acta* **1787**, 1122–1128 (2009).
40. Ueno, Y., Aikawa, S., Kondo, A. & Akimoto, S. Energy transfer in cyanobacteria and red algae: confirmation of spillover in intact megacomplexes of phycobilisome and both photosystems. *J. Phys. Chem. Lett.* **7**, 3567–3571 (2016).
41. Calzadilla, P. I. & Kirilovsky, D. Revisiting cyanobacterial state transitions. *Photoch. Photobiol. Sci.* **19**, 585–603 (2020).
42. Deng, G., Liu, F., Liu, X. & Zhao, J. Significant energy transfer from CpcG2-phycobilisomes to photosystem I in the cyanobacterium *Synechococcus* sp. PCC 7002 in the absence of ApcD-dependent state transitions. *FEBS Lett.* **586**, 2342–2345 (2012).
43. Mullineaux, C. W., Tobin, M. J. & Jones, G. R. Mobility of photosynthetic complexes in thylakoid membranes. *Nature* **390**, 421–424 (1997).
44. Yokono, M., Murakami, A. & Akimoto, S. Excitation energy transfer between photosystem II and photosystem I in red algae: larger amounts of phycobilisome enhance spillover. *Biochim. Biophys. Acta* **1807**, 847–853 (2011).
45. Huang, Z. et al. Structure of photosystem I-LHCI-LHCII from the green alga *Chlamydomonas reinhardtii* in State 2. *Nat. Commun.* **12**, 1100 (2021).
46. Cunningham, F. X., Dennenberg, R. J., Jursinic, P. A. & Gantt, E. Growth under red light enhances photosystem II relative to photosystem I and phycobilisomes in the red alga *Porphyridium cruentum*. *Plant Physiol.* **93**, 888–895 (1990).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2023

## Methods

### Red algae culture and cryo-EM sample preparation

*P. purpureum* was cultured as previously described<sup>22</sup>. Holey-carbon copper grids (Quantifoil R1.2/1.3, 200 mesh) were used for the cryo-EM sample preparation. Then 4  $\mu$ l of cell culture was applied onto the front side of glow-discharged grids, and 2  $\mu$ l of medium was subsequently applied on the back side of the grids. Then the grids were back-side blotted for 10 s in a Leica EM GP2 (Leica Company) plunger at 20 °C and 100% humidity and plunged into liquid ethane at −184 °C for vitrification. Cryo-EM grids were stored in liquid nitrogen until cryo-focused ion beam (FIB) milling.

### Cryo-FIB milling

Cryo-FIB milling was performed using an FIB–scanning EM dual-beam microscope (Aquilos, Thermo Fisher Scientific). The cryo-EM grids were clipped to the autogrids in the liquid nitrogen reservoir, and then transferred into the FIB–scanning EM chamber in a high vacuum at −185 °C. To protect the sample, a layer of Pt was sputtered to the surface of samples to increase the conductivity. Subsequently, the samples were deposited with organometallic platinum using a gas-injection system operated with a 35 s gas-injection time before milling at a work distance of 11 mm. The gas-injection system temperature was set to 46 °C. During the milling process, the stage was tilted at 22° (milling angle of roughly 15° related to the EM grid). Rough milling was performed with the accelerating voltage of the ion beam at 30 kV and a current of 0.5–3.0 nA, and a ‘micro-expansion joint’ was milled to eliminate tensions<sup>47</sup>. Cells were subsequently polished with an ion beam of 30–50 pA to produce lamellae with a final thickness of 100–200 nm.

### Cryo-ET data collection and tomogram reconstruction

The lamellas of *P. purpureum* PBSs prepared by cryo-FIB were loaded into a 300-kV Titan Krios microscopy G3i (Thermo Fisher Scientific) equipped with GIF quantum energy filter and K3 Summit direct electron detector (Gatan Inc.) for data collection. Images of lamellas at low and media magnification are presented in Extended Data Fig. 1a,b. For each lamella, the grid was tilted to either −15° or +15° before data collection, which makes the first acquired image of the tilt series at the angle parallel to the lamella plane. Tilt images were acquired by the K3 camera operated under the super-resolution mode at a nominal magnification of  $\times 53,000$  (calibrated pixel size 1.632 Å). A postcolumn energy filter (GIF, Gatan) with a slit of 20 eV was adjusted to zero lost peak and kept active during the data collection. Each tilt image was recorded as a video stack consisting of 15 frames with a dose of  $4.4 \text{ e}^-/\text{Å}^2$ . SerialEM v.3.7.13 (ref. <sup>48</sup>) was used to collect the tilt series data. The dose-symmetric tilting scheme was applied for data collection, starting at the angle of lamella plane while ending at  $\pm 42^\circ$  (relative to the lamella plane) with an increment step of 3°. Each tilt series contains 29 images with a total dose of roughly  $130 \text{ e}^-/\text{Å}^2$ . The defocus range was set from −2.5 to −5.5  $\mu$ m.

Video frames of each tilt were motion-corrected, summed and binned with a factor of two by using MotionCor2 v.1.1.0 (ref. <sup>49</sup>). Tilt series were merged into one stack and aligned by patch tracking in IMOD v.4.10.38 (ref. <sup>50</sup>). In total, 170 sets of tilt series were collected. Tilt series that contained images with large motion or lost tracking were excluded. Tomograms were three-dimensionally reconstructed by the weighted back-projection method, and selected tomograms were filtered with a deconvolution script ‘tomo\_deconv’ ([https://github.com/dtegunov/tomo\\_deconv](https://github.com/dtegunov/tomo_deconv)) for display and segmentation.

### Subtomogram averaging of PBS complexes

Forty-six tilt series with good alignment (alignment error smaller than 0.7 nm) were subjected to subtomogram averaging in emClarity v.1.5.3 (ref. <sup>51</sup>) (Extended Data Fig. 1c). Particle picking was performed by using template matching with previous reconstruction of PBS

complex from subtomogram averaging<sup>22</sup> (low-pass filtered to 20 Å) as a reference. The picked particles of each tomogram were assessed visually in IMOD v.4.10.38 (ref. <sup>50</sup>), and the false pickings were manually removed. In total, 23,637 particles were labelled out for further processing. The alignment of subtomogram particles was consecutively refined at bin 4, bin 3 and bin 2. After cycles of alignment at bin 2, two-dimensional (2D) particle images related to three-dimensional (3D) subtomograms were cropped out at bin 1 (with a box size of 400 pixels, roughly 65 nm) from the original first nine tilt images (tilt images with an exposed dose under  $40 \text{ e}^-/\text{Å}^2$ ), by calculating the 2D projection coordinates on the basis of 3D coordinates in tomograms. These 2D particles were transferred into RELION v.3.0 (ref. <sup>52</sup>) for local refinement with a mask of PBS shape, yielding a 3D reconstruction of 6.5 Å resolution.

To classify heterogeneous conformations, another branch job was carried out after cycles of alignment at bin 2. The subtomogram particles were written out with a larger volume size of 128 pixels at bin 6 (equal to roughly 125 nm) and transferred to RELION v.3.0 (ref. <sup>52</sup>). A 3D classification with local search at an angular sampling interval of 3.7° was used to separate single and double PBS–PSII–PSI–LHC megacomplexes. This 3D classification result was then used to demonstrate the distribution of single and double PBS–PSII–PSI–LHC in tomograms (Extended Data Fig. 3).

### Tomogram segmentation and 3D visualization

The software Amira v.2020.2 (ref. <sup>53</sup>) (FEI Visualization Sciences Group) was used to annotate the membrane layers in the 3D reconstruction of cell lamella. The reconstruction results of single and double PBS–PSII–PSI–LHC megacomplexes were repositioned back to tomograms according to the coordinates and Euler angles from refinement, and the segmentation result of thylakoid membranes was shown as extra image layer in ChimeraX v.1.3 (ref. <sup>54</sup>) (Extended Data Fig. 3). The histogram was plotted in MATLAB2021a (Mathworks).

### Data collection for single-particle analysis

The high-dose image data were collected at the same magnification as cryo-ET data by using SerialEM v.3.7.13 (ref. <sup>48</sup>) with a defocus range from −2.0 to −4.0  $\mu$ m. Micrographs were acquired as 80 video frames with a total dose of  $35 \text{ e}^-/\text{Å}^2$ . Motion correction was done in cryoSPARC v.3.3.1 (ref. <sup>55</sup>) and images were dose-weighted and binned with a factor of 2. Images with large blurred area, too low electron counts (usually means very large sample thickness) or too high or low defocus ( $< -6.0$  or  $> -1.0 \mu$ m) were discarded. In total, 2,245 images were selected from roughly 3,000 acquired images for further processing.

### Particle picking from images by isSPA

Owing to the crowded environment of cell lamella, the density of protein complexes overlaps with other densities from biomacromolecules, so protein complexes in the micrographs cannot be directly picked out by the traditional particle-picking method or deep learning method. To translationally and rotationally localize the target protein in these micrographs, we performed a high-resolution template matching method using the cross-correlate score in isSPA<sup>4</sup> that was optimized on the basis of overlapping densities. The result from subtomogram averaging was applied with a bandpass filter from 8 to 150 Å and a mask of PBS shape (only outside membrane region), and was then used as the template. The gold-standard Fourier shell correlation from the subtomogram averaging was intergraded into the score function, as described in isSPA<sup>4</sup>. The contrast transfer function parameters of defocus and astigmatism of the target images were estimated by using CTFFIND4 (ref. <sup>56</sup>) before performing the template matching. Both the input micrographs and template were binned by a factor of two to reduce the computational consumption. The defocus variation in the 2D images was not concerned at the current step. The threshold for particle detection was set as 6.3, which generated an average of

roughly 400 detections for each micrograph. Sampling points for three Euler angles were defined by an interval step of  $7.5^\circ$  and uniformly distributed in the Euler angle space (considering the C2 symmetry of PBS, only half of the full Euler angle combinations need to be searched) using the same Euler angle definitions as RELION v.3.0 (ref. <sup>52</sup>). In the peak identification procedure, local peaks within a distance of 30 pixels (unbinned) in cross-correlograms and an orientation spacing angle of  $30^\circ$  (in quaternion space) were considered to be one peak (one potential particle). Subsequently, the particle coordinates  $x, y$ , defocus and corresponding Euler angles were transferred into RELION v.3.0 (ref. <sup>52</sup>) for particle extraction.

### Structure refinement of PBS, PSII and PSI–LHC

The overall workflow of image processing is demonstrated in Extended Data Fig. 1c and was similar to that described in isSPA<sup>4</sup>. In total, 762,000 potential particles were picked out and extracted at bin 1 with a box size of 400 pixels. A local refinement (with an angular sampling interval of  $3.7^\circ$ ) yielded a  $3.8 \text{ \AA}$  reconstruction. As the FIB lamella usually has a thickness in the range of roughly 100–200 nm, three rounds of contrast transfer function refinement and 3D auto-refine were carried out to minimize the defocus offset of each individual particle. Phase-residue particle sorting was calculated by homemade MATLAB scripts (equivalent programs are available at <https://github.com/chengji-dot/isSPA>) between images of the original particles and projections generated from 3D reconstruction, as described in isSPA<sup>4</sup>. After roughly 30% particles were removed by particle sorting to exclude false-positive picking, the remaining 522,000 particles were re-extracted with a much larger box size of 200 pixels at bin 4 (roughly 130 nm) to include surrounding proteins that interacted with PBS. The 3D classification was carried out to separate particles of single and double PBS–PSII–PSI–LHC megacomplexes, as discovered in the subtomogram averaging study. Two separate refinements of single and double PBS–PSII–PSI–LHC megacomplexes (using particles selected from more steps of 3D classifications) yielded the reconstructions of single PBS–PSII–PSI–LHC at  $3.3 \text{ \AA}$  resolution (from 215,000 particles with C1 symmetry) and double PBS–PSII–PSI–LHC at  $4.3 \text{ \AA}$  resolution (from 87,000 particles with C2 symmetry applied), respectively (Extended Data Fig. 1c).

To obtain a high-resolution PSII structure, we performed focused refinement on PSII: we gathered the PSII region from both single and double PBS–PSII–PSI–LHC (each PBS has two PSII dimers connected to the bottom, Extended Data Fig. 2a), and aligned them in one refinement with a mask of the PBS shape. To centre the region of PSII density in the refinement, the PBS particle coordinates were shifted  $-217 \text{ \AA}$  along the  $z$  axis of the 3D reconstruction. The 3D classification and refinements of PSII were done first with a larger soft-edge mask containing a region of PSII and partial APC and then with a tighter mask, while applying symmetry relaxation from C2 to C1. A total of 288,000 particles of PSII were re-extracted from micrographs motion corrected in cryoSPARC v.3.3.1 (ref. <sup>55</sup>) and Fourier cropped by a factor of 3/4 (equivalent to a pixel size of  $1.088 \text{ \AA}$ ). Local refinement of these particles yielded a structure of  $3.2 \text{ \AA}$  resolution (Extended Data Fig. 2b).

Particles of PSI–LHC were extracted by shifting the coordinates on the basis of the 3D geometry of a single or double PBS–PSII–PSI–LHC structure (each single or double PBS–PSII–PSI–LHC contains two PSI–LHCs, as shown in Extended Data Fig. 2c). The PSI–LHC particles from the left or right side of a single or double PBS–PSII–PSI–LHC were reoriented and refined separately with round masks. Several steps were carried out to gradually classify and align homogeneous conformation. First, a local refinement for joined particles was performed with a mask containing a PSII dimer and the PSI–LHC region. Second, 3D classification was performed with a mask containing only the PSI–LHC region. Third, 3D classification was performed with a tighter mask containing only the PSI–LHC region. After local refinements in RELION v.4.0 (ref. <sup>52</sup>) and then in cryoSPARC v.3.3.1 (ref. <sup>55</sup>) for selected 112,000 particles, a 3D structure of  $3.6 \text{ \AA}$  resolution was obtained.

The particles of PSII dimer in the centre of the double PBS–PSII–PSI–LHC complex were extracted with coordinates shifted to target region and refined with a local mask (Extended Data Fig. 2d). The region of the hexamer that directly connected to this PSII dimer and located between the two PBS structures was refined with a local mask (resolution  $6.3 \text{ \AA}$ ) to aid model building (Extended Data Fig. 2d).

All the 3D classifications have been done in RELION v.4.0 (ref. <sup>52</sup>). The early refinements were done in RELION v.4.0 using 3D auto-refine with either a round mask or soft-edge local mask generated from density segmentation. The last refinements for each subregion were performed in cryoSPARC v.3.3.1 (ref. <sup>55</sup>) using local refinement. The resolutions were estimated in cryoSPARC v.3.3.1 by gold-standard Fourier shell correlation with a criterion of 0.143 (Extended Data Fig. 2e–j).

To obtain the intact high-resolution density maps of single and double PBS–PSII–PSI–LHC for visualization, local maps of PBS, PSII and PSI–LHC from separated refinements were fitted together and merged in MATLAB2021a.

### Model building and refinement

For model building of the PBS, the atomic model of *P. purpureum* PBS (PDB 6KGX)<sup>9</sup> was first docked into the EM maps using Chimera v.1.12 (ref. <sup>57</sup>). All of the PBP and linker proteins were fitted well. Sequence assignments of the PB-loop of  $L_{CM}$  and the C-terminal regions of  $L_{RC3}$  and  $L_{RC2}$  were guided by corresponding residues in Coot v.0.8.9.1 (ref. <sup>58</sup>) and manually adjusted to better fit with the map. The lateral phycoerythrin hexamers were well-fitted into the maps at a lower threshold level in Chimera v.1.12 (ref. <sup>57</sup>). To build the model of PSII dimers, structures of the red algal PSII (PDB 4YUU)<sup>59</sup>, cyanobacterial PSII (PDB 7RCV)<sup>60</sup>, *Pisum sativum* PSII–LHCII (PDB 5XNM)<sup>61</sup> and diatom PSII–FCP (PDB 6JLU)<sup>17</sup> were fitted into the local density maps using Chimera v.1.12 (ref. <sup>57</sup>) and the sequences were mutated to the counterparts in *P. purpureum* with Coot v.0.8.9.1 (ref. <sup>58</sup>) on the basis of the Basic Local Alignment Search Tool (BLAST) and the results of the mass spectrometry performed in the present study. An unidentified chain N (modelled as poly-Ala) is designated Psb34 owing to its similar position to Psb34 in the red algal and diatom PSII structures<sup>17,59</sup>. The density of Psb31 in the cryo-EM map is weaker (but clearly distinguishable) than the other PSII core subunits, indicating that they may have lower occupancy. Furthermore, de novo model building was performed on the unidentified subunits defined as the CNT and  $L_{pp1}$ , and the N or C termini were artificially defined in this study. CNT was located on the stromal surface of each PSII monomer and partly inserted into the plastoquinone  $Q_B$  exchange cavity formed by TMs PsbE, PsbF, PsbJ, PsbK, D1 and D2 (ref. <sup>25,62,63</sup>).  $L_{pp1}$  is located on the stromal side of PSII dimers and its C terminus is bound to the PBS core A1 layer. However, no protein density corresponding to CNT or  $L_{pp1}$  was observed in published PSII structures, indicating that they might be lost during purification. The density of  $L_{pp2}$  is eventually assigned as an appropriate protein (hypothetical protein FVE85\_5305) on the basis of the conserved C-terminal helix sequence searching and well-resolved bulky residues. For model building of the PSI–LHC moieties, the structure of the red algal PSI–LHCR complex (PDB 5ZGB)<sup>29</sup> was fitted into the  $3.6 \text{ \AA}$  local density map using UCSF Chimera v.1.12 (ref. <sup>57</sup>) and then mutated to the corresponding sequences of *P. purpureum* based on the BLAST program and mass spectrometry analysis (Supplementary Table 2). Compared to the PSI–LHCR complex, the structures of PSI core subunits are basically conserved except for Psar, which was found in diatom PSI<sup>16</sup>. The diatom Psar (PDB 6LY5)<sup>16</sup> was then fitted into the map and replaced by the corresponding sequence of *P. purpureum*. To assign the appropriate sequence to the antenna, we performed BLAST searches in the National Centre for Biotechnology Information database, which retrieved seven Lhcr sequences. The structures of these LHCRs were first predicted by AlphaFold2 (ref. <sup>64</sup>) to recognize specific helices and  $\beta$ -sheets and then were compared with the density maps to facilitate the identification of the antennas (defined as Lhcr1 to Lhcr7). The sequence assignments were further improved according to the

well-resolved bulky residues. Notably, the remaining antenna subunit showed an obvious structural difference compared with Lhcrs, indicating that it may belong to other LHC superfamilies. The structure-based alignment was executed by the DALI server<sup>65</sup>, which provides a FCI-1 structure from diatom *C. gracilis* (PDB 6LY5)<sup>16</sup> with the highest DALI Z scores of 8.1 and the r.m.s.d.s of 2.0. The antenna is eventually assigned as an appropriate protein (hypothetical protein FVE85\_8746) on the basis of BLAST searches and well-resolved bulky residues, which was classified into a red lineage-restricted LHC protein superfamily, termed red lineage Chl a/b-binding-like protein (RedCAP)<sup>66</sup>. The similarity of this structure to those of the fucoxanthin-chlorophyll a/c-binding proteins (FCPs) of diatom may show that FCPs may originate from this protein in the red algae. Similarly, the original structure of RedCAP was produced by AlphaFold2 (ref. <sup>64</sup>) and further manually adjusted in Coot v.0.8.9.1 (ref. <sup>58</sup>) according to the bulky residues. The electron carrier Fd was located on the stromal side of PSI. The structure of Fd provided by AlphaFold2 (ref. <sup>64</sup>) was used to dock into the density map using Chimera v.1.12 (ref. <sup>57</sup>) and manually adjusted in Coot v.0.8.9.1 (ref. <sup>58</sup>).

The model building of PBS, PSII-d1-d2, PSII-d3 and PSI-LHC was completed through iterative rounds of manual building with Coot v.0.8.9.1 (ref. <sup>58</sup>) and refinement with phenix.real\_space\_refine program in Phenix v.1.14-3260 (ref. <sup>67</sup>) with geometry and secondary structure restraints. Then, all parts were merged to the whole single or double PBS-PSII-PSI-LHC megacomplexes and the structural validation reports were produced against two whole artificially stitched maps using local maps in MATLAB2021a. The statistics for data collection and structure refinement are summarized in Extended Data Table 1. Figures were prepared with UCSF ChimeraX v.1.3 (ref. <sup>54</sup>) and PyMOL v.1.8.2.1 (<http://pymol.org>). The sequence alignments were performed by ClustalX v.2.0 (ref. <sup>68</sup>) and created by ESPript v.3.0 (ref. <sup>69</sup>). Phylogenetic analysis was based on the neighbour-joining algorithm and p distance model in MEGA v.11 (ref. <sup>70</sup>).

### Mass spectrometry analysis

Cells were collected by centrifugation for 10 min at 6,000g, and resuspended in buffer A (30 mM MOPS at pH 7.0, 10 mM CaCl<sub>2</sub>, 5 mM MgCl<sub>2</sub>, 25% glycerol) and 1 mM phenylmethylsulfonyl fluoride. Cells were disrupted by ultrasonication with 5 s on and 10 s off on ice at a power of 90 W. The suspension was centrifuged at 1,500g for 5 min to dispose the unbroken cells, and the supernatant was further centrifuged at 100,000g for 30 min to collect the thylakoid membranes. The thylakoid membrane was separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis using a 4–12% polyacrylamide gel (GenScript). Gel bands were excised for in-gel digestion, and proteins were identified by mass spectrometry. In brief, proteins were reduced with 25 mM dithiothreitol and alkylated with 55 mM iodoacetamide. In-gel digestion was performed using sequencing grade-modified trypsin in 50 mM ammonium bicarbonate at 37 °C overnight. The peptides were extracted twice with 1% trifluoroacetic acid in 50% acetonitrile aqueous solution for 30 min. The peptide extracts were then centrifuged in a SpeedVac to reduce the volume.

For liquid chromatography–tandem mass spectrometry (MS/MS) analysis, peptides were separated by a 120 min gradient elution at a flow rate of 0.300 µl min<sup>-1</sup> with a Thermo-Dionex Ultimate 3000 high-performance liquid chromatography (HPLC) system, which was directly interfaced with the Thermo Orbitrap Fusion mass spectrometer. The analytical column was a homemade fused silica capillary column (75 µm ID, 150 mm length; Upchurch) packed with C-18 resin (300 Å, 5 µm; Varian). Mobile phase A consisted of 0.1% formic acid, and mobile phase B consisted of 100% acetonitrile and 0.1% formic acid. The Orbitrap Fusion mass spectrometer was operated in the data-dependent acquisition mode using Xcalibur 3.0 software and there is a single full-scan mass spectrum in the Orbitrap (350–1,550 *m/z*, 120,000 resolution) followed by 3 s data-dependent MS/MS scans in an ion routing multipole at 30% normalized collision energy (high

collision dissociation). The MS/MS spectra from each LC–MS/MS run were searched against the selected database using Proteome Discovery searching algorithm (v.1.4) (Supplementary Table 2).

### Identification and quantification of pigments by HPLC

Pigments of *P. purpureum* were analysed by HPLC as described before<sup>16</sup>. In brief, pigments in the thylakoid membrane were extracted with cold 90% acetone (v/v), and the supernatant was centrifuged and loaded onto C-18 column (3.5 µm, 100 × 4.6 mm, Agilene) equipped in an Agilene 1260 Infinity system. Pigments were eluted at 20 °C at a flow rate of 1 ml min<sup>-1</sup> with the following procedure: linear gradient of buffer A (methanol:water 9:1) from 100% to 0 for 20 min; 100% buffer B (ethyl acetate) for 2 min; two linear gradient of buffer B from 100% to 0 for 1 min and 100% buffer A for 5 min. The elutes were detected at 445 nm using a diode array detector. The elution pattern shows a large amount of zeaxanthin (Zex), Chl *a* and β-carotene (Bcr) in the thylakoid of *P. purpureum*, whereas Zex was not identified in the maps owing to limited resolution (Extended Data Fig. 9c).

### 77 K fluorescence spectroscopy

*P. purpureum* cultured as previously described<sup>22</sup> was collected to measure fluorescence spectroscopy. Fluorescence emission spectra were measured at 77 K with a fluorescence spectrometer (F-7000, Hitachi) equipped with a xenon lamp source, and the spectra were recorded at a wavelength range from 500 to 800 nm with excitation wavelengths of 440 and 560 nm. The slit widths of both excitation and emission were set at 5.0 nm.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The cryo-EM density map and atomic models generated in this study have been deposited in the Electron Microscopy Data Bank (EMDB) and the PDB for the single PBS–PSII–PSI–LHC megacomplex structure at 3.3 Å resolution (EMDB ID code 33605 and PDB 7Y5E), the double PBS–PSII–PSI–LHC megacomplex structure at 4.3 Å resolution (EMDB ID code 33669 and PDB 7Y7A), the PBS structure at 3.3 Å resolution (EMDB ID code 33605 and PDB 7Y4L), the PSII-d1-d2 structure at 3.2 Å resolution (EMDB ID code 33597), the PSII-d3 structure at 3.4 Å resolution (EMDB ID code 33568), the PSI–LHC structure at 3.6 Å resolution (EMDB ID code 33561) and the lateral hexamer structure at 6.3 Å resolution (EMDB ID code 33558 and PDB 7Y1A). Two whole artificially stitched maps have been deposited in the EMDB (EMDB ID code 33618 for single PBS–PSII–PSI–LHC and 33658 for double PBS–PSII–PSI–LHC). For the publicly available atomic models used in this study, their accession codes in the PDB have been provided in the paper.

### Code availability

The code used in this study to calculate phase-residue particle sorting is available at <https://github.com/chengj-dot/isSPA>.

47. Wolff, G. et al. Mind the gap: micro-expansion joints drastically decrease the bending of FIB-milled cryo-lamellae. *J. Struct. Biol.* **208**, 107389 (2019).
48. Mastrorade, D. N. SerialEM: a program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. *Microsc. Microanal.* **9**, 1182–1183 (2003).
49. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
50. Kremer, J. R., Mastrorade, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
51. Himes, B. A. & Zhang, P. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nat. Methods* **15**, 955–961 (2018).
52. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

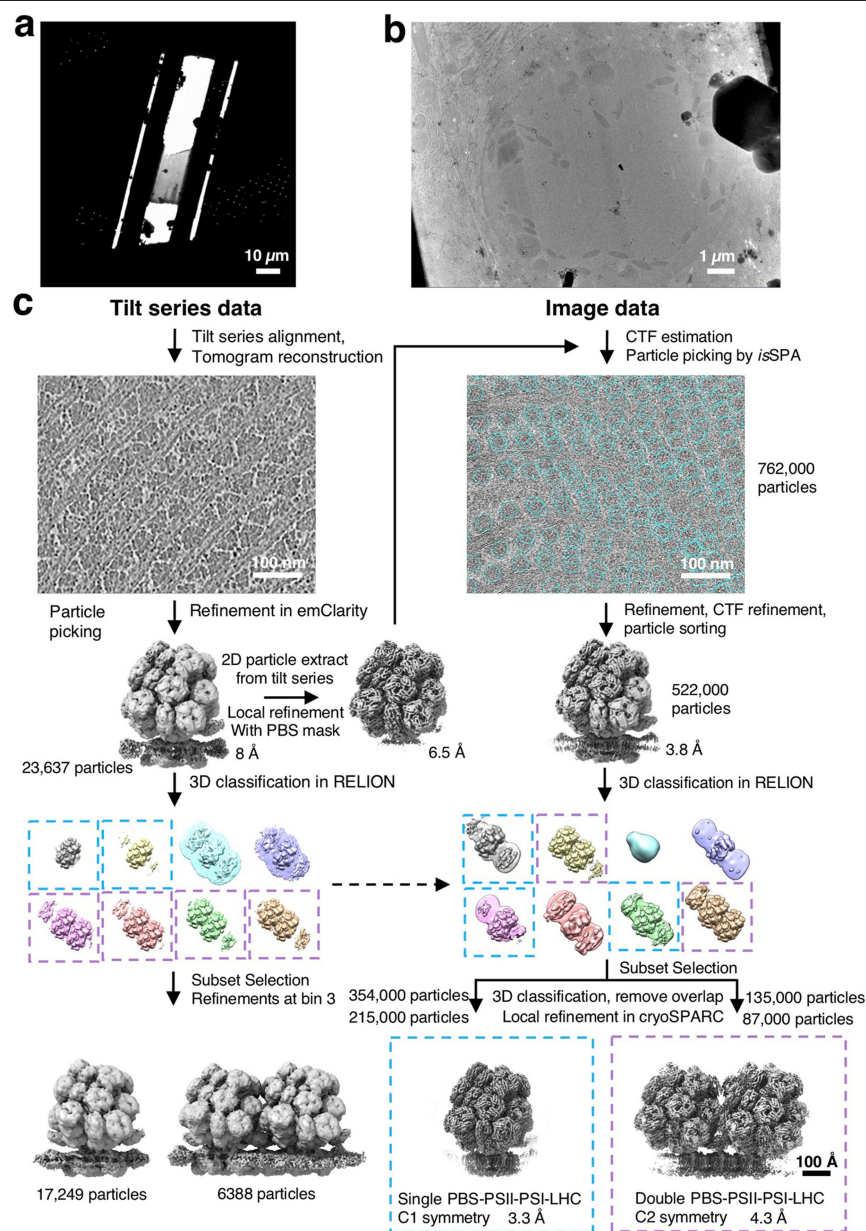
53. Stalling, D., Westerhoff, M. & Hege, H. C. in *The Visualization Handbook* (eds Hansen, C. D. & Johnson, C. R.) 749–767 (Elsevier, 2005).
54. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
55. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
56. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
57. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
58. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
59. Ago, H. et al. Novel features of eukaryotic photosystem II revealed by its crystal structure analysis from a red alga. *J. Biol. Chem.* **291**, 5676–5687 (2016).
60. Gisriel, C. J. et al. High-resolution cryo-electron microscopy structure of photosystem II from the mesophilic cyanobacterium, *Synechocystis* sp. PCC 6803. *Proc. Natl Acad. Sci. USA* **119**, e2116765118 (2022).
61. Su, X. et al. Structure and assembly mechanism of plant C2S2M2-type PSII-LHCII supercomplex. *Science* **357**, 815–820 (2017).
62. Guskov, A. et al. Cyanobacterial photosystem II at 2.9-Å resolution and the role of quinones, lipids, channels and chloride. *Nat. Struct. Mol. Biol.* **16**, 334–342 (2009).
63. Loll, B., Kern, J., Saenger, W., Zouni, A. & Biesiadka, J. Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II. *Nature* **438**, 1040–1044 (2005).
64. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
65. Holm, L. Using DALI for protein structure comparison. *Methods Mol. Biol.* **2112**, 29–42 (2020).
66. Sturm, S. 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).
67. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
68. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
69. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
70. Tamura, K., Stecher, G. & Kumar, S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).

**Acknowledgements** We thank the staff at the Tsinghua University Branch of the National Protein Science Facility (Beijing) for their technical support on the cryo-EM and the high-performance computation platforms. We thank H. Deng and X. Meng in Proteinomics Facility at Technology Center for Protein Sciences, Tsinghua University, for protein MS analysis. We thank the staff at the SUSTech Core Research Facilities (Shenzhen) for their technical support in HPLC analysis. We thank J.-R. Shen from the Institute of Botany, Chinese Academy of Sciences for useful discussion. We thank Y. Yin from Plant Science Facility of the Institute of Botany, Chinese Academy of Sciences, for her excellent technical assistance on 77 K fluorescence spectra. This work was supported by the National Natural Science Foundation of China (grant nos. 32271245, 32241030 and 32071192 to S.-F.S., 32150010 to Xinzhen Z., 31825009 to H.-W.W., and 32000848 to J.M.), the National Basic Research Program (grant no. 2017YFA0504600 to S.-F.S.), the National Key R&D Program of China (grant no. 2017YFA0504700 to Xinzhen Z.), and the Xplorer Prize to H.-W.W.

**Author contributions** S.-F.S. supervised the project. X.Y. and Y.X. froze samples, and performed cryo-FIB milling and the sequence analysis. Xing Z. optimized the data collection scripts. Xing Z., X.Y. and Y.X. collected the EM data. Xing Z. and J.C. performed the EM analysis under the supervision of H.-W.W. and Xinzhen Z. X.Y. performed the model building and the structure refinement. X.Y. and Y.X. performed the biochemical experiments. J.M. provided the red algae cells and did the initial tries. X.Y., Y.X., S.S. and S.-F.S. analysed the structure. All authors contributed to writing the manuscript.

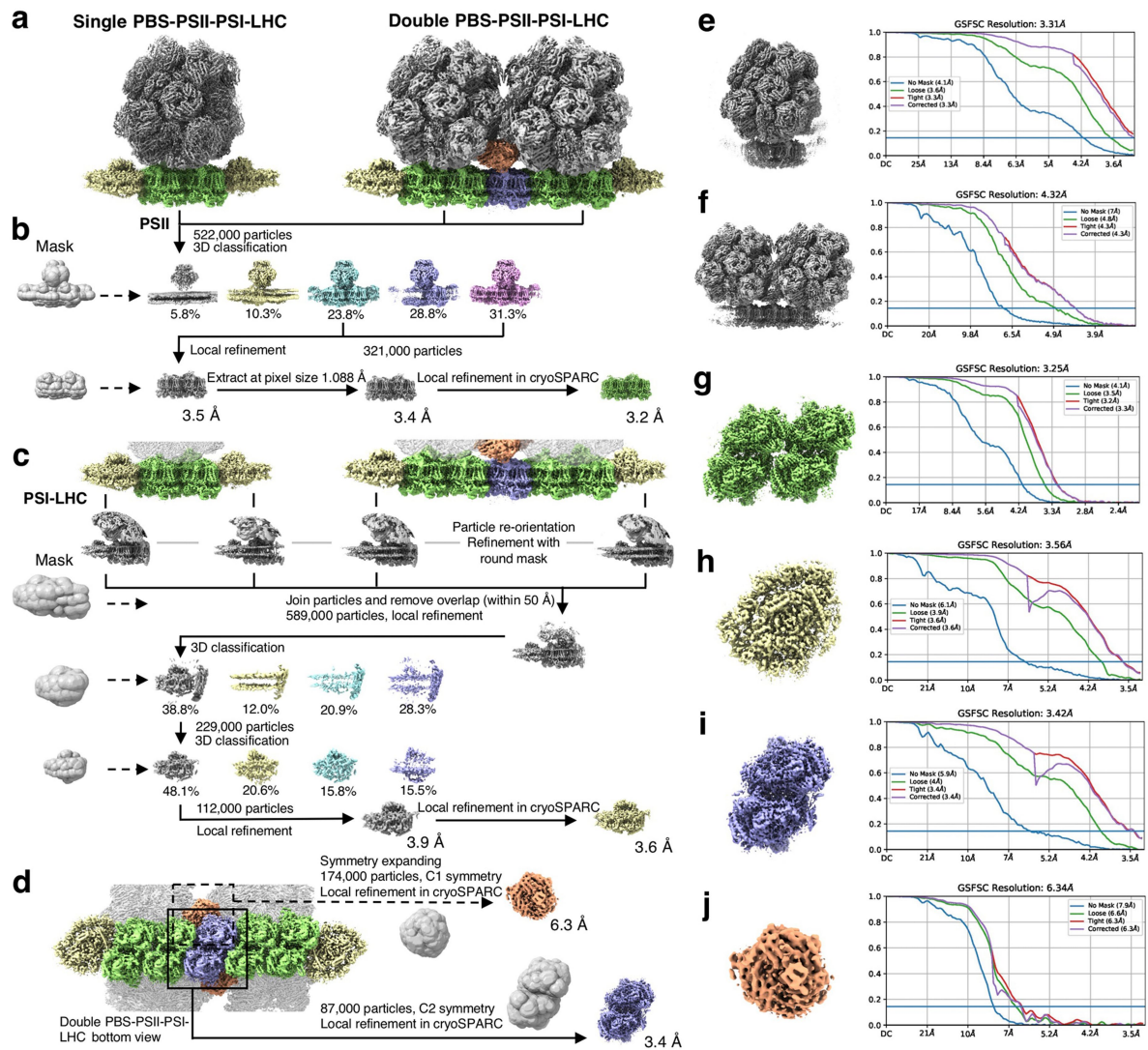
**Competing interests** The authors declare no competing interests.

**Additional information**  
**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-023-05831-0>.  
**Correspondence and requests for materials** should be addressed to Xinzhen Zhang, Hong-Wei Wang or Sen-Fang Sui.  
**Peer review information** Nature thanks Seiji Akimoto, Wolfgang Baumeister and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.  
**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



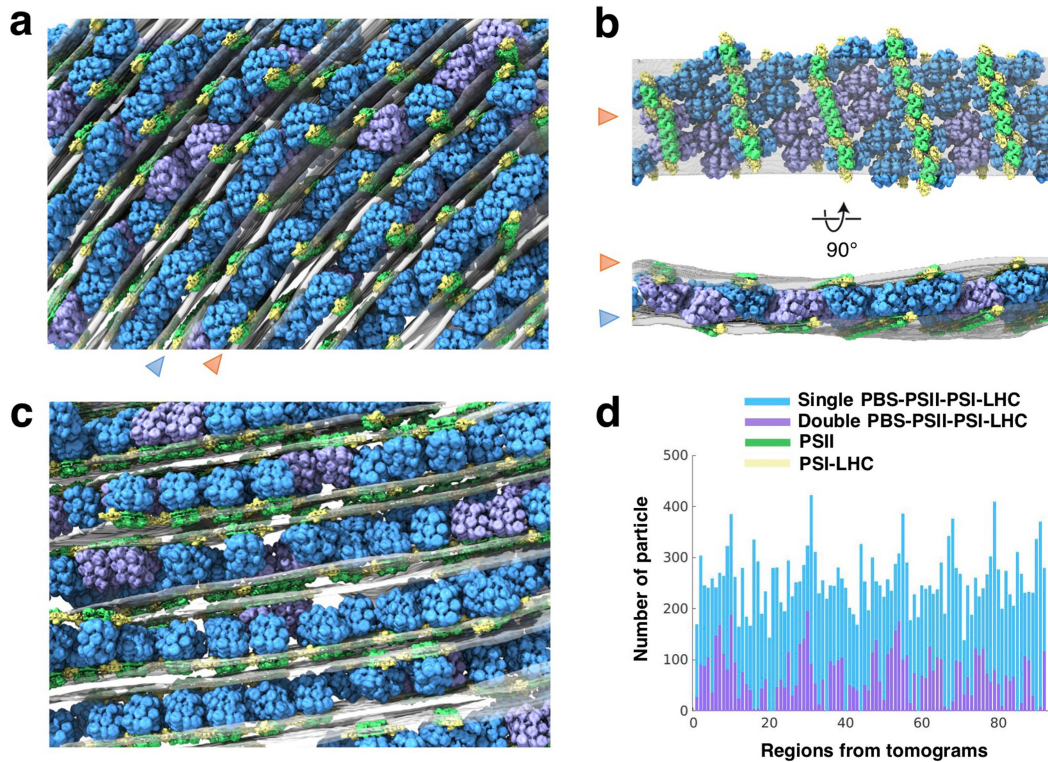
**Extended Data Fig. 1 | Overview of FIB lamella sample preparation and flowchart of cryo-EM study.** **a**, A representative EM image of *P. purpureum* lamella at low magnification. 46 tilt series images were collected with similar results. **b**, A representative EM image of *P. purpureum* lamella (selected from 46) at media magnification. 46 similar tilt series were collected. **c**, A flowchart of the cryo-EM analysis. 23,637 sub-tomogram particles were picked out from tomograms reconstructed from 46 tilt series in emClarity. Alignment and averaging of subtomograms yielded an 8 Å structure. 3D classification showed two distinct conformations, the single and double PBS-PSII-PSI-LHC

megacomplexes. The tilt images corresponding to the all subtomograms were extracted and subjected to local refinement in RELION with a soft-edge mask of PBS, yielding a 6.5 Å structure of PBS. This structure was then used as a high-resolution reference to pick out the potential particles (762,000) in high-dose image data (2,245) by method *isSPA*. After further particle filtering, 3D classifications and refinements, we obtained high resolution structures of single and double PBS-PSII-PSI-LHC megacomplexes from 215,000 and 87,000 particles, respectively.



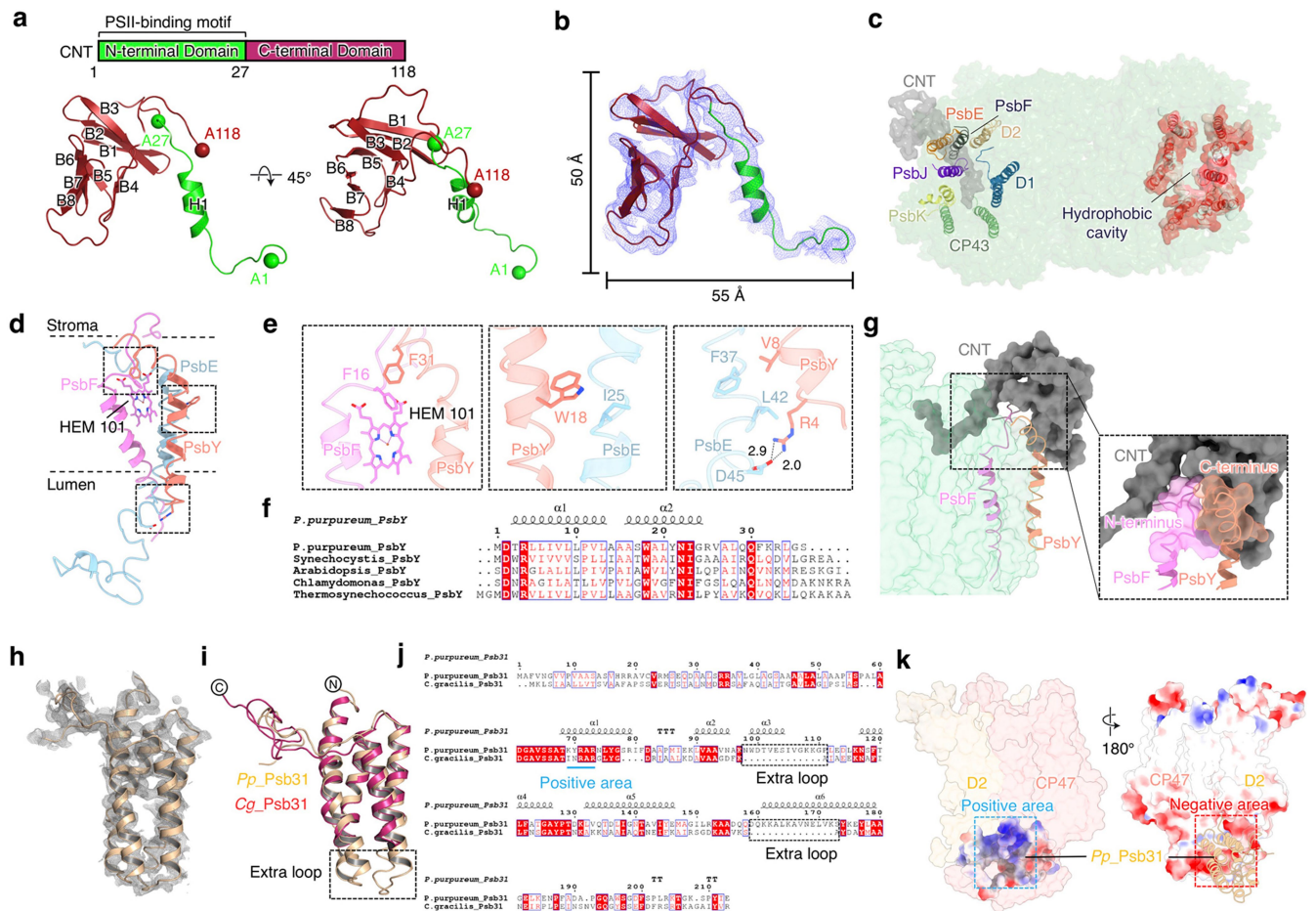
**Extended Data Fig. 2 | Single-particle analysis of each sub-region of PBS-PSII-PSI-LHC complex.** **a**, Density maps of single and double PBS-PSII-PSI-LHC megacomplexes that merged from final results, demonstrating the basic structural architectures. **b**, Structure-refinement procedure of

sub-region of PSII. **c**, Structure-refinement procedure of sub-region of PSI-LHC. **d**, Image processing of two sub-regions, a hexamer and a PSII dimer, at the center connecting position of double PBS-PSII-PSI-LHC megacomplex. **e-j**, Results of sub-regional refinements and the resolution estimations.



**Extended Data Fig. 3 | The distribution of single and double PBS-PSII-PSI-LHC megacomplexes in cell.** **a**, A tomogram (selected from 46) with repositioned single and double PBS-PSII-PSI-LHC megacomplexes showing the face view of PBS. Thylakoid membrane was represented as transparent density layers. **b**, A local area between two thylakoid membranes. Rotated 90°

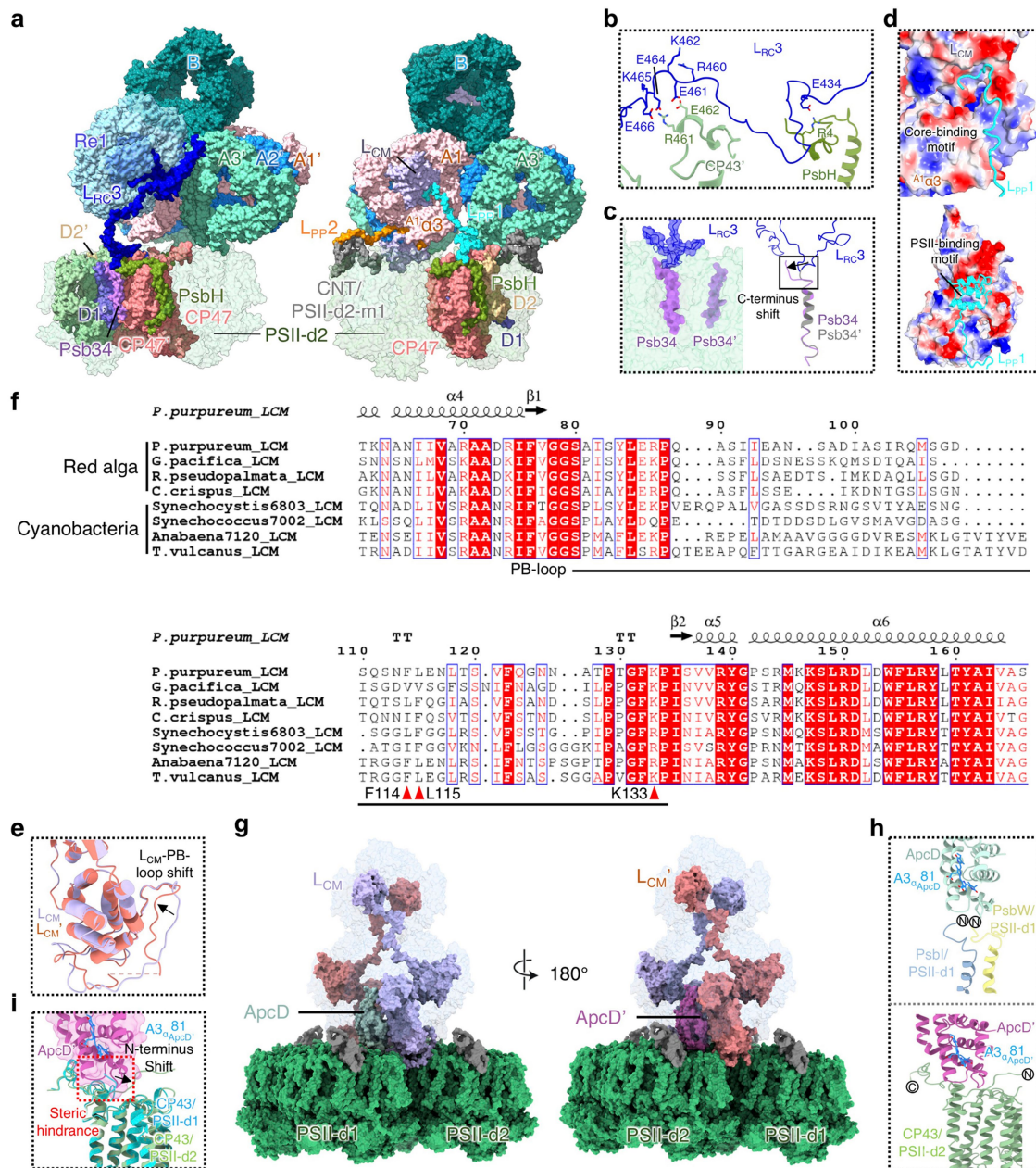
to present the PSII and PSI-LHCs organization in membrane. **c**, A tomogram (selected from 46) showing the side view of single and double PBS-PSII-PSI-LHC megacomplexes. **d**, Histogram of single and double PBS-PSII-PSI-LHC megacomplexes in 92 regions from 46 tomograms (each tomogram was separated into two regions).



**Extended Data Fig. 4 | Characterization of connector (CNT), PsbY and Psb31.**

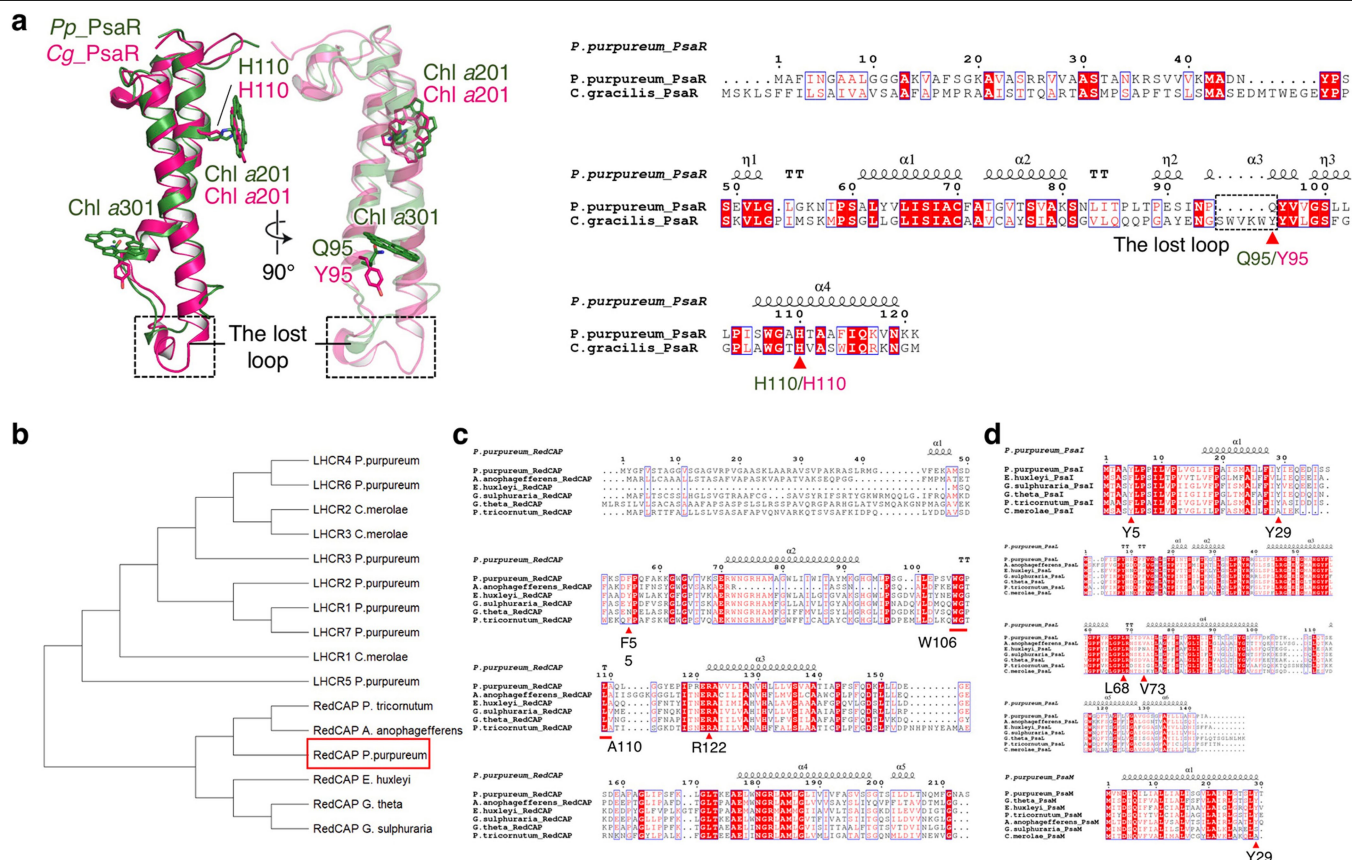
**a**, Structure of CNT is shown as cartoon representation. N- and C-terminal domains are colored in warm pink and green, respectively. **b**, Cryo-EM density (mesh) for the CNT superimposed with its atomic model (cartoon). **c**, Location of CNT shows that the N-terminal domain is inserted into a hydrophobic cavity created by transmembrane helices of D1, D2, CP43, PsbK, PsbJ, PsbE and PsbF. **d, e**, The interactions of PsbY with PsbE and PsbF. The amino acid residues involved in the interactions are shown as sticks in the enlarged boxes. **f**, Sequence alignment of PsbY from *P. purpureum* and other species.

**g**, The interactions between PsbY, PsbF and CNT. C-/N-terminus of PsbY and PsbF are shown as surface representation. **h**, Cryo-EM density (mesh) for the Psb31 superimposed with its atomic model (cartoon). **i**, Structural comparison between them is 1.34 Å. **j**, Sequence alignment of Psb31 from *P. purpureum* and *C. gracilis*. The positively charged area of Psb31 is marked with a blue line. The extra loop of Psb31 from *P. purpureum* is boxed in rectangles. **k**, Electrostatic interaction between Psb31 and PSII core. Positively and negatively charged areas are boxed in blue and red rectangles, respectively.



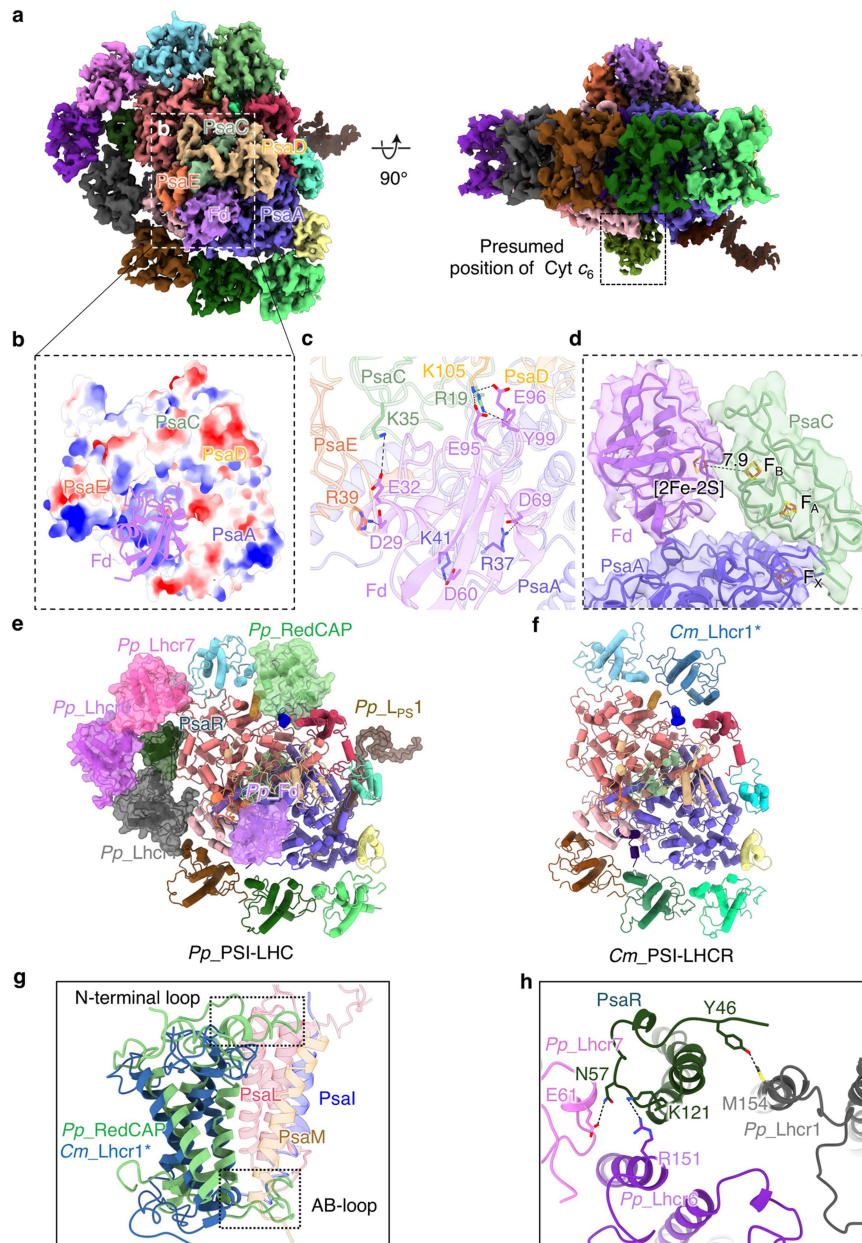
**Extended Data Fig. 5 | Interactions between PBS and PSII. a**, Overview of  $L_{RC3}$ ,  $L_{PP1}$  and  $L_{PP2}$  interacting with both PBS and PSII from two different views. **b**, The interactions of  $L_{RC3}$  with CP43' and PsbH. The amino acids involved in interaction are shown as sticks. **c**, A C-terminus shift of Psb34 occurs due to the interaction between  $L_{RC3}$  and Psb34. **d**, Electrostatic interaction of  $L_{PP1}$  with PBS core and PSII. **e**, Structural alignment of  $L_{CM}$  and  $L_{CM}'$ . The arrow indicates a shift of  $L_{CM}$ -PB-loop. **f**, Sequence alignment of PB-loop/ $L_{CM}$  from different red algae and cyanobacteria. PB-loop of  $L_{CM}$  is highlighted by a solid line. Key residues participating in interactions are marked as red triangle. Used species are red algae: *Porphyridium purpureum*, *Griffithsia pacifica*, *Rhodomenia*

*pseudopalmata* and *Chondrus crispus* and cyanobacteria: *Synechocystis* sp. PCC 6803, *Synechococcus* sp. strain PCC 7002, *Anabaena* sp. PCC 7120 and *Thermosynechococcus vulcanus* NIES-2134. **g**, Overview of  $L_{CM}$ / $L_{CM}'$  and ApcD/ApcD' interacting with PSII. **h**, Structural alignment of  $L_{CM}$  and  $L_{CM}'$ . The arrow indicates a shift of  $L_{CM}$ -PB-loop. **i**, Interaction of ApcD (left), ApcD' (right) with PSII-d1 and PSII-d2. **c**, Structural alignment of CP43 from PSII-d1 and PSII-d2. Steric hindrance created by ApcD' is shown as surface representation in red box. The arrow indicates the N-terminus shift of CP43/PSII-d2 due to the steric hindrance.



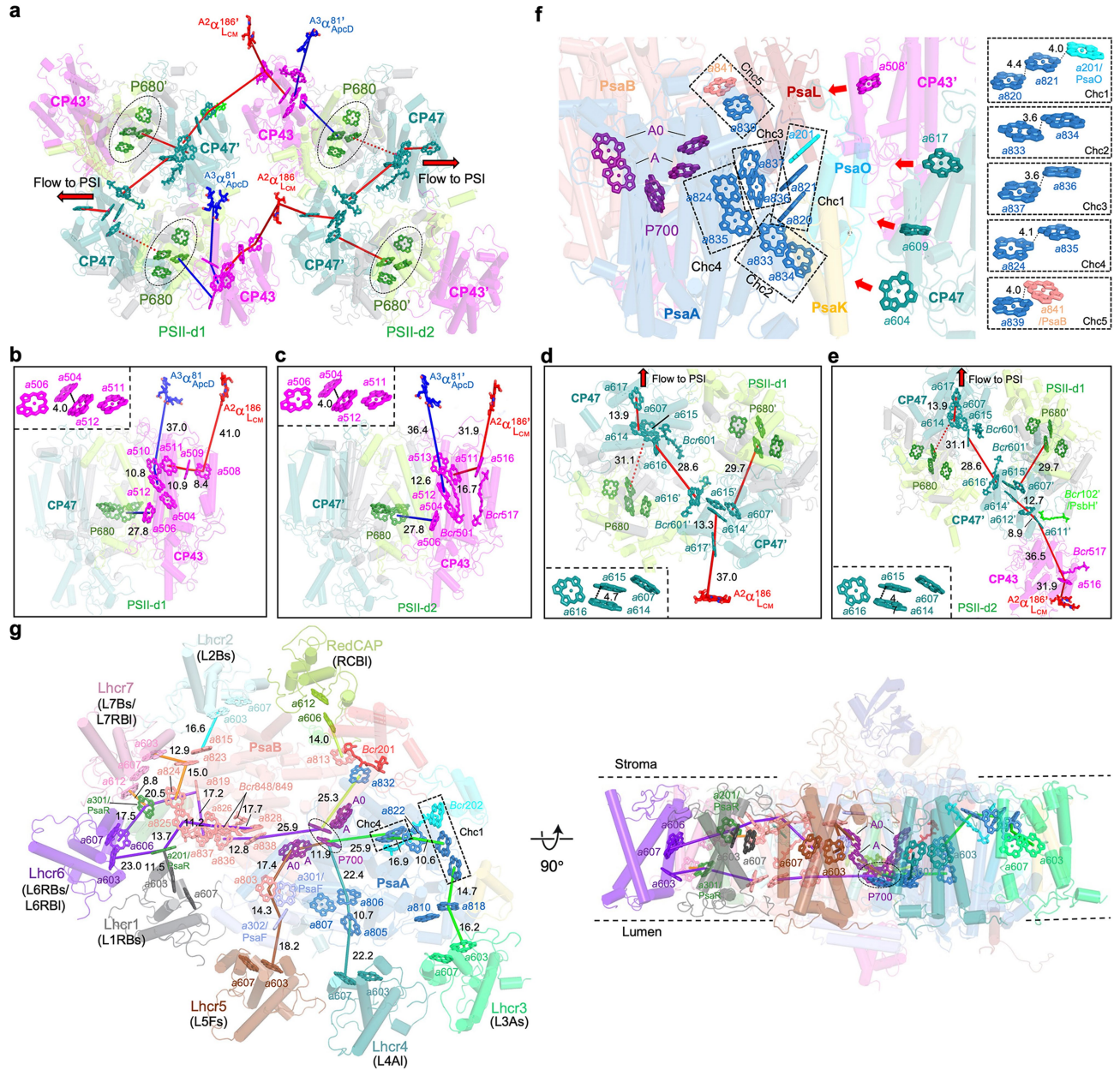
**Extended Data Fig. 6 | Structural and sequence analysis of PSI-LHC subunits. a**, Structural and sequence alignment of PsaR from *P. purpureum* and *Chaetoceros gracilis*. Left, structural alignment of PsaR from *P. purpureum* and *C. gracilis*. The Chls *a* and key residues are shown as sticks. The lost loop of *Pp\_PsaR* is boxed in rectangle. Right, sequence alignment of PsaR from *P. purpureum* and *C. gracilis*. Key residues participating in Chls *a* binding are marked as red triangle. **b**, Phylogenetic analysis of LHCR and RedCAP from red algal lineage

species, RedCAP from *P. purpureum* is boxed in red. **c-d**, Sequence alignments of RedCAP (c), PsaI, PsaL and PsaM (d) from *P. purpureum* and other algae with secondary plastids of red algal origin. Key residues participating in interactions are marked as red triangle or red line. Used species are *Porphyridium purpureum*, *Cyanidioschyzon merolae*, *Phaeodactylum tricornutum*, *Aureococcus anophagefferens*, *Emiliania huxleyi*, *Guillardia theta* and *Galdieria sulphuraria*.



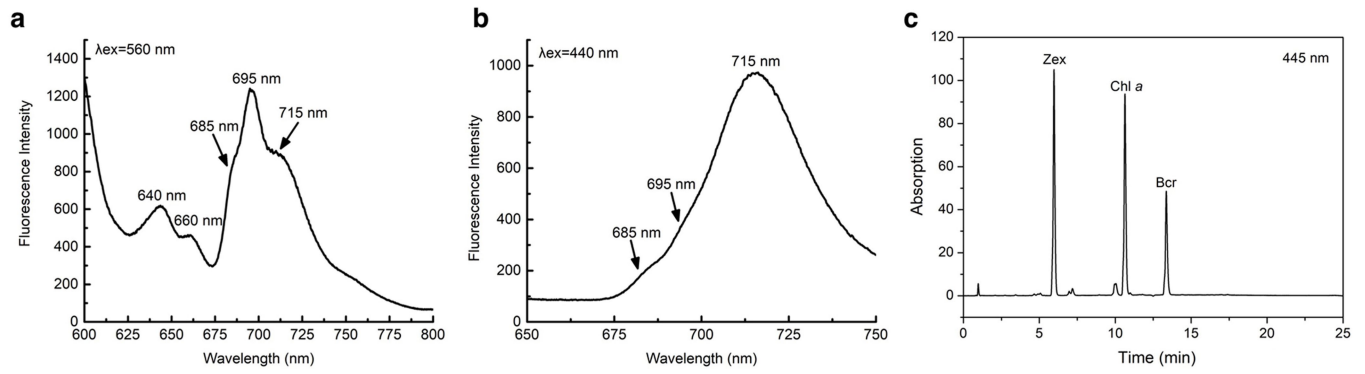
**Extended Data Fig. 7 | Structural analysis of PSI-LHC.** **a**, The cryo-EM density map of PSI-LHC is viewed from the stromal side (left) and along the membrane plane (right) at a threshold level of 0.203. All subunits are color-coded. The density map of the boxed area is presumed to be Cyt  $c_6$  according to the cryo-EM structure of Fd-PSI- $c_6$  complex from *Thermosynechococcus elongatus* BP-1. **b**, Enlarged view of the boxed area of **a** shows the electrostatic interaction of Fd with PsaA, PsaC, PsaD and PsaE. **c**, Key sites of Fd-PSI interactions. Six negatively charged residues D29/Fd, E32/Fd, D60/Fd, D69/Fd, E95/Fd and E96/Fd are salt-bridged with positively charged residues R39/PsaE, K35/PsaC, R37/PsaA, K41/PsaA and K105/PsaD, respectively. Furthermore, an additional

cation- $\pi$  interaction is formed by an aromatic residue Y99/Fd with R19/PsaC. The residues involved in the interactions are shown as sticks. **d**, The edge-to-edge distance (Å) between [2Fe-2S] cluster of Fd and [4Fe-4S] cluster of  $F_B$  in PsaC is measured (solid lines). **e**, **f**, The structures of *P. purpureum* PSI-LHC (**e**) and *C. merolae* PSI-LHCR (PDB entry 5ZGB) (**f**) are shown as cartoon representation. Compared to *C. merolae* PSI-LHCR, the extra parts of *P. purpureum* PSI-LHC are highlighted as surface representation in **e**. **g**, Structural alignment of RedCAP and Lhcr1\*. N-terminal loop and AB-loop of RedCAP are boxed in rectangles. **h**, Detailed interactions of PsbR with Lhcr1, Lhcr6 and Lhcr7.



**Extended Data Fig. 8 | Possible energy transfer pathways from PBS to PSII/PSI and in PSI-LHC.** **a**, Overall energy transfer pathways from two pairs of energy terminal emitters ( $A^3\alpha_{ApcD}$ ,  $A^3\alpha_{ApcD'}$ ,  $A^2\alpha_{LCM}$  and  $A^2\alpha_{LCM'}$ ) to reaction center of PSII (P680). **b-e**, Enlarged views of **a** show the energy transfer details from the terminal emitters to P680. The red arrows in **d** and **e** indicate that the energy further flow to PSI. **f**, The distribution and composition of five low energy state Chl clusters of PSI core. Chl clusters Chl1 to Chl5 are boxed in rectangles. The red arrows indicate the direction of the energy transfer. **g**, Possible energy transfer pathways of the PSI-LHC supercomplex viewed normal to the membrane plane from the stromal side (left) and along the membrane plane (right), respectively. Key Chl  $\alpha$  and Bcr are shown as bold-stick and the  $\pi$ - $\pi$  distances ( $\text{\AA}$ ) for the adjacent pigments are labelled in black. P700 and the low energy state Chl pairs are boxed in oval and rectangles, respectively. Possible energy transfer pathways are designated as L1RBs, L2Bs, L3As, L4Al, L5Fs, L6RBs/L6RBI, L7Bs/L7RBI and RCBI (based on Lhcr/RedCAP number, PSI core subunit, and stromal or luminal side) and described as below. L1RBs: Chl dimer  $a603/607/Lhcr1$  -  $a201/PsaR$  (11.5  $\text{\AA}$ ) - Chl dimer  $a824/825/PsaB$  (13.7  $\text{\AA}$ ) -  $a826/PsaB$  (11.2  $\text{\AA}$ ) - Chl dimer  $a828/838/PsaB$  (17.7  $\text{\AA}$ ) - P700 (25.9  $\text{\AA}$ ); L2Bs:

Chl dimer  $a603/607/Lhcr2$  - Chl dimer  $a815/823/PsaB$  (16.6  $\text{\AA}$ ) - Chl dimer  $a824/825/PsaB$  (15.0  $\text{\AA}$ ) -  $a826/PsaB$  (11.2  $\text{\AA}$ ) - Chl dimer  $a828/838/PsaB$  (17.7  $\text{\AA}$ ) - P700 (25.9  $\text{\AA}$ ); L3As: Chl dimer  $a603/607/Lhcr3$  - Chl dimer  $a810/818/PsaA$  (16.2  $\text{\AA}$ ) - Chl1 (14.7  $\text{\AA}$ ) -  $a822/PsaA$  (10.6  $\text{\AA}$ ) - Chl4 (16.9  $\text{\AA}$ ) - P700 (25.9  $\text{\AA}$ ); L4Al: Chl dimer  $a603/607/Lhcr4$  -  $a805/PsaA$  (22.2  $\text{\AA}$ ) - Chl dimer  $a806/807/PsaA$  (10.7  $\text{\AA}$ ) - P700 (22.4  $\text{\AA}$ ); L5Fs: Chl dimer  $a603/607/Lhcr5$  -  $a302/PsaF$  (18.2  $\text{\AA}$ ) - Chl dimer  $a803/301/PsaB$  -  $PsaF$  (14.3  $\text{\AA}$ ) - A0/A (11.9  $\text{\AA}$ ) - P700 (25.9  $\text{\AA}$ ); L6RBs: Chl dimer  $a603/607/Lhcr6$  -  $a201/PsaR$  (23.0  $\text{\AA}$ ) - Chl dimer  $a824/825/PsaB$  (13.7  $\text{\AA}$ ) -  $a826/PsaB$  (11.2  $\text{\AA}$ ) - Chl dimer  $a828/838/PsaB$  (17.7  $\text{\AA}$ ) - P700 (25.9  $\text{\AA}$ ); L6RBI:  $a606/Lhcr6$  -  $a301/PsaR$  (17.5  $\text{\AA}$ ) -  $a819/PsaB$  (20.5  $\text{\AA}$ ) - Chl dimer  $a836/837/PsaB$  (17.2  $\text{\AA}$ ) - Chl dimer  $a828/838/PsaB$  (12.8  $\text{\AA}$ ) - P700 (25.9  $\text{\AA}$ ); L7Bs: Chl dimer  $a603/607/Lhcr7$  - Chl dimer  $a815/823/PsaB$  (12.9  $\text{\AA}$ ) - Chl dimer  $a824/825/PsaB$  (15.0  $\text{\AA}$ ) -  $a826/PsaB$  (11.2  $\text{\AA}$ ) - Chl dimer  $a828/838/PsaB$  (17.7  $\text{\AA}$ ) - P700 (25.9  $\text{\AA}$ ); L7RBI:  $a612/Lhcr7$  -  $a301/PsaR$  (8.8  $\text{\AA}$ ) -  $a819/PsaB$  (20.5  $\text{\AA}$ ) - Chl dimer  $a836/837/PsaB$  (17.2  $\text{\AA}$ ) - Chl dimer  $a828/838/PsaB$  (12.8  $\text{\AA}$ ) - P700 (25.9  $\text{\AA}$ ); RCBI: Chl dimer  $a606/612/RedCAP$  - Chl dimer  $a813/832/PsaB$  -  $PsaA$  (14.0  $\text{\AA}$ ) - P700 (25.3  $\text{\AA}$ ).



**Extended Data Fig. 9 | 77 K fluorescence emission spectra of *P. purpureum* and pigment analysis of the thylakoid membrane from *P. purpureum*.**  
**a, b,** Low temperature (77 K) fluorescence emission spectra of *P. purpureum* cell

locked in solid line. The excitation wavelength was 560 nm (**a**) and 440 nm (**b**).  
**c,** Pigment analysis by HPLC. The elutes were recorded at 445 nm. Zex, zeaxanthin; Chl *a*, Chlorophyll *a*; Bcr,  $\beta$ -carotene.

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistic

|                                        | Single<br>PBS-PSII-<br>PSI-LHC<br>(EMDB-<br>33605)<br>(PDB<br>7Y5E) | Double<br>PBS-PSII-<br>PSI-LHC<br>(EMDB-<br>33669)<br>(PDB<br>7Y7A) | PBS<br>(EMDB-<br>33605)<br>(PDB<br>7Y4L) | PSII-d1-d2<br>(EMDB-<br>33597) | PSII-d3<br>(EMDB-<br>33568) | PSI-LHC<br>(EMDB-<br>33561) | Lateral<br>hexamer<br>(EMDB-<br>33558)<br>(PDB<br>7Y1A) |
|----------------------------------------|---------------------------------------------------------------------|---------------------------------------------------------------------|------------------------------------------|--------------------------------|-----------------------------|-----------------------------|---------------------------------------------------------|
| <b>Data collection and processing</b>  |                                                                     |                                                                     |                                          |                                |                             |                             |                                                         |
| Magnification                          | 53,000                                                              | 53,000                                                              | 53,000                                   | 53,000                         | 53,000                      | 53,000                      | 53,000                                                  |
| Voltage (kV)                           | 300                                                                 | 300                                                                 | 300                                      | 300                            | 300                         | 300                         | 300                                                     |
| Electron exposure (e-/Å <sup>2</sup> ) | 35                                                                  | 35                                                                  | 35                                       | 35                             | 35                          | 35                          | 35                                                      |
| Defocus range (μm)                     | -1.0 ~ -6.0                                                         | -1.0 ~ -6.0                                                         | -1.0 ~ -6.0                              | -1.0 ~ -6.0                    | -1.0 ~ -6.0                 | -1.0 ~ -6.0                 | -1.0 ~ -6.0                                             |
| Pixel size (Å)                         | 1.632                                                               | 1.632                                                               | 1.632                                    | 1.088                          | 1.632                       | 1.632                       | 1.632                                                   |
| Symmetry imposed                       | C1                                                                  | C2                                                                  | C1                                       | C1                             | C2                          | C1                          | C1                                                      |
| Initial particle image (no.)           | 762,000                                                             | 762,000                                                             | 762,000                                  | 762,000                        | 762,000                     | 762,000                     | 762,000                                                 |
| Final particle image (no.)             | 215,000                                                             | 87,000                                                              | 215,000                                  | 321,000                        | 87,000                      | 112,000                     | 87,000                                                  |
| Map resolution (Å)                     | 3.3                                                                 | 4.3                                                                 | 3.3                                      | 3.2                            | 3.4                         | 3.6                         | 6.3                                                     |
| FSC threshold                          | 0.143                                                               | 0.143                                                               | 0.143                                    | 0.143                          | 0.143                       | 0.143                       | 0.143                                                   |
| Map resolution range (Å)               | 3.3 ~ 11                                                            | 4.3 ~ 15                                                            | 3.3 ~ 11                                 | 3.2 ~ 5                        | 3.4 ~ 5                     | 3.5 ~ 7                     | 6.3 ~ 8                                                 |
| <b>Refinement</b>                      |                                                                     |                                                                     |                                          |                                |                             |                             |                                                         |
| Initial model used (PDB code)          | 6KGX, 4YUU, 5ZGB                                                    | 6KGX, 4YUU, 5ZGB                                                    | 6KGX                                     |                                |                             |                             | 6KGX                                                    |
| Model resolution (Å)                   | 3.3                                                                 | 4.3                                                                 | 3.3                                      |                                |                             |                             | 6.3                                                     |
| FSC threshold                          | 0.143                                                               | 0.143                                                               | 0.143                                    |                                |                             |                             | 0.143                                                   |
| Model resolution range (Å)             | 3.3 ~ 11                                                            | 4.3 ~ 15                                                            | 3.3 ~ 11                                 |                                |                             |                             | 6.3 ~ 8                                                 |
| Map sharpen                            | -65                                                                 | -55                                                                 | -65                                      |                                |                             |                             | -600                                                    |
| <i>B</i> factor (Å <sup>2</sup> )      |                                                                     |                                                                     |                                          |                                |                             |                             |                                                         |
| Model composition                      |                                                                     |                                                                     |                                          |                                |                             |                             |                                                         |
| Non-hydrogen atoms                     | 1,264,564                                                           | 2,499,152                                                           | 1,032,325                                |                                |                             |                             | 17,788                                                  |
| Protein residues                       | 152,419                                                             | 305,004                                                             | 127,924                                  |                                |                             |                             | 2,283                                                   |
| Ligands                                | 2,437                                                               | 4,593                                                               | 1602                                     |                                |                             |                             | 31                                                      |
| <i>B</i> factors (Å <sup>2</sup> )     |                                                                     |                                                                     |                                          |                                |                             |                             |                                                         |
| Protein                                | 89.04                                                               | 88.38                                                               | 88.26                                    |                                |                             |                             | 149.69                                                  |
| Ligands                                | 97.16                                                               | 95.97                                                               | 101.7                                    |                                |                             |                             | 180.89                                                  |
| R.m.s.deviation                        |                                                                     |                                                                     |                                          |                                |                             |                             |                                                         |
| Bond lengths (Å)                       | 0.012                                                               | 0.011                                                               | 0.012                                    |                                |                             |                             | 0.012                                                   |
| Bond angles (°)                        | 2.093                                                               | 2.063                                                               | 2.163                                    |                                |                             |                             | 2.506                                                   |
| Validation                             |                                                                     |                                                                     |                                          |                                |                             |                             |                                                         |
| MolProbity score                       | 2.2                                                                 | 2.16                                                                | 2.24                                     |                                |                             |                             | 3.02                                                    |
| Clashscore                             | 12.43                                                               | 11.39                                                               | 13.15                                    |                                |                             |                             | 9.27                                                    |
| Poor rotamers (%)                      | 1.78                                                                | 1.72                                                                | 2.05                                     |                                |                             |                             | 18.4                                                    |
| Ramachandran plot                      |                                                                     |                                                                     |                                          |                                |                             |                             |                                                         |
| Favored (%)                            | 93.99                                                               | 94.02                                                               | 94.61                                    |                                |                             |                             | 89.45                                                   |
| Allowed (%)                            | 5.65                                                                | 5.68                                                                | 5.14                                     |                                |                             |                             | 10.51                                                   |
| Disallowed (%)                         | 0.37                                                                | 0.30                                                                | 0.26                                     |                                |                             |                             | 0.04                                                    |

Corresponding author(s): Sen-Fang Sui

Last updated by author(s): Dec 28, 2022

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☒ ☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection serialEM 3.7.13

Data analysis CTFIND4, MotionCor2 1.1.0, IMOD 4.10.38, emClarity 1.5.3, Amira 2020.2, isSPA, RELION 3.0, RELION 4.0, cryoSparc 3.3.1, Phenix 1.14-3260, Coot 0.8.9.1, MATLAB2021a, Pymol 1.8.2.1, Chimera 1.12, ChimeraX 1.3, AlphaFold2, CLUSTAL 2.0, ESPrpt 3.0, MEGA11.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The cryo-EM density map and atomic models generated in this study have been deposited in the Electron Microscopy Data Bank and the Protein Data Bank for the single-PBS-PSII-PSI-LHCs megacomplex structure at 3.3 Å resolution (EMDB ID code 33605 and PDB ID code 7Y5E), the double-PBS-PSII-PSI-LHCs megacomplex structure at 4.3 Å resolution (EMDB ID code 33669 and PDB ID code 7Y7A), the PBS structure at 3.3 Å resolution (EMDB ID code 33605 and PDB ID code 7Y4L), the PSII-d1-d2 structure at 3.2 Å resolution (EMDB ID code 33597), the PSII-d3 structure at 3.4 Å resolution (EMDB ID code 33568), the PSI-LHCs structure at 3.6 Å resolution (EMDB ID code 33561) and the lateral hexamer structure at 6.3 Å resolution (EMDB ID code 33558 and PDB ID code 7Y1A). Two whole artificially stitched

maps have been deposited in the Electron Microscopy Data Bank (EMDB ID code 33618 for single-PBS-PSII-PSI-LHCs and 33658 for double-PBS-PSII-PSI-LHCs). For the publicly available atomic models used in this study, their accession codes in the Protein Data Bank have been provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

|                 |                                                                                                                                                                                                                                                                                                                                                                             |
|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sample size     | Amount of cryo-ET and single-particle micrographs collected were based on the previous knowledge that the reconstruction of the protein particles picked from these micrographs could reach to a near atomic resolution and also limited by the time allocation of the microscope.                                                                                          |
| Data exclusions | The exclusion criteria were not pre-established. 2D and 3D classification yielded multiple classes. Only the particles in the classes that showed clear structural signals and intact structures were selected, combined and used in the final reconstruction and refinement. Details are described in the flowchart of Extended Data Figure 1 and 2, and Methods.          |
| Replication     | The refinements were repeated at least two times with different angle searching range and all resulted in similar density maps (with different resolutions though). Experiments concerning the thylakoid membrane separation and HPLC analysis have been repeated for at least three times with similar results. Two batches of the thylakoid membrane were analyzed by MS. |
| Randomization   | Gold standard Fourier Shell Correlation method was used to estimate the resolution of the cryo-EM structures, in which the dataset is split into two sets and refined independently. The process of splitting the dataset, odd and even, is considered random.                                                                                                              |
| Blinding        | Blinding is not applicable to a cryo-EM experiment, because no human research participants are involved.                                                                                                                                                                                                                                                                    |

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a                                 | Involved in the study                                  |
|-------------------------------------|--------------------------------------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                    |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern  |

### Methods

| n/a                                 | Involved in the study                           |
|-------------------------------------|-------------------------------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |