

Structure of a green algal photosystem I in complex with a large number of light-harvesting complex I subunits

Xiaochun Qin^{1,2,5}, Xiong Pi^{3,5}, Wenda Wang^{1,2}, Guangye Han², Lixia Zhu¹, Mingmei Liu¹, Linpeng Cheng³, Jian-Ren Shen^{1,2,4}, Tingyun Kuang^{1,2*} and Sen-Fang Sui^{1,3*}

Photosystem I (PSI) is a highly efficient natural light-energy converter, and has diverse light-harvesting antennas associated with its core in different photosynthetic organisms. In green algae, an extremely large light-harvesting complex I (LHCI) captures and transfers energy to the PSI core. Here, we report the structure of PSI-LHCI from a green alga *Bryopsis corticulans* at 3.49 Å resolution, obtained by single-particle cryo-electron microscopy, which revealed 13 core subunits including subunits characteristic of both prokaryotes and eukaryotes, and 10 light-harvesting complex a (Lhca) antennas that form a double semi-ring and an additional Lhca dimer, including a novel four-transmembrane-helix Lhca. In total, 244 chlorophylls were identified, some of which were located at key positions for the fast energy transfer. These results provide a firm structural basis for unravelling the mechanisms of light-energy harvesting, transfer and quenching in the green algal PSI-LHCI, and important clues as to how PSI-LHCI has changed during evolution.

Photosynthesis converts light energy from the sun into the chemical energy that is indispensable for almost all life forms on the Earth. To compensate for the low density of light energy available on the Earth's surface, photosynthetic organisms have developed various light-harvesting antennas to harvest and transfer light energy to the core complex of two photosystems, photosystem I (PSI) and photosystem II (PSII), in which charge separation and electron transfer take place. PSI is responsible for light-driven charge separation and electron transfer from plastocyanin or cytochrome c_6 on the lumenal side of the membrane to ferredoxin on the stromal side. The core of PSI is associated with various types of peripheral antennas that are designated as membrane-intrinsic light-harvesting complexes (Lhcs)—encoded by the *lh* gene superfamily—with the exception of membrane-extrinsic soluble phycobilisome (PBS), which is found in cyanobacteria and some red algae^{1–3}. In green algae, mosses and higher plants within the green lineage of photosynthetic organisms, Lhcs are categorized into Lhca and Lhcb, which serve as the antennas for PSI and PSII, respectively; whereas in red algae, one of the most primitive photosynthetic eukaryotic algae, Lhcr subunits function as the light-harvesting antenna for PSI^{3,4}. During evolution from the primitive red algae to the green algae and the terrestrial plants, the number of antennas coupled with the PSI core has changed from 3–5 (ref. ⁴) to 9–14 (refs. ^{5–8}) and then to 4 (refs. ^{9–11}). Interestingly, PSI-LHCI of the green algae—with its larger antenna cross-section—has a similar or even shorter average transfer and decay time for excitation energy compared with the higher plant PSI-LHCI^{12,13}, suggesting that there is a higher efficiency for energy transfer in the green algal PSI-LHCI, contrary to the general expectation that a larger antenna size would mean slower excitation trapping. The architecture of the green algal PSI-LHCI and the detailed mechanism of energy transfer within it,

however, are unclear due to the lack of high-resolution structures. How the structure of this extremely large antenna system affects the excitation-energy transfer (EET) is unclear, and is a fundamental question for understanding photosynthesis. Thus, a detailed structure of the green algal PSI-LHCI is needed, not only to clarify LHCI assembly, but also to reveal the possible EET pathways and to fully understand the structure and function of PSI.

Bryopsis corticulans is a macroscopic, siphonaceous marine green alga living in intertidal zones, with a life cycle adapted to a complex light environment that changes frequently between strong light and low underwater light¹⁴. We purified the PSI-LHCI supercomplex from *B. corticulans*, determined the complementary DNA (cDNA) gene sequences of the PSI core and Lhca subunits, and solved its structure to an overall resolution of 3.49 Å using single-particle cryo-electron microscopy (cryo-EM). This structure, together with our previous biochemical and spectroscopic studies on *B. corticulans* LHCII and PSI-LHCI^{15,16}, sheds light on the mechanisms of light-energy harvesting, transfer and quenching in the green algal PSI-LHCI, and provides important clues to the evolutionary changes that have occurred in PSI-LHCI from red algae to higher plants.

Properties of the isolated PSI-LHCI supercomplex

We isolated the PSI-LHCI supercomplex from *B. corticulans* by density-gradient ultracentrifugation (see Methods and Supplementary Fig. 1a), and compared its polypeptide composition with those of the PSI-LHCI supercomplexes isolated from a well-studied, model green alga *Chlamydomonas reinhardtii* and a higher plant pea *Pisum sativum* (Supplementary Fig. 1b). The result showed that the three samples have different subunit compositions and/or molecular weights of individual subunits in the region of LHCI (15–25 kDa). Since the genomic sequence of *B. corticulans* has not been determined,

¹School of Biological Science and Technology, University of Jinan, Jinan, China. ²Photosynthesis Research Center, Key Laboratory of Photobiology, Institute of Botany, Chinese Academy of Sciences, Beijing, China. ³State Key Laboratory of Membrane Biology, Beijing Advanced Innovation Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing, China. ⁴Institute for Interdisciplinary Science, Okayama University, Okayama, Japan.

⁵These authors contributed equally: Xiaochun Qin, Xiong Pi. *e-mail: kuangty@ibcas.ac.cn; suif@mail.tsinghua.edu.cn

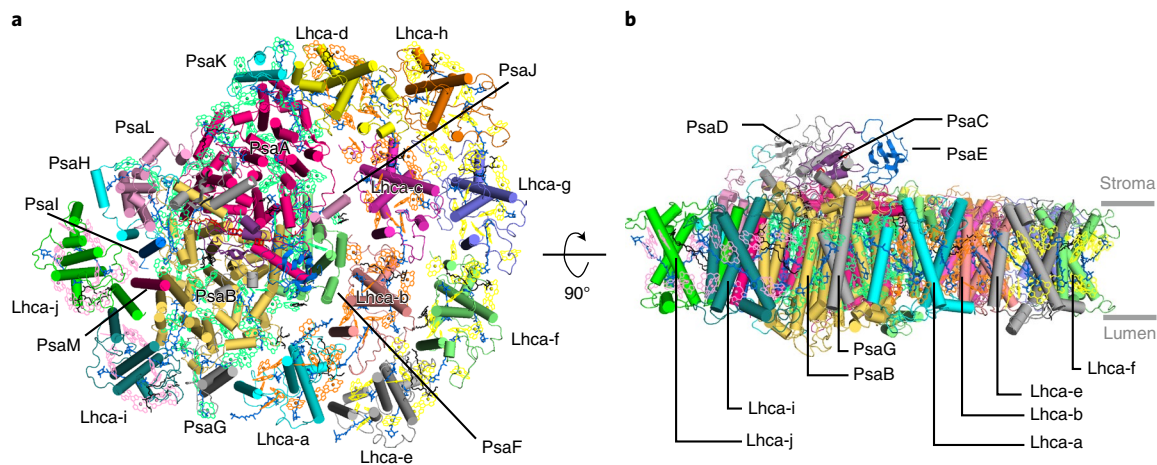


Fig. 1 | Overall structure of the PSI-LHCI supercomplex from *B. corticulans*. **a**, View from the stromal side. **b**, Side view along the membrane plane, with the phytol chains of Chls omitted. Colour codes, PSI core subunits: bright pink, PsaA and PsaM; yellow-orange, PsaB; violet-purple, PsaC; grey, PsaD and PsaG; marine, PsaE and PsaI; lime, PsaF; cyan, PsaH and PsaK; pink, PsaJ and PsaL. Lhca proteins: cyan, Lhca-a; salmon, Lhca-b; magenta, Lhca-c; yellow, Lhca-d; grey, Lhca-e; lime, Lhca-f; blue, Lhca-g; orange, Lhca-h; deep teal, Lhca-i; green, Lhca-j. Cofactors: lime, Chls of the PSI core complex; orange, Chls of the inner ring Lhcas (Lhca-a to Lhca-d); yellow, Chls of the outer ring Lhcas (Lhca-e to Lhca-h); pink, Chls of Lhca-i and Lhca-j; blue, carotenoids; red, phytylquinones and Fe_4S_4 clusters.

we determined the cDNA sequences of 13 PSI core subunits (Supplementary Fig. 2) and 9 Lhcas (Supplementary Fig. 3).

Cryo-EM images of the isolated PSI-LHCI supercomplex were collected and the resulting particles were subjected to two-dimensional (2D) and three-dimensional (3D) classification (see Methods and Supplementary Figs. 4 and 5). Among the 59,525 particles selected from the 3D classification, 78% of the particles were intact PSI-LHCI supercomplexes containing 10 Lhca subunits, whereas 22% had a smaller size and lacked the density corresponding to the LHCI dimer bound at the PsaB/H/I/M side (Supplementary Fig. 4), which is similar to the case of PSI from a red alga *Cyanidioschyzon merolae*⁴. All 59,525 particles were used to solve the structure of the PSI-LHCI supercomplex at a resolution of 3.49 Å (Methods, Supplementary Fig. 5 and Supplementary Table 1). The following sections describe the structural features of the complete PSI-LHCI supercomplex and the functional implications.

Architecture of the PSI-LHCI supercomplex

The overall structure of the green algal PSI-LHCI resembles a fan shape, with a core containing ten transmembrane subunits (PsaA, PsaB, PsaF, PsaG, PsaH, PsaI, PsaJ, PsaK, PsaL and PsaM) and three extrinsic subunits (PsaC, PsaD and PsaE) on the stromal side (Fig. 1). PsaN, which is present in the higher plant PSI-LHCI, and PsaO, which is found in both red algal PSI-LHCR⁴ and higher plant PSI-LHCI¹⁷, were not present in the core of the purified complex, suggesting that these two subunits were lost during preparation owing to their loose association with the PSI core. A major LHCI belt containing two approximately parallel semi-rings of LHCI subunits was associated at the PsaG–PsaF–PsaJ–PsaK side, with each of the rings containing four different Lhcas, Lhca-a to Lhca-d in the inner ring, and four Lhcas, Lhca-e to Lhca-h, in the outer ring. In addition, a Lhca dimer was found to be associated with the PsaB–PsaH–PsaI–PsaM side, opposite to the side of the main Lhca semi-ring belt, the subunits of which were designated Lhca-i and Lhca-j (Fig. 1). This attachment site for the two additional Lhcas is similar to that identified recently in a red algal PSI-LHCI⁴. However, the green algal PSI-LHCI is remarkably different from both red algal and higher plant PSI-LHCIs in that it has a much larger number of LHCI subunits attached, making its antenna cross-section much larger than that of the other species.

Structural features of the green algal PSI core

Although the PSI core is conserved during evolution¹⁸, the green algal PSI core (Protein Data Bank (PDB) ID code 6IGZ) shows some unique features in terms of its subunit composition and local structures of some subunits compared with other organisms (Supplementary Fig. 6). Ten core subunits (PsaA to PsaF and PsaI to PsaL) are found in all reported PSI structures from cyanobacteria to higher plants, whereas PsaM is present in cyanobacteria, red algae and green algae but not in higher plants, and PsaG and PsaH did not appear until the green algae evolved^{4,9–11,17–19}. PsaG provides an anchor site for Lhca-a (Lhca1) in both green algae and higher plants^{9–11}, whereas it is absent in the red algal PSI, which has a three-Lhcr-subunit belt⁴, reinforcing the important role of PsaG in maintaining the four-Lhca-subunit LHCI belt in the green algal lineage. In accordance with this, the three-Lhcr subunit in the red algal PSI is shifted slightly towards the PsaK side compared with the four-Lhca subunit in the higher plant PSI⁴. Furthermore, the stromal loop of PsaL is much longer in the green algae and higher plants than in cyanobacteria and red algae. This long loop, together with the N-terminal loop of PsaL, clamps PsaH (Supplementary Fig. 6a–c), providing a binding site for LHCI in a process called state transition²⁰, as revealed by the cryo-EM structure of PSI-LHCI-LHCI from maize¹⁷. The stromal loop of PsaK is also longer in green algae and higher plants than in red algae (here and refs. 4,17). Leu26 and Asp51 in this loop are coordinated to chlorophyll (Chl) a4006 and Chl a4005, respectively (Supplementary Fig. 6d); Chl a4006 plays an important role in EET from LHCI to the PSI core upon state transition¹⁷. Thus, the highly conserved structures of PsaL, PsaH and PsaK in both green algae and higher plants suggest a common basis for state transition, that is, LHCI shuttling between PSI and PSII²⁰, whereas the different structures between the red and green algal lineages reflect the absence of a transmembrane LHCI in primitive eukaryotic red algae, which instead use the extrinsic phycobilisomes as their antenna, and thus undergo the state transition by a different mechanism²¹. Another related difference is found in the structures of PsaF, all of which except that from cyanobacteria bind a Chl molecule (Chl a308) (Supplementary Fig. 6f) to funnel energy from LHCI to the PSI core in eukaryotes^{4,9–11,19}. In summary, the composition and structures of the PSI core subunits have changed in response to the evolution of their respective light-harvesting antennas.

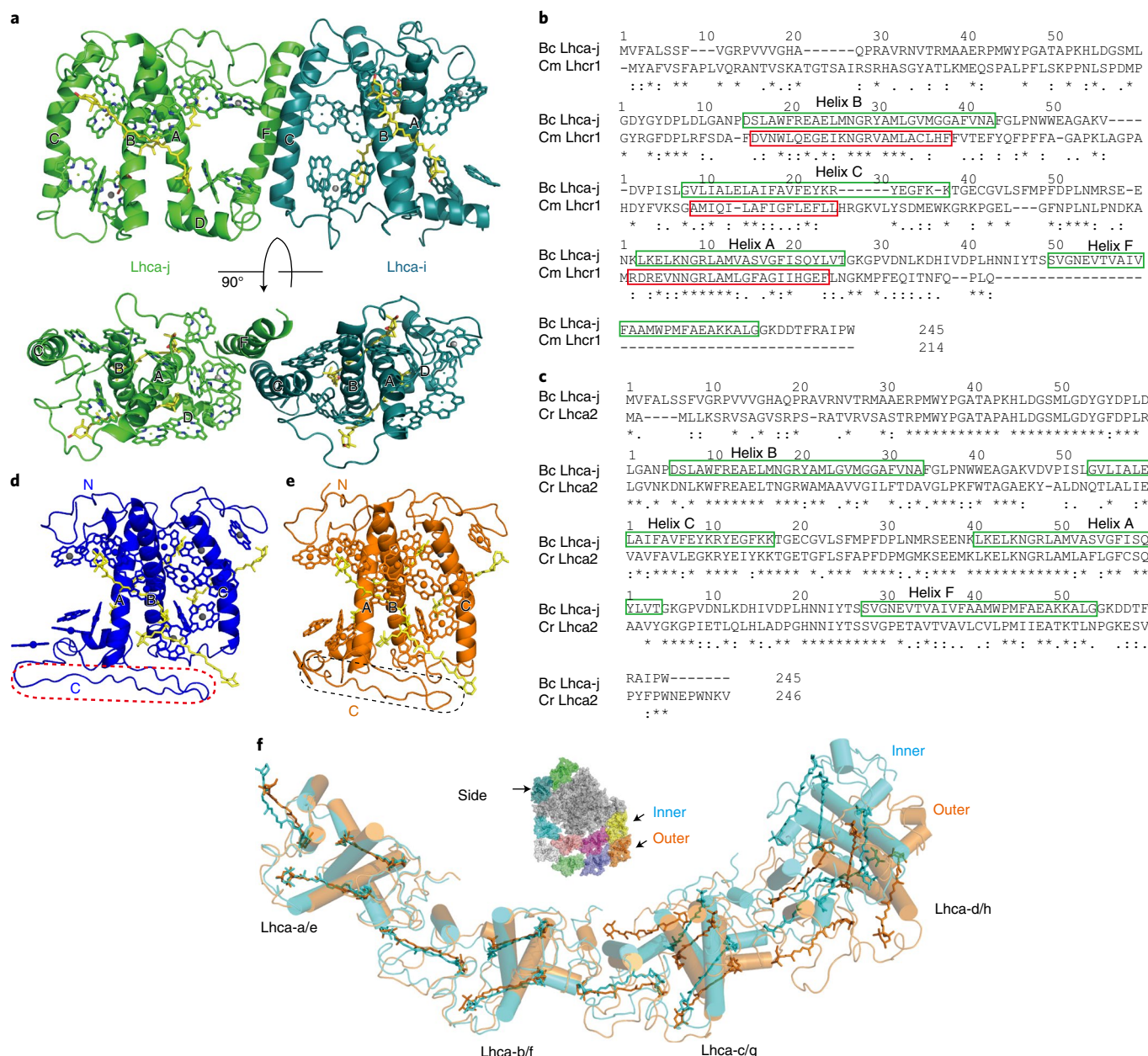


Fig. 2 | Structures of Lhcas. **a**, Structure of the Lhca-i-Lhca-j dimer. **b,c**, Sequence of Lhca-j of *B. corticulans* (Bc) and sequence alignments with Lhcr1 of *C. merolae* (Cm) (**b**) and with Lhca2 of *C. reinhardtii* (Cr) (**c**). **d,e**, Structures of Lhca-g (**d**) and Lhca-h (**e**). The long C-terminal loops are circled by red (**d**) and black (**e**) dashed lines. **f**, Superposition of the apoproteins and carotenoids between the inner LHCI ring (Lhca-a to Lhca-d, cyan) and the outer LHCI ring (Lhca-e to Lhca-h, orange).

Assembly of the large LHCI complex

Among the ten Lhca subunits found in the *B. corticulans* PSI-LHCI, the overall folding of nine subunits is similar to that of LHCII and LHCI^{9–11,22}, with three transmembrane helices (A, B and C) and one or two short amphipathic helices (D and E) on the luminal side (Supplementary Figs. 7 and 8), in agreement with their sequence similarities (Supplementary Fig. 9). The remaining subunit, Lhca-j, has an additional transmembrane helix (F) at its C-terminus (Supplementary Fig. 8a), which interacts with transmembrane helix C of Lhca-i to form a Lhca-i-Lhca-j dimer bound at the opposite side of the main Lhca double-ring belt (Figs. 1 and 2a). This dimer interacts with the PSI core subunits PsaH, PsaI, PsaM and PsaB; PsaM contributes to this association via extensive interactions with helix F of Lhca-j (Fig. 3a–c). Although Lhca-j is in a similar location in the *B. corticulans* PSI-LHCI supercomplex to Lhcr1* in the red

alga *C. merolae*, these two Lhcs have different helix arrangements, with Lhcr1* lacking a fourth transmembrane helix, which is consistent with their sequence alignment (Fig. 2b). However, sequence alignment of Lhca-j with *C. reinhardtii* Lhca2 revealed a high similarity (Fig. 2c), suggesting that Lhca2 also contains a fourth helix. Psbs is also a member of the four-transmembrane-helix LHC family; however, it does not bind pigments²³ and its transmembrane-helix arrangement is distinctly different from that of Lhca-j (Supplementary Fig. 10). Lhca-j is unusual among pigment-binding Lhc proteins in having four transmembrane helices, and may therefore represent the ancient form of the three-transmembrane-helix Lhcs commonly found in the antennae, since the three-transmembrane-helix Lhcs are suggested to have evolved from a four-transmembrane-helix ancestor through loss of the C-terminal transmembrane helix²⁴.

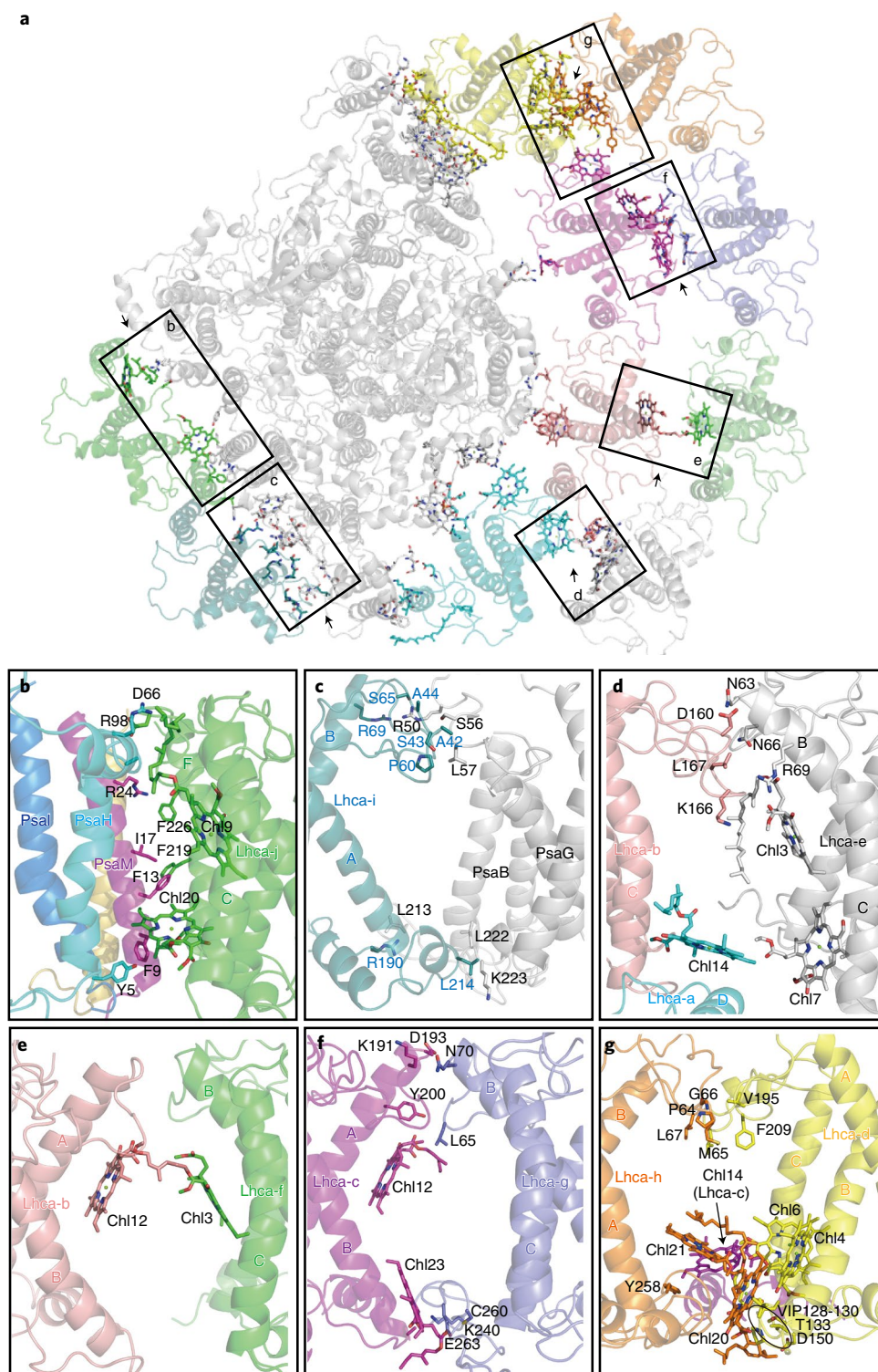


Fig. 3 | Interactions between the outer and inner Lhcas and between the additional Lhca dimers and the PSI core. **a**, The interaction regions in the overall PSI-LHCI supercomplex. Boxed areas are enlarged in the indicated panels. **b–g**, Protein–protein, pigment–pigment and protein–pigment interactions between: Lhca-j and the PSI core (**b**), Lhca-i and the PSI core (**c**), Lhca-b and Lhca-e (**d**), Lhca-b and Lhca-f (**e**), Lhca-c and Lhca-g (**f**), and Lhca-d and Lhca-h (**g**). Orientations of **b–g** are indicated by arrowheads in **a**. In **b**, PsaH, PsaM, PsaI and PsaB are shown using the same colour codes as in Fig. 1, whereas grey is used for the PSI core subunits in the other panels. Side chains of amino acid residues, pigments and lipids that are involved in the interactions are shown.

Further notable differences are found in the long C-terminal loops of Lhca-g and Lhca-h (Fig. 2c–e) which protrude towards Lhca-c and Lhca-d of the inner ring on the luminal side (Supplementary Fig. 11). This results in less inward bending of the outer ring than of the inner ring in LHCI (Fig. 2f).

A double-ring arrangement has been suggested for the *C. reinhardtii* Lhca subunits on the basis of low-resolution electron microscopy and biochemical cross-linking studies^{25–28}. However, our high-resolution structure reveals details of the assembly of the complex and the interactions between subunits (Fig. 3). Within the

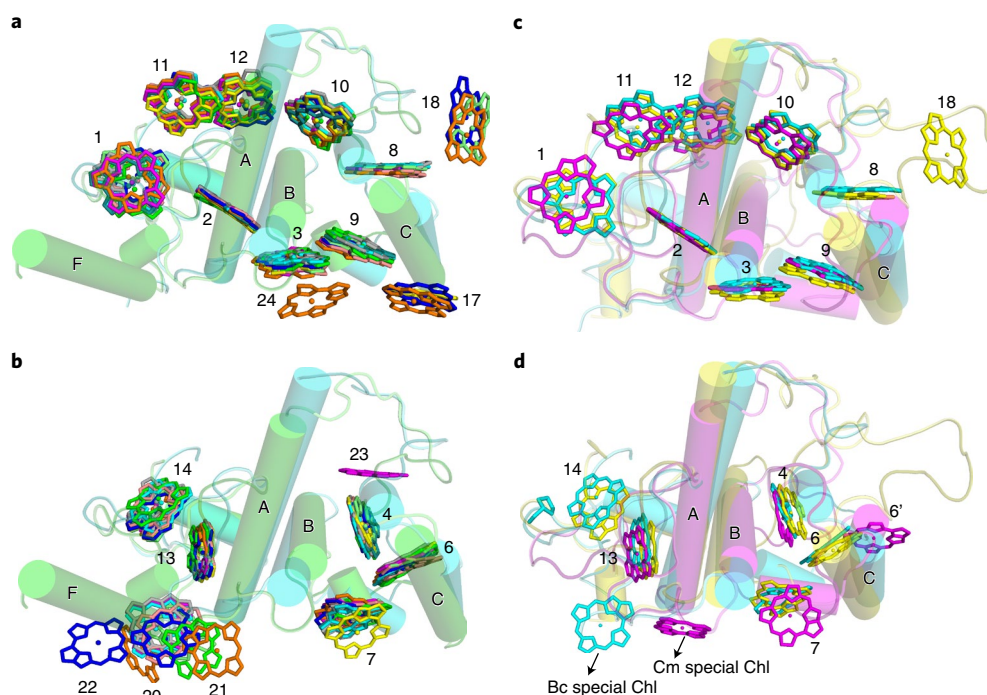


Fig. 4 | Chlorophyll arrangement in the Lhca subunits of *B. corticulans*. **a,b**, Superposition of the structures of the ten Lhca proteins of *B. corticulans*. Shown are chlorophylls at the stromal layer (**a**) and the luminal layer (**b**) using the same colour codes as in Fig. 1, with the apoprotein structure of Lhca-b to Lhca-i omitted. **c,d**, Comparison of chlorophyll arrangement in Lhcas of green algae (*B. corticulans*) and higher plants (*P. sativum*, pea), and in Lhcrs of red algae (*C. merolae*) at the stromal layer (**c**) and at the luminal layer (**d**). Nomenclature of chlorophylls is based on LHCI structure²² and special chlorophylls are labelled. Transmembrane helices are labelled in all of the panels.

double ring structure of the major Lhca belt, loops AC of Lhca-b, Lhca-c and Lhca-d interact with the N-terminus of helix B of Lhca-e, and the N-terminal loops of Lhca-g and Lhca-h, respectively (Fig. 3d,f,g), and Chl 12 of Lhca-b interacts with Chl 3 of Lhca-f (Fig. 3e), on the stromal side. On the luminal side, Chl 14 of Lhca-b interacts with Chl 7 of Lhca-e (Fig. 3d), Chl 23 of Lhca-c interacts with the C-terminal loop of Lhca-g (Fig. 3f), Chl 14 of Lhca-c interacts with Tyr258 at the C-terminal loop of Lhca-h (Fig. 3g), and Chls 4 and 6 and the loop BC region of Lhca-d interact with Chls 20 and 21 of Lhca-h (Fig. 3g). Interactions between each inner Lhca and the PSI core subunits are similar to those in higher plant PSI-LHCI (Supplementary Figs. 12 and 13), with the exception of Chl 20, which is at a position where it can enhance the interactions between the inner Lhca and the PSI core on the luminal side in green algae (Supplementary Figs. 12c and 13d). Sequence alignment of Lhcas from *C. reinhardtii* and *B. corticulans* shows high similarity between the pairs Lhca7/Lhca-c, Lhca3/Lhca-d, Lhca6/Lhca-g, Lhca5/Lhca-h, Lhca2/Lhca-j and Lhca9/Lhca-i (Supplementary Fig. 3). On the basis of these results, we propose the model of subunit organization of PSI-LHCI shown in Supplementary Fig. 14, where subunit-subunit interactions within Lhca and between Lhca and PSI core subunits are shown.

Arrangement of Chls and carotenoids in LHCI

Most of the Chls in the ten Lhcas have counterparts in Lhcas of higher plants and Lhcrs of red algae (Fig. 4). These are Chls 1, 2, 3, 8, 9, 10, 11 and 12 on the stromal side, and Chls 4, 6, 7, 13 and 14 on the luminal side (based on the nomenclature of LHCI²²). Chls 3 and 9 are assigned as red-form dimers, as in higher plant PSI-LHCI⁹. Red-form Chls have redshifted absorption and emission peaks compared to the bulk Chls, and therefore EET from the red Chls to bulk Chls is an energetically 'uphill' process²⁹. In addition to these common Chls, some additional Chls were observed. As shown

in Fig. 4, all Lhcas except Lhca-d and Lhca-i bind one or two Chls (named Chl 20, 21 or 22) at the C-terminal loop area, similar to the Lhcrs of red algae⁴ (Supplementary Fig. 15a). However, in higher plants, only Lhca1 retains this ancient feature^{9–11} (Supplementary Fig. 15b). These 'special' Chls greatly change the luminal Chl arrangement, which may have important functional implications as discussed below. Other uncommon Chls are Chl 24, which lies on the outside of Chl 3 in Lhca-h, and Chl 23 in the stromal loop BC area in Lhca-c (Fig. 4).

The whole PSI-LHCI supercomplex contains 54 carotenoids; 22 are located in the PSI core and the others are located in the ten Lhca antennas. The PSI core of *B. corticulans* binds α -carotenes but not β -carotenes, similar to the PSI core of higher plants and most cyanobacteria. Nevertheless, these organisms share very similar carotenoid arrangements in the PSI core (Supplementary Fig. 16). Carotenoid-binding sites 325 (L1 site) and 326 (L2 site) are conserved in all ten Lhcas, as in LHCI²² (Fig. 5), and the connecting carotenoid 327 between the adjacent Lhcas are found in the inner and outer LHCI rings, similar to the LHCI band of higher plants^{9–11} (Fig. 5 and Supplementary Fig. 16). Different numbers of carotenoids are bound to each Lhca, and two to five carotenoids are found in the individual Lhca subunits (Supplementary Table 2). Previously unknown carotenoid-binding sites were observed in Lhca-g, Lhca-h and Lhca-d (Fig. 5). Carotenoid-binding site 329 in Lhca-g and Lhca-h is close to the luminal side and located between the long C-terminal loop and the red dimer of Chls 3 and 9, which bridges the gap from these subunits to the Lhca-c and Lhca-d subunits. The other new binding site, 329', found in Lhca-d, is located with one end close to the luminal side of helix A and the other end close to the red dimer of Chls 3 and 9, which constitutes a carotenoid cluster, together with its L2-binding site and another four α -carotenes from PsaK and PsaA. This cluster therefore strengthens the interactions between the Lhca subunits and the PSI core. Carotenoid-binding

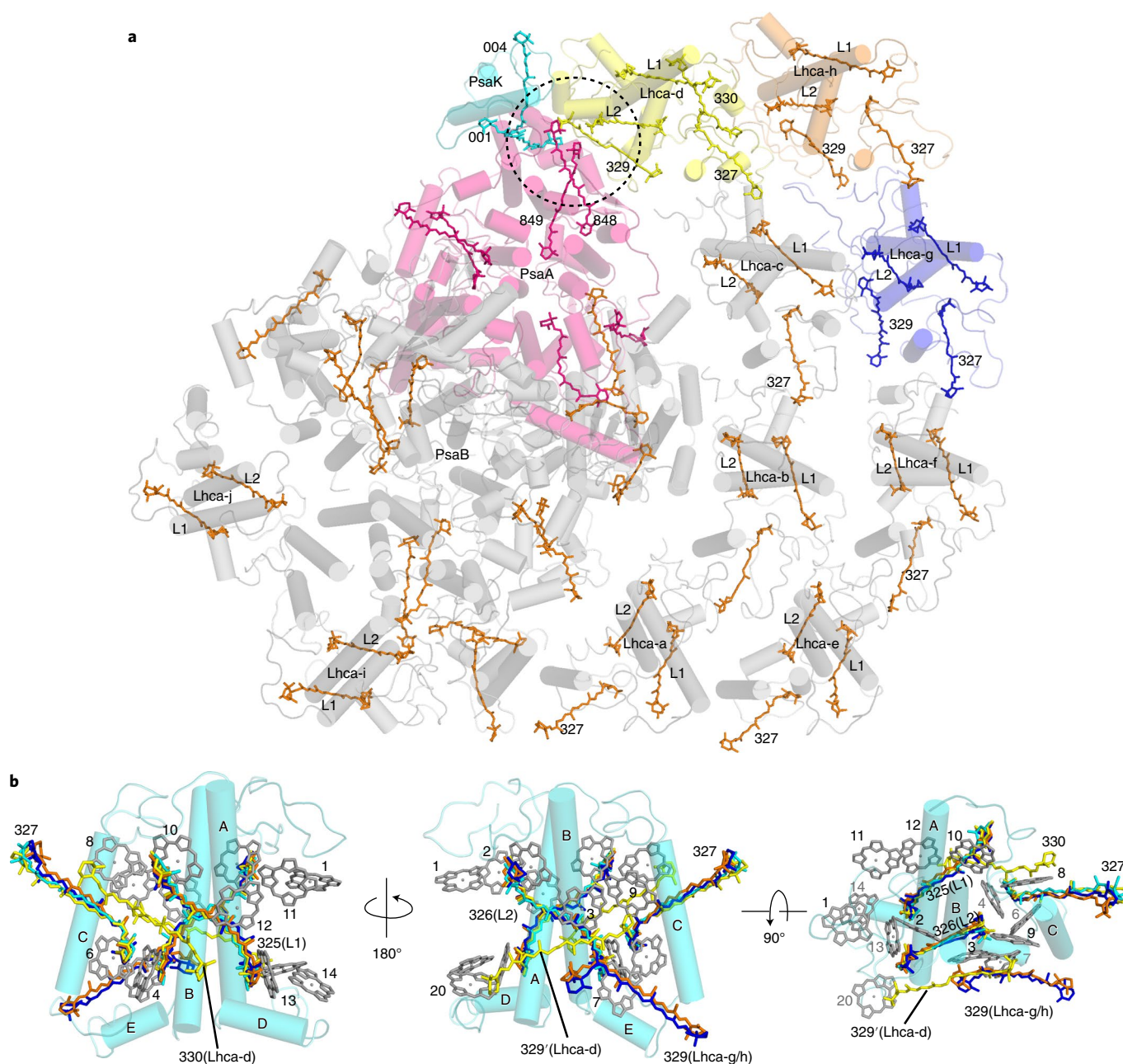


Fig. 5 | Carotenoid arrangement in the green algal PSI-LHCI supercomplex. **a**, Arrangement of carotenoids in the whole PSI-LHCI supercomplex. Carotenoid-binding sites L1 and L2 in the Lhca subunits are labelled based on the LHCI structure²². A carotenoid cluster located in the area of Lhca-d, PsaA and PsaK is circled with black dashed line. **b**, Comparison of the carotenoid arrangement of Lhca-a, Lhca-d, Lhca-g and Lhca-h. The structure of the Lhca-a apoprotein is shown in cyan, Chls of Lhca-a are shown in grey, and carotenoids are shown using the colour code for Lhcas shown in Fig. 1. Residue numbers of carotenoid and Chls (PDB 6IGZ) are labelled in **b**.

site 330 binds at the outside of helix B and is nearly parallel to the connecting carotenoid 327. The condensed carotenoid arrangement in Lhca-d may facilitate its function as one of the main EET exits to the PSI core, and may also enhance the quenching capacity, which is needed for *B. corticulans* to survive under the highly fluctuating light conditions in the intertidal zones that it inhabits.

Possible energy-transfer pathways

In our structure, the ten Lhcas contained 119 molecules of Chl *a*, 26 molecules of Chl *b* and 33 carotenoids (Supplementary Table 2). The total number of Chls is similar to the number determined chemically; however, the amount of Chl *b* assigned is less than that determined chemically, as some Chls *b* may have not been identified

due to the limited resolution. Overall, the Chls are distributed in two layers. One layer is close to the stromal surface and contains 84 Chls; the other layer is close to the luminal surface and contains 61 Chls (Fig. 6). Comparison with that of the higher plant LHCI revealed 8 and 11 unique Chls at the stromal layer and the luminal layer, respectively, in the green algal LHCI (Fig. 6 and Supplementary Fig. 13). Among all of the Chls, those with the shortest Mg–Mg distances can be categorized into three types: (1) between Chls in Lhcas and the PSI core; (2) between Chls in the outer and inner rings of Lhcas; and (3) between Chls in adjacent Lhcas. In (1), the red Chl 3–Chl 9 dimers of Lhca-a and Lhca-d are key Chls that connect Lhcas to the PSI core on the stromal side. However, the Mg–Mg distances between these red Chl dimers and the PSI core are not homogeneous; the dimers of

Lhca-a and Lhca-d are closer to the PSI core than those of Lhca-b and Lhca-c. On the basis of the Mg–Mg distances, we propose that the following Chl pairs provide the main EET pathways from the inner LHCI ring to the PSI core on the stromal side: Lhca-a Chl 3–PsaB Chl 841 (14.3 Å), Lhca-a Chl 9–PsaB Chl 822 (16.6 Å), Lhca-d Chl 3–PsaA Chl 811 (16.4 Å) and Lhca-d Chl 2–PsaA Chl 813 (16.5 Å) (Fig. 6a). Distances in category (2) are between the red dimers in the outer LHCI ring and the Chl 10–Chl 11–Chl 12 clusters in the inner LHCI ring, which include two Lhca pairs, Lhca-e–Lhca-a and Lhca-f–Lhca-b. In addition, the red Chl pairs of Lhca-h and Lhca-c are connected through two intermediate Chls, Lhca-h Chl 17 and Lhca-h Chl 24. These pathways are thus important for energy transfer from the outer ring to the inner ring on the stromal side. This arrangement may reduce the energy gap between the red forms of the outer ring and energy-accepting Chls of the inner ring and therefore smooth EET between them; the Chl 10–Chl 11–Chl 12 cluster has been suggested to be a low-energy site of LHCI³⁰. Distances in category (3) may be responsible for the EET between adjacent Lhcas mediated by the red dimer and Chl 8 of one Lhca, and Chl 1 of the adjacent Lhca. Therefore, the Chl arrangement in the stromal layer indicates multiple functions of some red dimers in the EET pathways from LHCI to the PSI core and between different Lhcas.

The Chl arrangement at the luminal layer of *B. corticulans* is remarkably different from that of the higher plant PSI–LHCI (Supplementary Fig. 13d). In higher plants, Lhca2, Lhca3 and Lhca4 proteins harbour a Chl 4–Chl 6–Chl 7 cluster and a Chl 13–Chl 14 cluster at the luminal layer (Supplementary Fig. 15b), and their sparse packing results in a large gap between Lhcas and the PSI core^{9–11}. This gap may be partly filled by the PsaN subunit revealed in the maize PSI–LHCI–LHCII structure that binds two Chls and connects Lhca2 with the PSI core¹⁷, but a large gap remains between Lhca4 and the higher plant PSI core at the luminal layer. By contrast, all Lhcas in green algae except Lhca-d and Lhca-i make use of the Chl 13–14–20–21–22 cluster to enhance the pigment density (Figs. 4b and 6b). Eleven special Chls at the luminal layer serve as the ‘gap Chls’ to connect the EET pathways from the inner and side Lhcas to the PSI core, including the pairs Lhca-a Chl 20–Psa F Chl 308^{PsaF} (13.0 Å), Lhca-b Chl 20–Psa J Chl 3002^{PsaJ} (17.8 Å), Lhca-c Chl 20–Psa A Chl 817 (18.8 Å) and Lhca-j Chl 20–PsaB Chl 807 (18.0 Å), and to connect the EET pathways from outer Lhcas to the inner Lhcas including the Chl pairs Lhca-e Chl 20–Lhca-b Chl 4 (16.8 Å), Lhca-f Chl 20–Lhca-b Chl 14 (19.5 Å), Lhca-g Chl 20–Lhca-c Chl 23 (17.3 Å), Lhca-g Chl 22–Lhca-c Chl 14 (18.0 Å), Lhca-h Chl 20–Lhca-d Chl 4 (9.7 Å) and Lhca-h Chl 21–Lhca-d Chl 6 (14.1 Å) (Fig. 6b). In addition, the average shortest Mg–Mg distance between the adjacent Lhcas is 16.8 Å (Fig. 6b).

The average Mg–Mg distance between these special Chls at the luminal layer and the red dimers at the stromal side is around 20.7 Å (Supplementary Fig. 17), which is not too far for them to couple; these Chls may therefore also affect the energy transfer between the two layers. Notably, the shortest Mg–Mg distances between the two layers were found between Chl 17 at the stromal layer and Chl 20 at the luminal layer (Lhca-d Chl 17–Lhca-c Chl 20, 13.5 Å; Lhca-h Chl 17–Lhca-g Chl 20, 12.4 Å), which may facilitate the energy transfer between the two layers.

Discussion

The energy-transfer efficiency of PSI depends on the cross-sectional area of LHCI and the energy levels and distances between adjacent Chls; the energy levels and distances between Chls in turn depend on the Chl arrangements within the protein environment. Structural comparison between the higher plant PSI–LHCI and the green algal PSI–LHCI suggests that each inner Lhca in the green algal PSI–LHCI has direct pathways to the PSI core. Among these direct pathways, Lhca-a and Lhca-d located at the two side edges of the inner LHCI ring make close interactions with the Chls of PsaB

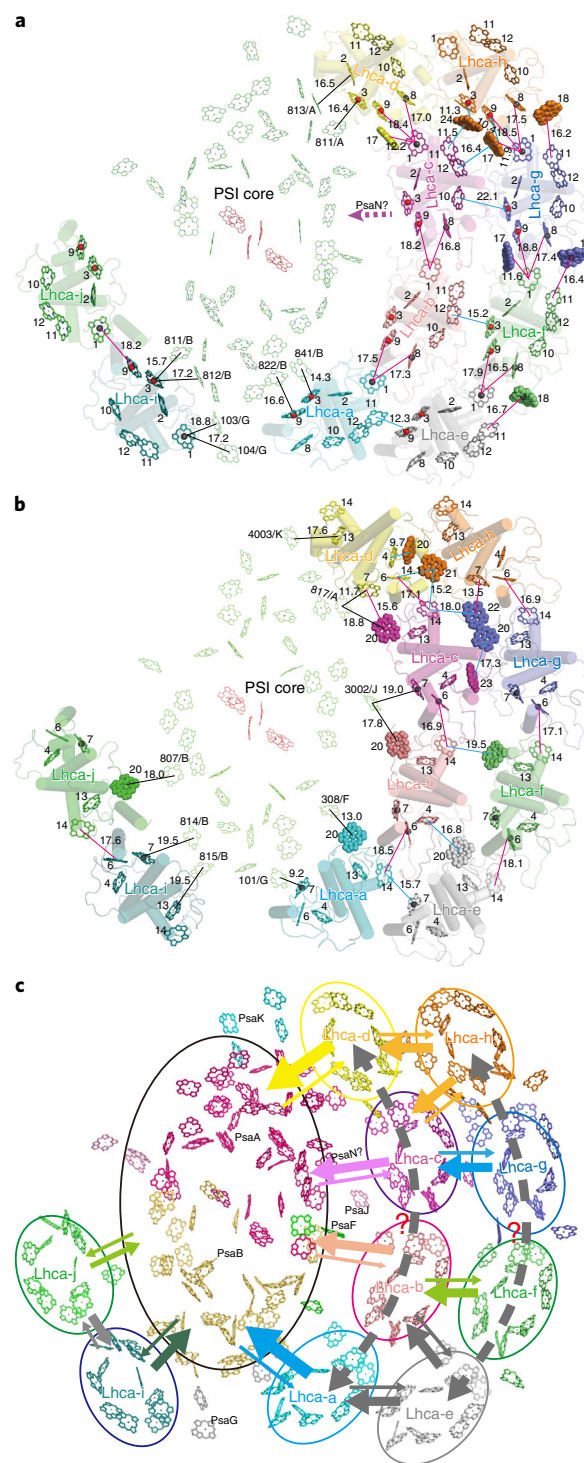


Fig. 6 | Possible energy-transfer pathways in the PSI–LHCI supercomplex of *B. corticulans*. **a, b**, Positions of Chls that are key to energy-transfer pathways at the stromal layer (**a**) and the luminal layer (**b**); both are viewed from the stromal side. Special Chls are shown in spheres. The central magnesium atoms are shown in red spheres for the red dimers, grey spheres for Chl *b* and omitted for other Chls. Chls of the PSI core are shown in red for the four electron transfer chain Chls and in green for the bulk Chls. Black lines indicate pathways from the Chls of inner Lhcas to the Chls of the PSI core, pink lines indicate pathways from Chls between adjacent Lhcas, and blue lines indicate pathways from Chls of the outer Lhca ring to Chls of the inner Lhca ring. **c**, Possible energy-transfer pathways from LHCI to the PSI core and within LHCI. The colour code is the same as in Fig. 1.

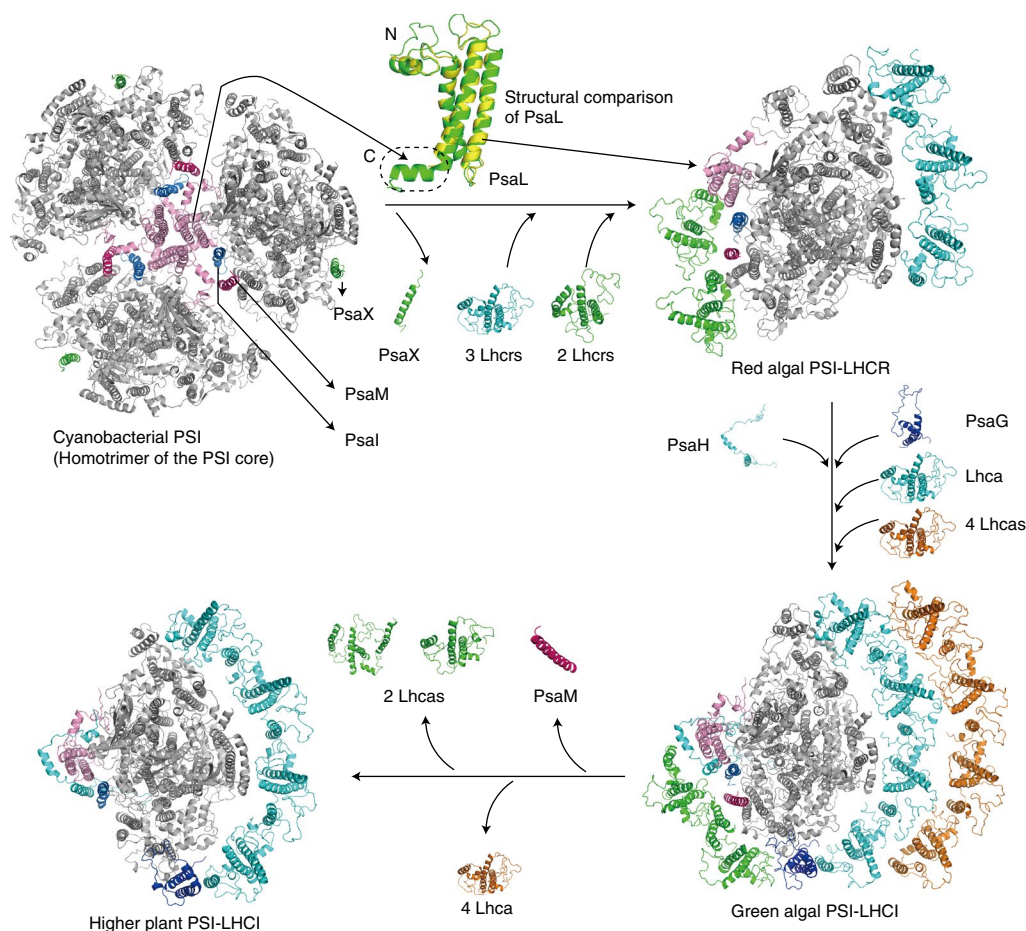


Fig. 7 | Evolutionary scheme for the structural changes of PSI-LHCI from cyanobacteria to higher plants. PDB IDs: 1JB0, PSI of a cyanobacterium (*Thermosynechococcus elongatus*); 5ZGB, PSI-LHCR of a red alga (*C. merolae*); 6IGZ, PSI-LHCI of a green alga (*B. corticulans*); 4XK8, PSI-LHCI of a higher plant (*P. sativum*).

and PsaA, respectively, which may provide the main energy-transfer pathways from LHCI to the PSI core, similar to that seen for Lhca1 and Lhca3 in the higher plant PSI-LHCI (Fig. 6). Lhca-b and Lhca-c, located in the middle of the inner LHCI ring, may also contribute to the energy transfer owing to the presence of Chl 20, which covers the large spatial gap between the inner LHCI ring and the PSI core on the luminal side (Supplementary Fig. 13). Furthermore, PsaN may also contribute to the energy transfer from Lhca-c to the PSI core, since it is found in a position between the space of Lhca-c and the PSI core in higher plant PSI-LHCI¹⁷ and may have been lost in our preparation. In addition, Lhca-i may transfer energy from the LHCI dimer located on the PsaG–PsaB side to the PSI core.

Green algal PSI-LHCI has a shorter fluorescence emission wavelength than that of higher plants at low temperature^{6,8,27,31}, suggesting that the green algal PSI-LHCI have red Chl forms with higher energy than higher plant PSI-LHCI, reducing the energy gap between the red forms and bulk Chls and hence facilitating the uphill EET process in green algae^{12,13}. The exact reason for this is unknown at present, but our structure may provide a basis for future studies. In summary, PSI-LHCI with an enlarged antenna cross-section in the green algae bears a highly efficient EET system, which is vital for light energy absorption under low intensity conditions in water. Binding of xanthophylls (siphonaxanthin, siphonein, all-*trans* and 9'-*cis* neoxanthin) in Lhcas and α -carotene in the PSI core, together with several new binding sites for the carotenoids in Lhcas revealed here, may also provide strategies for green algae to survive in the dynamically changing light environment^{15,16}.

Interestingly, red algal PSI-LHCRs appear to have more efficient EET pathways than green algal PSI-LHCIs, since the red forms of PSI-LHCI in the red algae are closer to the PSI core at the stromal side, and the special Chls (Chl 615) are also closer to the PSI core at the luminal side than the respective Chls in the green algal PSI-LHCI (Supplementary Fig. 18) in both the major LHCI belt and the additional Lhca-dimer side. This may have been caused by the shift of the major Lhc belt towards the PsaG side due to the appearance of PsaG, and the move of the additional Lhc dimer away from the PSI core due to the appearance of PsaH, in the green algal PSI-LHCI. These changes may be correlated with the gain of the additional Lhca subunits in the double-Lhca ring of the green algal PSI-LHCI to compensate for the reduced light-harvesting efficiency. Loss of the additional Lhca dimer from the higher plant PSI-LHCI may have been caused by the loss of PsaM during evolution, whereas loss of the outer LHCI belt may be a result of adaptation to the land environment, where sufficient light intensity is available, and for protection of terrestrial plants from photodamage by decreasing the capacity of light harvesting under high irradiation. Thus, the cryo-EM structure of the green algal PSI-LHCI supercomplex provides important insights on the evolution of PSI-LHCI as a result of adaptation to various light environments inhabited by photosynthetic organisms. The major structural differences in the PSI and PSI-LHCI supercomplex from cyanobacteria to higher plants are summarized in Fig. 7, in which the gain and loss of PSI core subunits, and the changes in the composition and organization of the Lhca subunits are visualized.

Methods

Sample purification and characterization. *B. corticulans* was collected from intertidal zones around the seashore of Qingdao city, China, and washed with distilled water. Isolation of thylakoids was performed as previously described⁴⁶, and the thylakoid membranes were solubilized with 1.0% (w/v) *n*-dodecyl- β -D-maltoside (β -DDM) at 1 mg Chl per ml in an ice bath for 20 min. The unsolubilized materials were removed by centrifugation at 40,000g for 15 min and the supernatant was loaded onto a 0.3–1.2 M continuous sucrose density gradient (containing 20 mM Tricine-NaOH (pH 7.8), 0.03 % (w/v) β -DDM) and centrifuged at 37,000 r.p.m. for 16 h with a SW40 rotor at 4°C (Beckman). The lower band containing PSI–LHCI (Supplementary Fig. 1a) was collected and precipitated by centrifugation at 15,000g for 5 min after addition of PEG 6,000 to a final concentration of 10% (w/v). The final pellet was dissolved in 20 mM Tricine-NaOH (pH 7.8) and used for cryo-EM specimen preparation.

PSI–LHCI samples were also prepared from pea and *C. reinhardtii* as previously described^{8,9}. To prevent the peripheral antenna proteins from being released during the purification procedure of PSI–LHCI from *C. reinhardtii*, samples obtained by sucrose density gradient centrifugation were not purified further by column chromatography. The polypeptide compositions of PSI–LHCI from *B. corticulans*, pea and *C. reinhardtii* were analysed by SDS–PAGE. Samples were treated with a sample buffer containing 2% (w/v) lithium dodecyl sulfate, 60 mM dithiothreitol and 60 mM Tris–HCl (pH 8.5) at 60°C for 10 min, and subjected to SDS–PAGE with a 16% gel containing 7.5 M urea as previously described³².

Sequencing of the PSI core and Lhca genes from *B. corticulans*. Total RNA samples were isolated from *B. corticulans* for RNA sequencing. De novo assembly was used in data processing. The predicted CDS sequences of PSI core subunits and Lhcas were picked out and aligned with corresponding genes of other green algae. To confirm the CDS sequences of Lhcas, RNA (approximately 2 μ g) was converted to cDNA, and each gene was amplified by PCR using primers designed according to the predicted coding sequence (CDS). Five clones from each PCR product were sequenced, and the sequences obtained were confirmed for consistency. Moreover, some Lhcas CDSs were cloned by homology sequencing analysis in which, the *Lhca* genes from different species were aligned and their conserved regions were selected to design primers to amplify the *Lhca* homologues from *B. corticulans*. Twenty clones from each PCR product were sequenced, and a cDNA amplification kit for 5' and 3' RACE (SMARTer RACE 5'/3' Kit, cat. nos. 634858 and 634859, respectively; Takara) were used to obtain the flanking sequences in the upstream and downstream regions according to the manufacturers' instructions.

Chloroplast genome sequencing and assembly. To obtain amino acid sequences of the chloroplast-coded PSI subunits, PsaA, PsaB, PsaC, PsaI, PsaJ and PsaM, *B. corticulans* chloroplasts were isolated from freshly collected plant materials and were used for chloroplast genome sequencing. Genomic DNA was extracted using the SDS method, and the collected DNA was detected by agarose gel electrophoresis and quantified using Qubit. Whole-genome sequencing was performed on the Illumina HiSeq PE 150 platform. A paired-end library was constructed according to the manufacturer's instructions (Illumina). The library, which contained an insert size of 350 bp, was sequenced using the Illumina HiSeq PE 150 platform by Beijing Novogene Bioinformatics Technology Co., Ltd. Illumina PCR adapter reads and low-quality reads from the paired-end were filtered. All good-quality paired reads were assembled using SOAPdenovo^{33,34} (<http://soap.genomics.org.cn/soapdenovo.html>) into a number of scaffolds, and then filter reads were handled by gap closing. Sequence prediction of protein-coding genes was performed by genome analysis of the closely related species *Bryopsis hypnoides* using MIT biom v.1.8 software.

Mass spectrometry analysis. Bands containing the targeted proteins were cut out from SDS–PAGE and digested as described previously³⁵. Online nano-reversed phase liquid chromatography (nano-RPLC) was used for mass spectrometry analysis on the Eksigent nanoLC-Ultra 2D System (AB SCIEX). The samples were loaded on a C18 nanoLC trap column (100 μ m $\times$ 3 cm, C18, 3 μ m, 150 Å) and washed with nano-RPLC buffer A (0.1% formic acid, 2% acetonitrile) at 2 μ l min^{−1} for 10 min. A gradient of 5–35% acetonitrile (0.1% formic acid) was used to elute the sample for 90 min on an analytical ChromXP C18 column (75 μ m $\times$ 15 cm, C18, 3 μ m 120 Å) with a spray tip. Data acquisition was performed with a Triple TOF 5600 System (AB Sciex) fitted with a Nanospray III source (AB Sciex) and a pulled quartz tip as the emitter (New Objectives). Data were acquired using an ion-spray voltage of 2.5 kV, curtain gas of 30 PSI, nebulizer gas of 5 PSI, and an interface heater temperature of 150°C. For information-dependent acquisition (IDA), survey scans were acquired in 250 ms and as many as 35 product ion scans were collected if they exceeded a threshold of 150 counts per second (counts s^{−1}) with a 2⁺ charge state. The total cycle time was fixed to 2.5 s. A rolling collision-energy setting was applied to all precursor ions for the collision-induced dissociation (CID). Dynamic exclusion was set for half of peak width (18 s) and the precursor was then refreshed off the exclusion list. On the basis of the combined MS and MS/MS spectra, protein identification was performed by the PEAKS 8.5 software (Bioinformatics Solutions).

Cryo-EM data collection. An aliquot of 4 μ l of the green algal PSI–LHCI sample at a concentration of ~3 mg Chl per ml was applied to glow-discharged Quantifoil holey carbon Cu grids (R 1.2/1.3, 400 mesh). The grids were blotted for 3 s and plunged into liquid ethane cooled by liquid nitrogen with an FEI Vitrobot Mark IV at 100% humidity and 8°C. The samples were imaged on an FEI Titan Krios electron microscope equipped with a Cs corrector and operated at 300 kV with a nominal magnification of $\times 75,000$. Micrographs were recorded with an FEI Falcon II direct electron detector with a pixel size of 0.8727 Å, with defocus values varied from 1.5 to 2.5 μ m. Each micrograph was dose-fractionated to 27 frames with a total exposure time of 1.6 s and a total dose rate of approximately 50 e[−] Å^{−2}. AutoEMationII³⁶ was used for fully automated data collection, and 3,665 micrographs were collected. The micrographs were processed for motion correction by MotionCorr³⁷, and then further processed by MotionCor2³⁸ with dose weighting.

Data processing. The defocus extents of the images were estimated by CTFFIND4³⁹ using the motion-corrected micrographs without dose weighting. A small dataset of 3,416 particles of the PSI–LHCI supercomplex were manually picked and processed by reference-free 2D classification using RELION⁴⁰. Five 2D averages were selected as references for auto-picking of all 3,665 micrographs, and 473,432 particles were automatically picked using RELION and subjected to the reference-free 2D classification. On the basis of the 2D classification, 279,111 particles were selected and subjected to 3D classification. An atomic model of the PSI–LHCI⁹ was low-pass filtered to 60 Å and used as the initial model for the 3D classification, and 59,525 particles were selected from the 3D classification for further 3D auto-refinement in RELION.

The number of classes in the 3D classification was set to 4 (Supplementary Fig. 4), and the results obtained showed that class 2, class 3 and class 4 correspond to intact PSI–LHCI structure with ten Lhcas, which collectively accounted for 78% of the particles, whereas class 1 represents the structure of PSI–LHCI with eight Lhcas but without the additional dimer of Lhca-i and Lhca-j, which accounted for 22% of the particles. To improve the resolution of the PSI–LHCI, we merged the total of 59,525 particles and finally obtained the 3.49 Å map based on the gold-standard Fourier shell correlation = 0.143 criteria⁴¹ (Supplementary Fig. 5b), and the local resolution map was calculated with ResMap⁴² (Supplementary Fig. 5d).

Model building and structural refinement. For model building of the green algal PSI–LHCI supercomplex, the structure of the pea PSI core taken from the PDB 4XK8 coordinates⁹ was automatically fitted into the 3.49 Å cryo-EM map with UCSF Chimera⁴³. The amino acid sequences were then mutated to its counterparts in *B. corticulans*. The models of LHC proteins were first predicted using the Phyre2⁴⁴ web server (<http://www.sbg.bio.ic.ac.uk/phyre2>) and then fitted into the 3.49 Å cryo-EM map with UCSF Chimera. The overall model of PSI–LHCI was manually adjusted against the map and additional features were added on the basis of the cryo-EM density map with COOT⁴⁵, followed by real-space automatic refinement against the cryo-EM map by Phenix⁴⁶. The resulting model was checked and corrected manually again with COOT, and this process was repeated until the structure matched the cryo-EM density map to the maximum extent. During the real-space refinement, the distances between the central magnesium ion of Chl molecules and the coordinating ligands were restrained with the values obtained from high-resolution crystal structures.

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

Data availability

The cryo-EM density map and atomic models for the PSI–LHCI supercomplex structure at 3.49 Å have been deposited in the Electron Microscopy Data Bank and the Protein Data Bank under accession codes EMD-9670 and 6IGZ, respectively. The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Received: 5 October 2018; Accepted: 4 February 2019;

Published online: 8 March 2019

References

- Watanabe, M. & Ikeuchi, M. Phycobilisome: architecture of a light-harvesting supercomplex. *Photosynth. Res.* **116**, 265–276 (2013).
- Zhang, J. et al. Structure of phycobilisome from the red alga *Griffithsia pacifica*. *Nature* **551**, 57–63 (2017).
- Büchel, C. Evolution and function of light harvesting proteins. *J. Plant Physiol.* **172**, 62–75 (2015).
- 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).
- Germano, M. et al. Supramolecular organization of photosystem I and light-harvesting complex I in *Chlamydomonas reinhardtii*. *FEBS Lett.* **525**, 121–125 (2002).

6. Kargul, J., Nield, J. & Barber, J. Three-dimensional reconstruction of a light-harvesting complex I-photosystem I (LHCI-PSI) supercomplex from the green alga *Chlamydomonas reinhardtii*—insights into light harvesting for PSI. *J. Biol. Chem.* **278**, 16135–16141 (2003).
7. Kargul, J. et al. Light-harvesting complex II protein CP29 binds to photosystem I of *Chlamydomonas reinhardtii* under state 2 conditions. *FEBS J.* **272**, 4797–4806 (2005).
8. Takahashi, Y., Yasui, T., Stauber, E. J. & Hippler, M. Comparison of the subunit compositions of the PSI-LHCI supercomplex and the LHCI in the green alga *Chlamydomonas reinhardtii*. *Biochemistry* **43**, 7816–7823 (2004).
9. Qin, X., Suga, M., Kuang, T. & Shen, J.-R. Structural basis for energy transfer pathways in the plant PSI-LHCI supercomplex. *Science* **348**, 989–995 (2015).
10. Mazor, Y., Borovikova, A., Caspy, I. & Nelson, N. Structure of the plant photosystem I supercomplex at 2.6 Å resolution. *Nat. Plants* **3**, 17014 (2017).
11. Suga, M., Qin, X., Kuang, T. & Shen, J.-R. Structure and energy transfer pathways of the plant photosystem I-LHCI supercomplex. *Curr. Opin. Struct. Biol.* **39**, 46–53 (2016).
12. Giera, W. et al. Excitation dynamics in photosystem I from *Chlamydomonas reinhardtii*. Comparative studies of isolated complexes and whole cells. *Biochim. Biophys. Acta* **1837**, 1756–1768 (2014).
13. Quiniou, C. L. et al. PSI-LHCI of *Chlamydomonas reinhardtii*: increasing the absorption cross section without losing efficiency. *Biochim. Biophys. Acta* **1847**, 458–467 (2015).
14. Giovagnetti, V. et al. A siphonous morphology affects light-harvesting modulation in the intertidal green macroalga *Bryopsis corticulans* (Ulvophyceae). *Planta* **247**, 1293–1306 (2018).
15. Wang, W. et al. Spectral and functional studies on siphonaxanthin-type light-harvesting complex of photosystem II from *Bryopsis corticulans*. *Photosynth. Res.* **117**, 267–279 (2013).
16. Qin, X. et al. Isolation and characterization of a PSI-LHCI super-complex and its sub-complexes from a siphonaceous marine green alga *Bryopsis corticulans*. *Photosynth. Res.* **123**, 61–76 (2015).
17. Pan, X. et al. Structure of the maize photosystem I supercomplex with light-harvesting complexes I and II. *Science* **360**, 1109–1113 (2018).
18. Busch, A. & Hippler, M. The structure and function of eukaryotic photosystem I. *Biochim. Biophys. Acta* **1807**, 864–877 (2011).
19. Jordan, P. et al. Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* **411**, 909–917 (2001).
20. Rochaix, J. D. Regulation and dynamics of the light-harvesting system. *Annu. Rev. Plant Biol.* **65**, 287–309 (2014).
21. Kana, R. et al. Phycobilisome mobility and its role in the regulation of light harvesting in red algae. *Plant Physiol.* **165**, 1618–1631 (2014).
22. Liu, Z. et al. Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution. *Nature* **428**, 287–292 (2004).
23. Fan, M. et al. Crystal structures of the PsbS protein essential for photoprotection in plants. *Nat. Struct. Mol. Biol.* **22**, 729–735 (2015).
24. Green, B. R. & Pichersky, E. Hypothesis for the evolution of three-helix Chl *a/b* and Chl *a/c* light-harvesting antenna proteins from two-helix and four-helix ancestors. *Photosynth. Res.* **39**, 149–162 (1994).
25. Drop, B., Yadav, K. N. S., Boekema, E. J. & Croce, R. Consequences of state transitions on the structural and functional organization of photosystem I in the green alga *Chlamydomonas reinhardtii*. *Plant J.* **78**, 181–191 (2014).
26. Steinbeck, J. et al. Structure of a PSI-LHCI-cyt *b₆f* supercomplex in *Chlamydomonas reinhardtii* promoting cyclic electron flow under anaerobic conditions. *Proc. Natl Acad. Sci. USA* **115**, 10517–10522 (2018).
27. Drop, B. et al. Photosystem I of *Chlamydomonas reinhardtii* contains nine light-harvesting complexes (Lhca) located on one side of the core. *J. Biol. Chem.* **286**, 44878–44887 (2011).
28. Ozawa, S.-I. et al. Configuration of ten light-harvesting chlorophyll *a/b* complex I subunits in *Chlamydomonas reinhardtii* photosystem I. *Plant Physiol.* **178**, 583–595 (2018).
29. Croce, R. & van Amerongen, H. Light-harvesting in photosystem I. *Photosynth. Res.* **116**, 153–166 (2013).
30. Novoderezhkin, V. I., Palacios, M. A., van Amerongen, H. & van Grondelle, R. Excitation dynamics in the LHCI complex of higher plants: modeling based on the 2.72 Å crystal structure. *J. Phys. Chem. B* **109**, 10493–10504 (2005).
31. Kuang, T., Argyroudi-Akoyunoglou, J. H., Nakatani, H. Y., Watson, J. & Arntzen, C. J. The origin of the long-wavelength fluorescence emission band (77 K) from photosystem I. *Arch. Biochem. Biophys.* **235**, 618–627 (1984).
32. Ikeuchi, M. & Inoue, Y. A new 4.8-kDa polypeptide intrinsic to the PSII reaction center, as revealed by modified SDS-PAGE with improved resolution of low-molecular-weight proteins. *Plant Cell Physiol.* **29**, 1233–1239 (1988).
33. Li, R., Li, Y., Kristiansen, K. & Wang, J. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**, 713–714 (2008).
34. Li, R. et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265–272 (2010).
35. Katayama, H., Nagasu, T. & Oda, Y. Improvement of in-gel digestion protocol for peptide mass fingerprinting by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **15**, 1416–1421 (2001).
36. Lei, J. & Frank, J. Automated acquisition of cryo-electron micrographs for single particle reconstruction on an FEI Tecnai electron microscope. *J. Struct. Biol.* **150**, 69–80 (2005).
37. Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
38. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
39. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
40. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
41. Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).
42. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
43. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
44. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
45. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
46. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

Acknowledgements

We thank J. Lei and the staff at the Tsinghua University Branch of the National Center for Protein Sciences Beijing for providing facility support, and the Explorer 100 cluster system of the Tsinghua National Laboratory for Information Science and Technology for providing computation resources. We thank S. Qin and Z. Liu from Yantai Institute of Coastal Zone Research, CAS for assistance in collecting *B. corticulans*, L. Shu from Shanghai Luming Biotechnology for mass spectrometry analysis. The project was funded by the National Key R&D Program of China (2017YFA0503700, 2016YFA0501101, 2015CB150101); the National Natural Science Foundation of China (31622007, 31670237, 31600191); a Strategic Priority Research Program of CAS (XDB17000000); State Key Laboratory of Membrane Biology; and Taishan Scholars Project.

Author contributions

T.K. and S.-F.S. conceived the project; X.Q. performed the sample preparation, characterization and sequence analysis; X.P. processed the cryo-EM data, built and refined the structure model; W.W. and G.H. assisted in sample preparation; L.Z. and M.L. cloned *Lhcas* from *B. corticulans* and assisted in sequence analysis; X.Q. and X.P. analysed the structure; X.Q., X.P., J.-R.S., T.K. and S.-F.S. wrote the manuscript and all authors discussed and commented on the results and the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41477-019-0379-y>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to T.K. or S.-F.S.

Journal peer review information: *Nature Plants* thanks Egbert Boekema, Jean-David Rochaix and the other anonymous reviewer(s) for their contribution to the peer review of this work.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019

Reporting Summary

Nature Research 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 Research policies, see [Authors & Referees](#) 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 (<i>n</i>) 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
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☒ | ☐ 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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ 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 <i>d</i> , Pearson's <i>r</i>), 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	AutoEMationII
Data analysis	ResMap, PEAKS 8.5, MIT biom V1.8, SOAP denovo, RELION 1.4, Coot, PHENIX, PyMol, Chimera, MotionCorr, MotionCorr2, CTFFIND4, Phyre2

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [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 list of figures that have associated raw data
- A description of any restrictions on data availability

The cryo-EM density map and atomic models have been deposited in the Electron Microscopy Data Bank and the Protein Data Bank for the PSI-LHCI supercomplex structure at 3.49 Å (EMD ID code 9670 And PDB ID code 6IGZ)

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	No statistical methods were used to predetermine sample size. The sample size was chosen following the guidelines from the instrument manufacturer.
Data exclusions	A fraction of the acquired cryo-EM movies were discarded by visual inspection, 2D classification and 3D classification because of excessive movement, low defocus, high defocus, or over-focus. These discarded movies lost the high-frequency signal for the high resolution reconstruction.
Replication	cryo-EM Data collection and structure calculation were not repeated
Randomization	Samples were not allocated into experimental groups because the goal of this study was not to evaluate the impact of a particular factor, but to simply determine the structure of a protein.
Blinding	Not applicable. There was no existing protein structure to refer to and the data analysis was ab initio.

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
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	The study did not involve laboratory animals.
Wild animals	The study did not involve wild animals.
Field-collected samples	The material used in this study, a green alga of <i>B. corticulans</i> , was collected from intertidal zones around seashore of Qingdao city, China
Ethics oversight	No ethical approval was required because the material used in the study was a green alga collected from intertidal zones around seashore.

Note that full information on the approval of the study protocol must also be provided in the manuscript.