ARTICLE

doi:10.1038/nature24278

Structure of phycobilisome from the red alga *Griffithsia pacifica*

Jun Zhang[1]†*, Jianfei Ma[1]*, Desheng Liu[1], Song Qin[2], Shan Sun[1], Jindong Zhao[3,4] & Sen-Fang Sui[1]

Life on Earth depends on photosynthesis for its conversion of solar energy to chemical energy. Photosynthetic organisms have developed a variety of light-harvesting systems to capture sunlight. The largest light-harvesting complex is the phycobilisome (PBS), the main light-harvesting antenna in cyanobacteria and red algae. It is composed of phycobiliproteins and linker proteins but the assembly mechanisms and energy transfer pathways of the PBS are not well understood. Here we report the structure of a 16.8-megadalton PBS from a red alga at 3.5 Å resolution obtained by single-particle cryo-electron microscopy. We modelled 862 protein subunits, including 4 linkers in the core, 16 rod–core linkers and 52 rod linkers, and located a total of 2,048 chromophores. This structure reveals the mechanisms underlying specific interactions between linkers and phycobiliproteins, and the formation of linker skeletons. These results provide a firm structural basis for our understanding of complex assembly and the mechanisms of energy transfer within the PBS.

The cyanobacteria and red algae are two important groups of photosynthetic organisms that share a light-harvesting antenna known as the phycobilisome (PBS)[1–4]. PBSs are among the largest protein complexes in the living world and consist of phycobiliproteins (PBPs), including phycocyanin, phycoerythrin and allophycocyanin (APC), and linker proteins[1]. Absorption of light by PBSs is accomplished by open-chain tetrapyrroles (bilins) acting as chromophores attached covalently to PBPs and linkers[1]. Two subunits of PBPs, the α- and β-subunits, form an αβ heterodimer that is conventionally called an (αβ) monomer. The monomer is then assembled into an $(\alpha\beta)_3$ trimer, the basic unit of PBS hierarchical assembly. The trimers of various PBPs are organized into a highly ordered supramolecular complex with the help of the linker proteins[1,5,6]. Four morphological types of PBS[1] are known: hemidiscoidal[7], hemiellipsoidal[8], block-type[9] and bundle-type[10]. The hemidiscoidal PBS contains a central core surrounded by peripheral rods[1,11,12]. The chromophores in hemidiscoidal PBS are arranged in such a way that a photon absorbed by a chromophore in the peripheral rods is rapidly funnelled to chromophores in the core[13] and eventually to the terminal emitters (the core–membrane linker protein (L_{CM})[14–16] or allophycocyanin D (ApcD)[17,18]). The terminal emitters then transfer the energy to photosynthetic reaction centres[16,19–23]. Currently, the mechanism of PBS assembly is poorly understood and the energy transfer routes within PBSs are not well defined. Although 3D structures of some individual PBPs have been reported (reviewed in refs 2, 3), the structures of most linker proteins are unknown and the complete structure of a PBS has not been published, to our knowledge. Here we report the cryo-electron microscopy (cryo-EM) structure of a PBS from the red alga *Griffithsia pacifica* at a resolution of 3.5 Å, which reveals details of the PBS architecture.

Overall structure

The PBS from *G. pacifica* was purified and its intactness confirmed by its protein composition and spectroscopic features (Extended Data Fig. 1a–f). We reconstructed a 3D structure of the intact PBS by single particle cryo-EM with an overall resolution of 3.5 Å (Extended Data Fig. 2a–e and Extended Data Table 1). Applying individual local masks improved the resolutions of local maps to 3.4–4.3 Å (Extended Data Fig. 2g). The PBS is one of the largest supramolecular complexes that has been reported, with a calculated molecular mass of approximately 16.8 megadaltons. The overall appearance is block-type[9] with two-fold symmetry oriented perpendicularly to the thylakoid membranes (Fig. 1a–c and Extended Data Figs 1g–l, 3a–d). This PBS is larger than the hemiellipsoidal PBS isolated from *Porphyridium cruentum*[8] and has dimensions of approximately 680 Å length, 390 Å height, and 450 Å thickness (Fig. 1a–c).

The PBS contains a triangular core with the top cylinder B (formed by two APC trimers) sitting above two basal cylinders A and A′ (each formed by three APC trimers) surrounded by peripheral rods arranged in a staggered fashion (Fig. 1a–c and Extended Data Fig. 3c–e). In addition to the core and rods, there are other $(\alpha\beta)_6$ hexamers as well as individual α-subunits and β-subunits that are interspersed throughout the whole PBS (Extended Data Fig. 3a). The 3.5 Å cryo-EM map allowed us to build the atomic model with side chains clearly visible for most residues from almost all protein components (Supplementary Table 1). In total, we modelled 48 APC subunits in the core, including ApcF[24], L_{CM} and ApcD, 528 phycoerythrin subunits and 72 phycocyanin subunits in the rods, and 144 phycoerythrin subunits in the separate hexamers, α-subunits, and β-subunits. In addition, 52 rod linker proteins (L_R), 16 rod–core linker proteins (L_{RC}), 2 L_{CM} proteins and 2 core linker proteins (L_C) were modelled and assigned. The fourteen peripheral rods can be divided into two types on the basis of their PBP composition. The first type consists of three phycoerythrin hexamers at the part of the rod distal to the core and one phycocyanin hexamer proximal to the core. The second type contains only phycoerythrin proteins (Fig. 1d). We defined a total of 2,048 bilins with 48 phycocyanobilins (PCBs) in the core and the rest of the bilins in the rods (Fig. 1e). The presence of extra peripheral rods and separated hexamers composed of phycoerythrin expands the cross-section of light absorption[25] and results in the block shape of the PBS. If these rods and hexamers were removed, the PBS from *G. pacifica* (Extended

[1]State Key Laboratory of Membrane Biology, Beijing Advanced Innovation Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China. [2]Yantai Institute of Coast Zone Research, Chinese Academy of Sciences, Yantai 264003, China. [3]State Key Laboratory of Protein and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing 100871, China. [4]Key Laboratory of Phycology of CAS, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei 430072, China. †Present address: MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK.
*These authors contributed equally to this work.

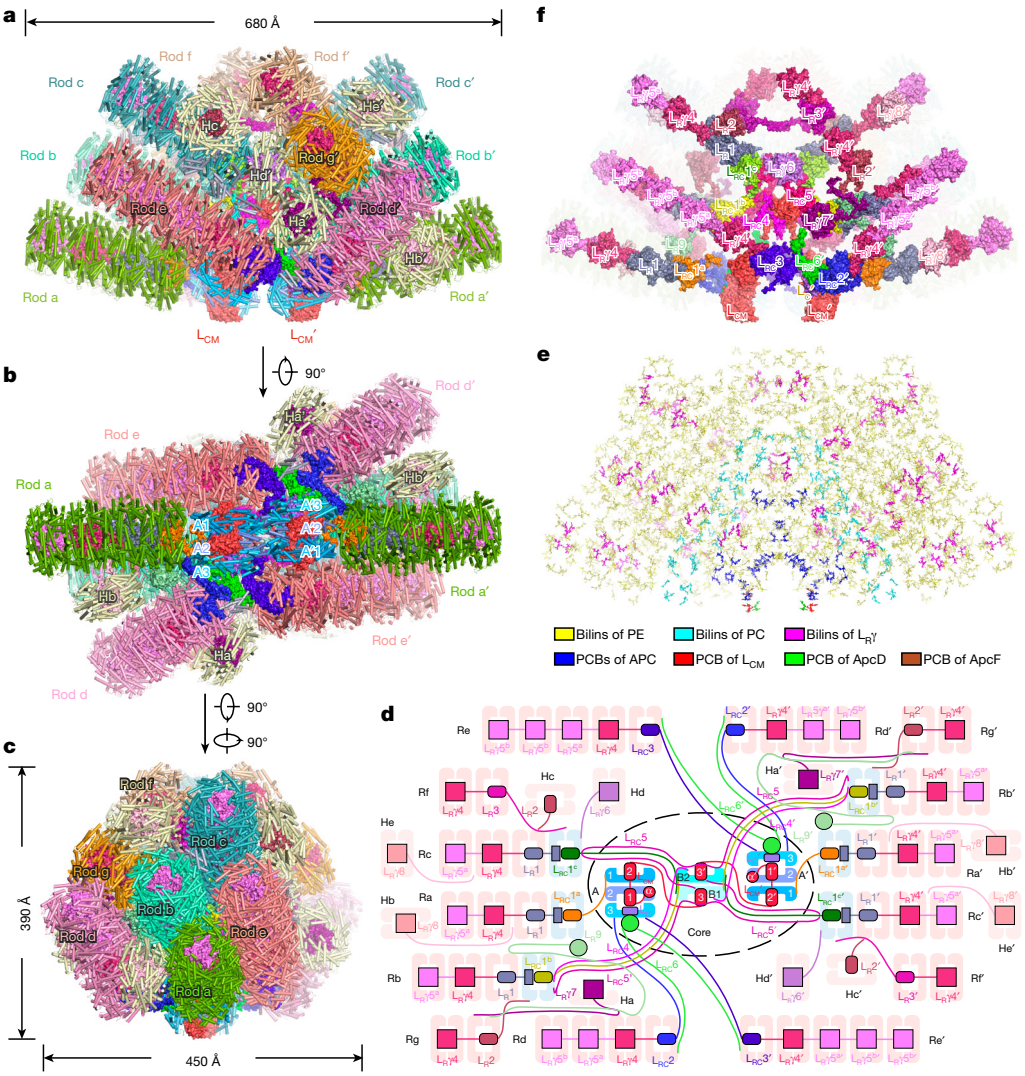


Figure 1 | Overall architecture of the PBS from *G. pacifica*. **a**–**c**, Structure of the PBS from three perpendicular views. PBPs are shown in cartoon representation, whereas linker proteins are shown as surface representations. **d**, Schematic model of the PBS architecture, showing the connections between PBS components. Light salmon, phycoerythrin hexamer; light blue, phycocyanin hexamer; large rectangular box, Pfam00427 domain; small rectangular box, Pfam01383 domain; square box, CBDγ. **e**, Bilin distribution in the PBS. All bilins are shown in the stick representation from the same view as in **a**. **f**, Structures of all linker proteins shown in the surface representation from the same view as in **a**. PE, phycoerythrin; PC, phycocyanin.

Data Fig. 3e) would be similar to the hemidiscoidal PBS[1,7], suggesting that the same principle of architecture for light absorption and energy transfer is adapted in both the hemidiscoidal and block-type PBSs.

Rod linker proteins

The linker proteins play very important roles in rod assembly[26–28]. We determined all linker proteins in the PBS and found that they form a skeleton for PBS assembly. One striking feature is that the L_R proteins assembled into skeletons by interacting with each other but with little participation of PBPs, as shown in Fig. 1f. The L_R proteins can be grouped into three classes on the basis of their structures (Extended Data Fig. 3f, g): the first class includes L_R1–L_R3 and contains the Pfam00427 domain; the second class possesses a previously unidentified, conserved chromophore-binding domain; and the third class has only one member (L_R9) and bears the FAS1 domain, which is critical for cell adhesion[29].

Figure 2 illustrates how a rod (Rb is used here) is assembled. The rod linker protein L_R1, ubiquitous in PBSs of cyanobacteria and red algae, contains a Pfam00427 domain in its N-terminal region (NTR) and a Pfam01383 domain in its C-terminal region (CTR) (Fig. 2a and Extended Data Figs 3f, g, 4a, b). The CTR of L_R1 possesses two long loops, the first between β-strands 1 and 2 and the second at the N-terminal extension (NTE) (Extended Data Fig. 4b). These two loops, especially the M326–K333 region, wrap around the surface of the adjacent rod–core linker protein $L_{RC}1^b$ (Fig. 2b), and also interact with the trimer Rb1II (Fig. 2c and Extended Data Fig. 5a, b). The NTR of L_R1 contains two extensions—the N-terminal extension (NTE) and the C-terminal extension (CTE) (Fig. 2a). Spatially, they are intertwined at the core–proximal side of the Pfam00427 domain (Fig. 2a). As shown in Fig. 2a, one α-helix from the NTE and two α-helices plus the C-terminal long loop from the CTE contact the α-helices F and F′ of the three β-subunits in the trimer Rb2I (Fig. 2d and Extended Data Fig. 5c–e). The Pfam00427 domain directly contacts the next rod linker protein, $L_R\gamma4$, by polar contacts and hydrogen bonds (Fig. 2a, e). It also interacts extensively with the trimers Rb2I and Rb2II (Fig. 2f, g and Extended Data Fig. 5f–h). Thus L_R1 makes a key contribution to rod formation by forming an $L_{RC}1^b$–L_R1–$L_R\gamma4$ connection and by making specific interactions with three trimers: Rb1II, Rb2I and Rb2II. The linkers L_R2 and L_R3 also contain N-terminal Pfam00427 domains located in the cavities of hexamers but have no C-terminal Pfam01383 domains (Extended Data Fig. 3f, g). Their C termini stretch out from the hexamers and contact with the neighbouring rods. As a result, rod Rg is associated with Rb and rod Rf with Rc (Fig. 1d and Extended Data Fig. 4c, d).

The next class of linkers has been previously described as the γ-subunits of phycoerythrins[1,30,31] and are therefore labelled as $L_R\gamma$ linkers ($L_R\gamma4$–$L_R\gamma8$; Extended Data Fig. 6a). The structure of $L_R\gamma$ proteins is distinguished by a conserved chromophore binding domain (labelled CBDγ) containing about 210 residues (Fig. 2a, h and Extended Data Figs 4e–h, 6a). The CBDγ consists of ten α-helices (H1–H10), which are organized in such a way that the five N-terminal helices (H1–H5) are related to the C-terminal five helices (H6–H10) by a 180°

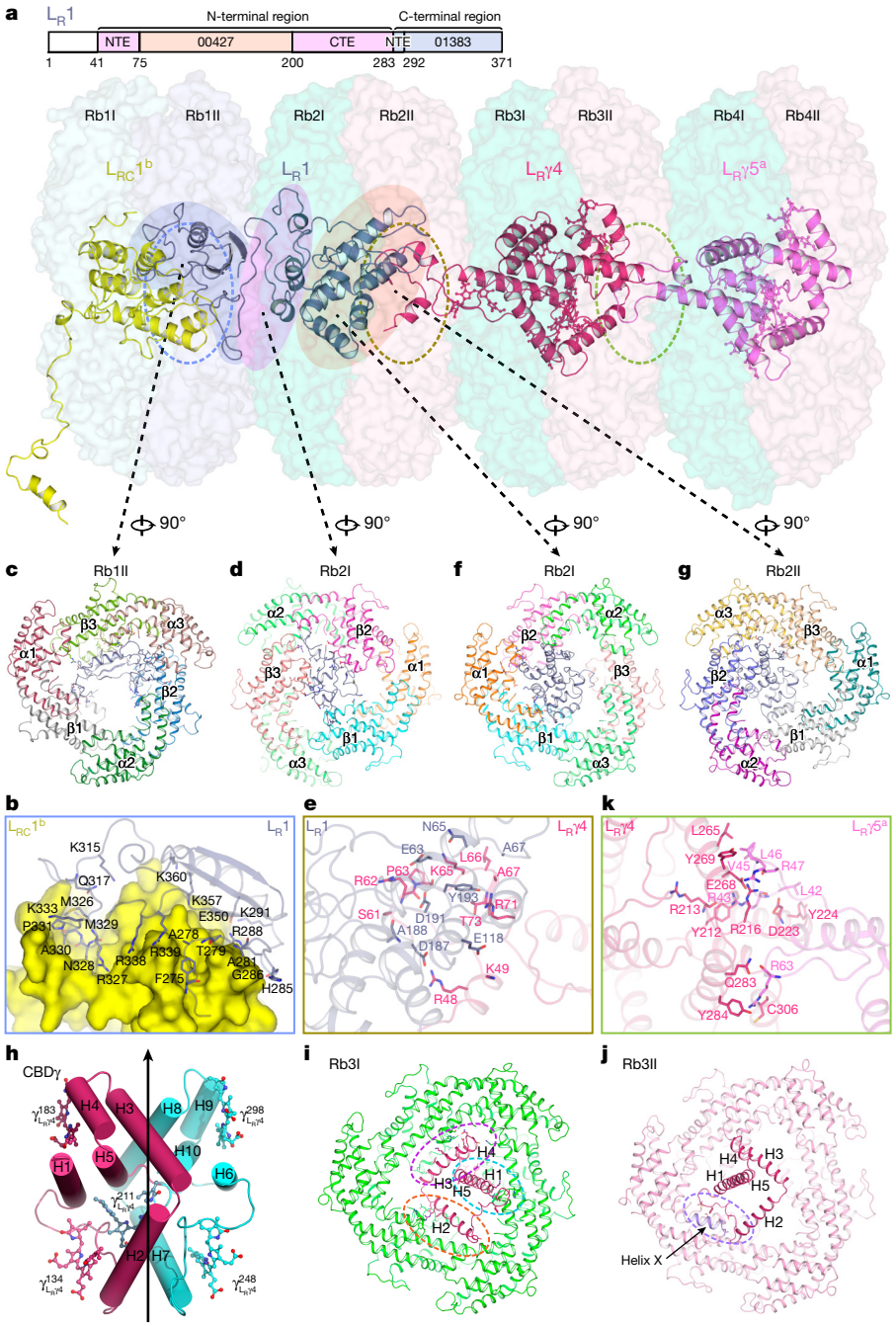


Figure 2 | Rod linker proteins in Rb. **a**, Overall structures of the linker proteins $L_{RC}1^b$, L_R1, $L_R\gamma4$ and $L_R\gamma5^a$ shown in cartoon representation. The bilins of $L_R\gamma4$ and $L_R\gamma5^a$ are shown in ball-stick representation and the hexamers of Rb are shown in surface representation. The phycocyanin hexamer (Rb1) and the phycoerythrin hexamers (Rb2–4) are coloured differently. A representative diagram of structural elements of L_R1 is shown above the structure. **b**, Interaction between $L_{RC}1^b$ and L_R1. Residues involved in the interaction are shown in stick representation and labelled. **c**, Interaction between the C-terminal region of L_R1 and trimer Rb1II. **d**, Interaction between the intertwined NTE and CTE of the N-terminal region of L_R1 and trimer Rb2I. **e**, Interaction between L_R1 and $L_R\gamma4$. **f**, Interaction between the Pfam00427 domain of L_R1 and trimer Rb2I. **g**, Interaction between the Pfam00427 domain of L_R1 and trimer Rb2II. **h**, Structure of the $L_R\gamma4$ CBDγ. The two repeats are coloured hot pink and cyan. Bilins are shown in ball-stick representation. All helices (H1–H10) are labelled. **i**, **j**, Interactions between the $L_R\gamma4$ CBDγ and trimer Rb3I (**i**), and between the $L_R\gamma4$ CBDγ and trimer Rb3II (**j**) with contacting interface areas circled. **k**, Interaction between $L_R\gamma4$ and $L_R\gamma5^a$.

rotation around an axis parallel to the hexamer plane (Fig. 2a, h). The CBDγs of most $L_R\gamma$ proteins bear five bilins, with each half binding two bilins. The remaining bilin is located at the interface between the two halves (Fig. 2h). The exception is the CBDγ of $L_R\gamma7$, which contains only four bilins. As a linker protein, $L_R\gamma4$ has three interactive functions. First, the CBDγ makes extensive contacts in a symmetrical fashion with the inner face of the rod hexamer (Extended Data Fig. 5b). Helices H1, H2, and H3/H4 contact the three inner sides of trimer Rb3I: helices X from the α-subunits and F and F′ from the β-subunits (Fig. 2i and Extended Data Fig. 5i–k). The loop separating helices H1 and H2 interacts mainly with helix X from the α-subunit of trimer Rb3II (Fig. 2j and Extended Data Fig. 5l). Second, the N terminus of $L_R\gamma4$ contains a long loop that folds back to the CBDγ and bears an additional bilin (Fig. 2a and Extended Data Fig. 3f). Two short helices following this loop insert into the neighbouring hexamer (Rb2) and interact with the Pfam00427 domain of the adjacent L_R1 (Fig. 2a, e). Third, the CBDγ of $L_R\gamma4$ is contacted by the extended loop at the N-terminus of $L_R\gamma5^a$ (one of the two $L_R\gamma5$ forms; see Extended Data Fig. 3f) that is located in the neighbouring hexamer (Rb4) (Fig. 2a, k). Thus, a common theme of peripheral rod assembly is that the linker proteins contain a rigid domain that occupies the central cavity of a hexamer[3] and interacts with extended helices or loops from core–distal neighbour linkers, leading to the formation of the rod linker skeleton.

The occurrence of multiple phycoerythrobilins (PEBs) and phycourobilins (PUBs)[30–32] on $L_R\gamma$ proteins is an evolutionary adaptation to habitats with low light, because they increase the cross-section of the antennae without expanding PBS spacing. It is also possible that the overall rate of energy transfer in peripheral rods could be increased by the presence of the bilins on the linker proteins owing to the shorter distances between bilins (Extended Data Fig. 6b–f).

Linker L_R9 is unique in that it has a rigid FAS1 domain with a very long extension at the C terminus. It sits between the rods and may function as a 'glue' molecule, holding the peripheral rods that it contacts together (Extended Data Fig. 4i, j).

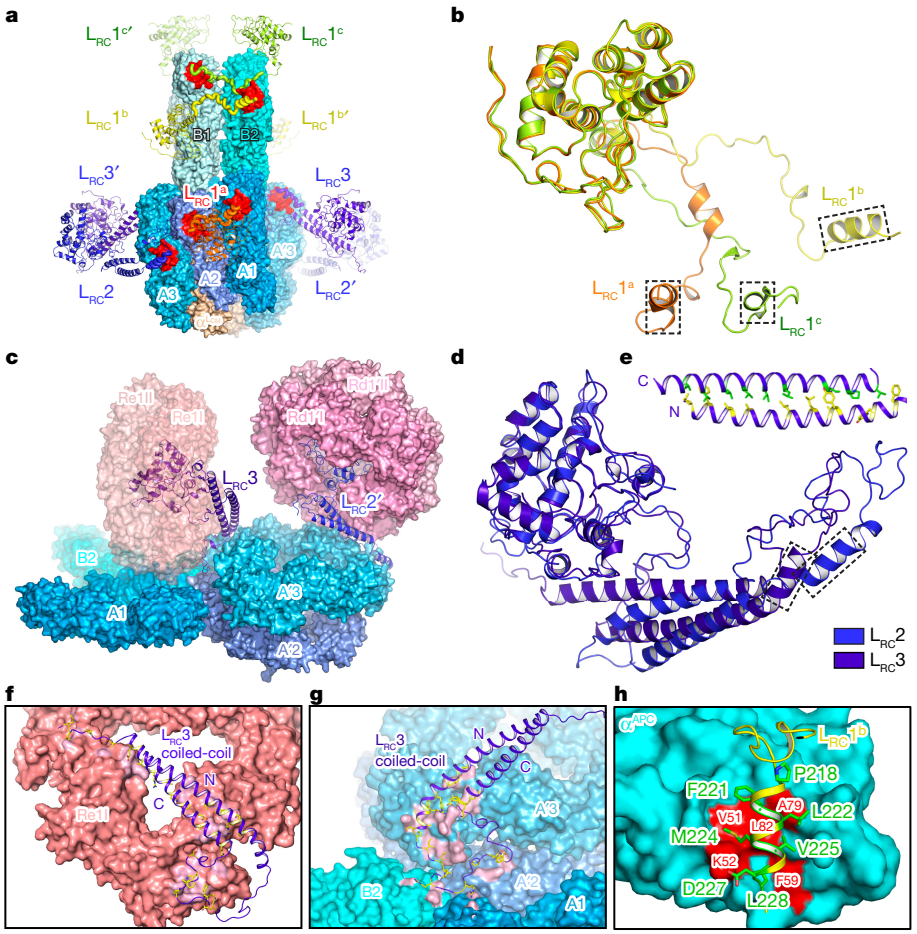


Figure 3 | Rod–core linker proteins. **a**, Organization of L_{RC} proteins $L_{RC}1$–3/$L_{RC}1'$–$3'$ and the core. The grooves on the α-subunits that contact the linker helices are shown in red. **b**, Structural similarity and differences among $L_{RC}1^a$, $L_{RC}1^b$ and $L_{RC}1^c$. These rod–core linkers are superimposed relative to the Pfam00427 domain. The helices interacting with the core are boxed. **c**, Organization of $L_{RC}2'$ (in Rd1′), $L_{RC}3$ (in Re1) and the core. **d**, Structural similarity of $L_{RC}2$ and $L_{RC}3$, as demonstrated by superimposition of the Pfam00427 domain at the N termini and the coiled-coil motif at the C termini. The helices interacting with the core are boxed. **e**, Hydrophobic interaction between the two helices of the coiled-coil motif. The residues (stick representation) involved in the interaction are coloured yellow (the N-terminal helix) and green (the C-terminal helix). **f**, **g**, Interactions between the coiled-coil motif of $L_{RC}3$ and trimer Re1I (**f**), and the core (**g**). The residues of $L_{RC}3$ involved in the interaction (stick representation) are coloured yellow. The trimer and core are shown in surface representation and the residues involved in the interaction are coloured light pink. **h**, Interaction between the C-terminal helix of $L_{RC}1^b$ and the α^{APC} of core layer B2. The residues of $L_{RC}1^b$ involved in the interaction (stick representation) are coloured green. α^{APC} are shown in surface representation and the residues involved in the interaction are coloured red.

Rod–core linker proteins

Rods Ra, Rb and Rc interact directly with the core through phycocyanin hexamers, while rods Rd and Re do not have a core–proximal phycocyanin hexamer and connect with the core by phycoerythrin hexamers (Fig. 1d). These two types of rod use different rod–core linkers to associate with the core.

Rods Ra, Rb and Rc use the linker $L_{RC}1$ to attach to the core (Fig. 3a). $L_{RC}1$ contains an N-terminal Pfam00427 domain buried in its phycocyanin hexamers and a C-terminal extension containing two separated helices and interacting with the core cylinders (Fig. 3a, b and Extended Data Fig. 3f). $L_{RC}1^a$ contacts layers A1 and A2; $L_{RC}1^b$ and $L_{RC}1^c$ both contact with the layers B1 and B2 (Fig. 3a). The different positions of the three $L_{RC}1$ proteins lead to different angles of the C-terminal extensions relative to the N-terminal domains, as illustrated by superimposing their structures (Fig. 3b).

Rods Rd and Re use linkers $L_{RC}2$ and $L_{RC}3$, respectively, in association with the core (Fig. 3c). Both $L_{RC}2$ and $L_{RC}3$ have N-terminal Pfam00427 domains located in the central cavity of phycoerythrin hexamers and a coiled-coil motif at the C termini protruding out from the hexamers (Fig. 3c, d and Extended Data Fig. 3f, g). The two helices in the coiled-coil motif are anti-parallel, and bound tightly together through an extensive hydrophobic interaction (Fig. 3e). In $L_{RC}3$, these two helices and the extended loop separating them make extensive interactions with the rod phycoerythrin trimer (Re1I) (Fig. 3f) and with the layers A1, A′2, A′3 and B′ of the core (Fig. 3g). A similar mechanism is used by $L_{RC}2$ in the interaction of rod Rd with the core.

Examination of the interactions between $L_{RC}1$–3 and the core components suggests that they all use helices to interact with the α-subunits of the core APC (Fig. 3a). $L_{RC}1$ proteins use the helices at their C termini (Fig. 3b) whilst the interactive helices of $L_{RC}2$ and $L_{RC}3$ are located at the C termini of the N-terminal helices in the coiled-coil motifs (Fig. 3d). When the α-subunits with which these L_{RC} proteins are associated are superimposed, the contacting helices are also aligned very well (Extended Data Fig. 7a). Indeed, these helices all contact with the grooves formed by helices B and E of the α-subunits via extensive hydrophobic interactions and some electrostatic interactions (Fig. 3h and Extended Data Fig. 7a–c). The residues of the L_{RC} proteins and the α-subunits involved in the interactions are highly conserved as hydrophobic or charged and/or polar amino acids throughout the red algae and cyanobacteria (Extended Data Fig. 7d, e). These results show that the rod–core linker proteins use a common mechanism when attaching rods to the core during PBS assembly.

The next group of proteins that function in the linking of the rods to the core, namely $L_{RC}4$, $L_{RC}5$ and $L_{RC}6$, were previously unknown and their structure is greatly different from that of the other linkers (Extended Data Figs 3f, 8). These proteins share a similar conformation distinguished by a structural element in the middle and extensions at both sides (Extended Data Figs 3f, g, 8). The structural element comprises a long α-helix in $L_{RC}4$ and $L_{RC}5$ and a FAS1[29,33] domain in $L_{RC}6$ (Extended Data Figs 3f, g, 8). The long α-helix of $L_{RC}4$ and $L_{RC}5$ spans one α-subunit of the core trimer (Extended Data Fig. 8b, c) and both extensions cover a few more proteins from both the core and the rods (Extended Data Fig. 8a, d–f). The FAS1 domain of $L_{RC}6'$ is inserted into the space between trimers Rd1′I and Re1I and trimer A′3, and together with both extensions makes wide contacts with the surrounding proteins (Extended Data Fig. 8i). Together, the revealed structures suggest that $L_{RC}4$–6 function as linkers by anchoring themselves to the core via their middle structural elements and using the extensions as ropes to maintain the stability of the assembled complex.

Core structures

The central core of the PBS is different from that of a typical hemidiscoidal PBS in that each of the two basal cylinders (A and A′) consists of only three APC trimers, with trimers A2 and A3 stacked face

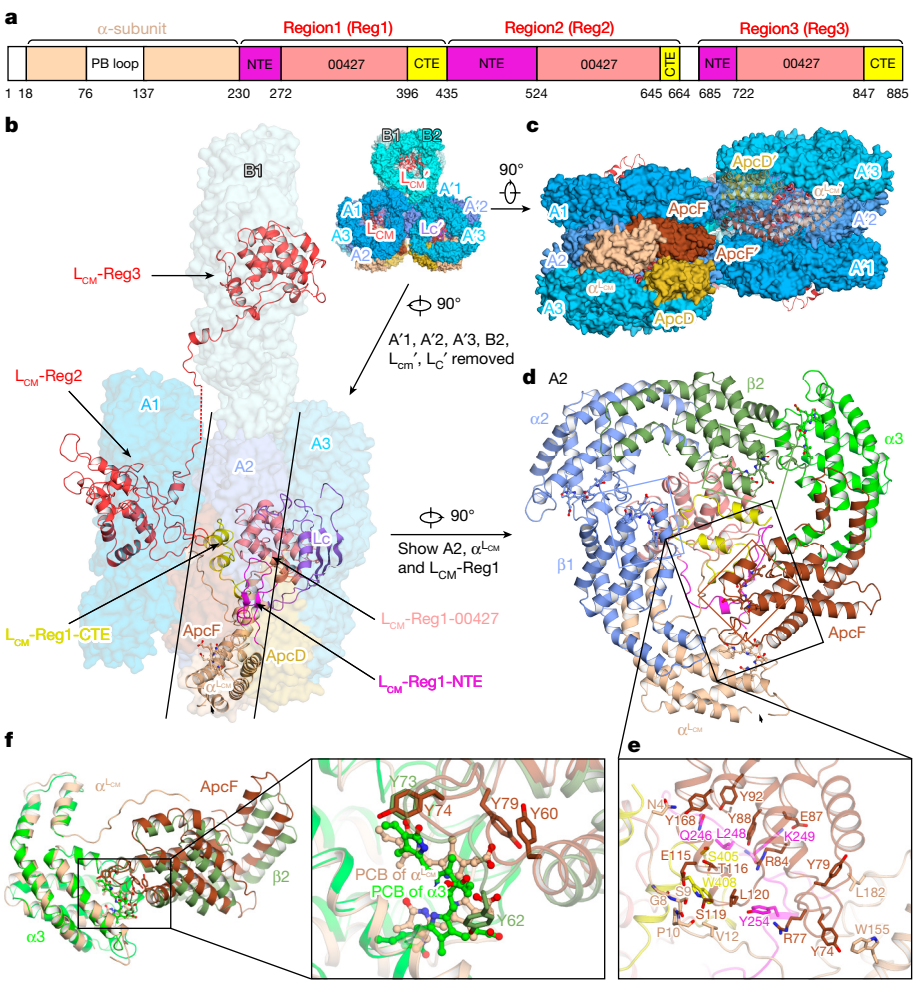


Figure 4 | Core structure. a, Diagram of the structural elements of L_{CM}. **b**, Half core structure, including L_{CM}, L_C, and the core layers A1, A2, A3 and B1. The PB loop in the α^{LCM} (black triangle) and the region between L_{CM}–Reg2 and L_{CM}–Reg3 (red dotted line) are not observed in the structure. **c**, Bottom view of the core. **d**, Structure of core layer A2 and L_{CM}–Reg1. The three PCB pockets of the β-subunits are boxed by dashed lines. **e**, Details of the interaction between L_{CM}–Reg1 and ApcF. **f**, Structural alignment of α^{LCM} with ApcF and the α3 subunit with the β2 subunit. The alignment was performed using α^{LCM} and the α3 subunit. Right, close-up view of the PCB pockets of α^{LCM} and α3. The residues from ApcF and β2 within 5 Å of the PCBs are shown in stick representation. Bilins are shown in ball-stick representation.

to face and the third trimer (A1) attaching to the A2 trimer in a back-to-back manner (Extended Data Fig. 9a, b). These two basal cylinders are arranged in a staggered antiparallel fashion (Fig. 4c and Extended Data Fig. 9a). As a result, two side-by-side planes are formed for attachment of the top cylinder, which contains only two APC trimers (B1 and B2) stacked back to back (Extended Data Fig. 9a). The interaction of the core trimers is symmetrical: trimers B1, A2 and A′1 interact with one another and trimers B2, A′2 and A1 interact with one another (Extended Data Fig. 9c). This core structure is likely to be evolutionarily derived from the core of hemidiscoidal PBS by the elimination of the exterior trimers of all three cylinders except for those with terminal emitters. The formation of a more compact core structure leads to the loss of 24 PCBs in the core and it could be an adaptation to habitats in which the red light absorbed by PCBs is limited.

We have determined an almost complete structure of L_{CM} (Fig. 4a, b). Starting from the N terminus, L_{CM} contains one α^{LCM} (the α domain in L_{CM}) positioned at the bottom of trimeric layer A2, followed by three successive regions, Reg1–3, located at the centres of layers A2, A1 and B1, respectively (Fig. 4a, b). The structure of each region is similar to that of the NTR of L_R1, containing two parts: a Pfam00427 domain and a part that is composed of the NTE and CTE (Fig. 4b and Extended Data Fig. 9d). The structures of the Pfam00427 domains from the three regions are all aligned well with the crystal structure of the Pfam00427 domain, whereas the NTEs and CTEs are largely variable in conformation (Fig. 4b and Extended Data Fig. 9d). They interact with the surrounding APC subunits and could provide distinct environments for the PCB in the core (Extended Data Fig. 9e–g). The two regions of L_{CM} that are not observed are the loops located within the α^{LCM} and between Reg2 and Reg3 (Fig. 4b). The loops in the α^{LCM} are near the bottom extrusions of trimer A2 (Fig. 4b) and we cannot rule out the possibility that this loop has some significant function. In addition to the structure of L_{CM}, we have also determined the structure of L_C, which contains an N-terminal loop region and a C-terminal Pfam01383 domain (Extended Data Fig. 3f, g). L_C is located in trimer A3 (Fig. 4b) and interacts with the surrounding β-subunits, as previously reported[27]. It also interacts with L_{CM}-Reg1 in a manner similar to the interaction between the CTR of L_R1 and $L_{RC}1$[b].

Three APC variants in the core, α^{LCM}, ApcF and ApcD, are clearly identified (Extended Data Fig. 10). Whilst the overall structure of α^{LCM} is very similar to that of recombinant α^{LCM} (ref. 16), α^{LCM} is more strongly influenced by ApcF, whose function is critical to energy transfer from PBS to photosystem II (PSII)[34]. ApcF is located in the same trimer as α^{LCM} but in a different αβ monomer from α^{LCM} and close to the interface of the two basal cylinders (Fig. 4b–d). When ApcF is superimposed with ApcB, most of the structures align well except that ApcF has an extension at the tip of the G–H helix hairpin (Extended Data Fig. 9h). This extension protrudes out of the trimer and could be important to the functions of PBSs. L_{CM} makes extensive contacts with ApcF: α^{LCM} and the CTE of L_{CM}–Reg1 interact with helix F′ and the N-terminal loop of helix E of ApcF, and the NTE of L_{CM}–Reg1 interacts with helix E of ApcF (Fig. 4e). Most importantly, the PCB of α^{LCM} is very close to the adjacent ApcF subunit (Fig. 4d, f). Two aromatic residues from ApcF, Tyr60 and Tyr79, are present within 5 Å of the PCB in α^{LCM} and one residue corresponding to the aromatic amino acid Tyr62 from the β2-subunit is missing when compared with the PCB pocket of the other α-subunits (Fig. 4f and Extended Data Fig. 9i, j). The interaction between α^{LCM} and ApcF may strongly affect the spectroscopic property of the PCB in α^{LCM} (ref. 34) and play a key role in energy transfer from α^{LCM} to PSII[22].

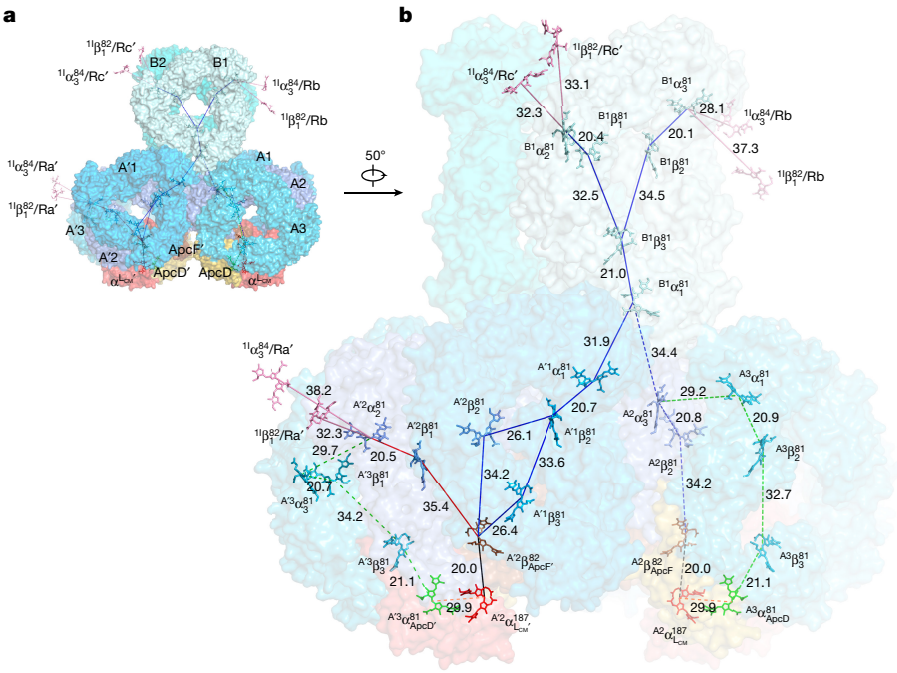


Figure 5 | Positions of bilins that are key to energy transfer pathways in the core. a, Face-view of the core with all layers labelled. The bilins of the three peripheral rods (Ra′, Rb and Rc′) that are closest to the core are shown as individual molecules. **b**, Side view of the core with possible energy transfer pathways. Bilins in core layers A′1, A2 (A′2), A3 (A′3) and B are shown in different colours. Other colour codes used: red, bilin in α^{LCM} ($\alpha^{LCM'}$); green, bilin in ApcD (ApcD′); brown, bilin in ApcF (ApcF′); pink, the bilins in phycocyanin hexamers nearest to the core. Blue lines, two possible pathways from the bilins nearest to the core in the phycocyanin hexamers of Rb and Rc′ to the bilin of ApcF (dashed line) or the bilin of ApcF′ (solid line); red line, a pathway from the bilins nearest to the core in the phycocyanin hexamers of Ra′ to the bilin of ApcF′; black lines, pathways from the bilin of ApcF to the bilin of α^{LCM} (dashed line) or from the bilin of ApcF′ to the bilin of $\alpha^{LCM'}$ (solid line); green dashed lines, the two possible alternative pathways to the bilins of ApcD or ApcD′; orange lines, possible pathways from the bilin of α^{LCM} ($\alpha^{LCM'}$) to the bilin of ApcD (ApcD′). The numbers near the lines indicate the distances (Å) between the bilin pairs.

α^{LCM} and ApcD are both located at the bottom of the core and are present in opposite trimers of the same hexamer; α^{LCM} resides in trimer A2 and ApcD in trimer A3 (Fig. 4b, c). The structure of ApcD in our model is identical to the structure of the recombinant ApcD/ApcB trimer[18]. These results are consistent with the functions of L_{CM} and ApcD in the transfer of energy to the reaction centres (Fig. 4b, c) and should shed light on the mechanisms of state transitions that regulate light energy distribution between two photosystems[35–38].

Plausible energy transfer pathway

Plausible energy transfer routes could be predicted from the types of bilins and distances among the chromophores, as revealed in this study, because energy transfer efficiency is inversely related to the sixth power of the distance between the chromophores. Several pairs of bilins, having the shortest distances between rods and the core, are probably key bridges in energy transfer from rods to core. They are the following pairs: $^{1I}\alpha_3^{84}$/Ra′ and $^{A'2}\alpha_2^{81}$ (38.2 Å), $^{1I}\beta_1^{82}$/Ra′ and $^{A'2}\alpha_2^{81}$ (32.3 Å) for Ra; $^{1I}\alpha_3^{84}$/Rb and $^{B1}\alpha_3^{81}$ (28.1 Å), $^{1I}\beta_1^{82}$/Rb and $^{B1}\alpha_3^{81}$ (37.3 Å) for Rb; and $^{1I}\alpha_3^{84}$/Rc′ and $^{B1}\alpha_2^{81}$ (32.3 Å), $^{1I}\beta_1^{82}$/Rc′ and $^{B1}\alpha_2^{81}$ (33.1 Å) for Rc (Fig. 5, light pink solid line). For rods Rb and Rc, the energy flows into trimer B1 and is transferred to $^{B1}\alpha_1^{81}$, the nearest bilin to the basal cylinders. From $^{B1}\alpha_1^{81}$ the energy may flow either to the layer A′1 bilin $^{A'1}\alpha_1^{81}$ (31.9 Å) or to the layer A2 bilin $^{A2}\alpha_3^{81}$ (34.4 Å) (Fig. 5b). Both distances could allow energy transfer through the side-to-side interfaces of layers B1 and A′1, or B1 and A2 (Fig. 5b). Energy arriving at layer A′1 will have to travel across the back-to-back inter-trimer interface to layer A′2 (Fig. 5b). Energy from rod Ra travels to A′2 directly (Fig. 5b); upon arriving at either A2 or A′2, it will be transferred to the bilin on ApcF or ApcF′, and finally to the terminal emitter L_{CM} or L_{CM}' (Fig. 5b). The distance between bilins $^{A2}\beta_{ApcF}{}^{82}$ and $^{A2}\alpha_{LCM}{}^{187}$ ($^{A'2}\beta_{ApcF'}{}^{82}$ and $^{A'2}\alpha_{LCM'}{}^{187}$) is 20.0 Å, the shortest distance between bilin pairs in the core (Fig. 5b). This is consistent with the previous conclusion that ApcF plays a crucial role in energy transfer in PBS[34]. We also note a probable alternative energy transfer route after $^{A2}\alpha_3{}^{81}$ whereby energy passes across the face-to-face inner-trimer space (29.2 Å) between core layers A2 and A3[39], reaches bilin $^{A3}\alpha_1{}^{81}$, and eventually reaches the terminal emitter ApcD ($^{A3}\alpha_{ApcD}{}^{81}$) (Fig. 5b). Similarly, there is an alternative route alongside the pathway from Ra′ to $^{A'2}\alpha_{LCM'}{}^{187}$. Energy passes from Ra′ to $^{A'3}\alpha_{ApcD'}{}^{81}$ via the bilin pair of $^{A'2}\alpha_2{}^{81}$ and $^{A'3}\alpha_3{}^{81}$ (29.7 Å) (Fig. 5b).The distance between bilins of L_{CM} and ApcD is 29.9 Å (Fig. 5b), which also suggests that direct energy transfer from L_{CM} to ApcD could be possible. Functionally, the presence of these routes is an important feature because ApcD is responsible for the delivery of energy to PSI[21]. Understanding how energy flow to ApcD is regulated could provide insight into the mechanism that regulates light energy distribution between photosystems in state transitions[37].

The presence of individual phycoerythrin hexamers and rods Rd and Re in the PBS (Fig. 1) was an unexpected discovery, given the data from previously reported PBSs[3]. They do not make contact with the core and the energy they absorb would travel to the peripheral rods through a side-to-side interface[40] as exemplified by the distances of bilin pairs $^{HaI}\beta_1{}^{50/61}$ on Ha and $^{Rd2I}\beta_1{}^{158}$ on Rd2 (16.1 Å), and $^{HbI}\beta_1{}^{50/61}$ on Hb and $^{Ra2II}\alpha_1{}^{139}$ on Ra2 (12.7 Å) (Extended Data Fig. 11a). Rods Rd and Re contain no phycocyanin hexamers but are associated with the core directly through L_R2 and L_R3 (Fig. 3c). Light energy absorbed by these rods could be transferred to the core, although energy transfer from phycoerythrin to APC would have low efficiency owing to the effect of the small overlap integral between a PEB in phycoerythrin and a PCB in APC. This low efficiency could, however, be somewhat offset by the proximity of the bilin pairs $^{Re1'I}\alpha_1{}^{139}$ and $^{A'1}\alpha_1{}^{81}$ (19.1 Å), and $^{Rd1I}\alpha_1{}^{82}$ and $^{A3}\alpha_3{}^{81}$ (26.6 Å) (Extended Data Fig. 11b). Alternatively, side-to-side transfer from Rd to Rb or Re to Ra or Rc could occur. The distances between bilins of Rd to Rb and Re to Ra or Rc, such as $^{Rd1II}\beta_3{}^{158}$ to $^{Rb1I}\beta_3{}^{153}$, $^{Rd2II}\beta_3{}^{158}$ to $^{Rb2I}\beta_3{}^{158}$, $^{Re3II}\alpha_3{}^{139}$ to $^{Ra1II}\beta_3{}^{153}$, and $^{Re3II}\alpha_2{}^{139}$ to $^{Rc1II}\beta_3{}^{153}$ are 13.8 Å, 13.2 Å, 12.9 Å and 15.7 Å, respectively (Extended Data Fig. 11c, d).

Energy transfer in PBSs is very fast with a high quantum yield[41,42]. The determination of the 3D structure of the PBS from *G. pacifica* will provide a firm basis for understanding what roles the proteins' environments, such as the covalent linkages of the chromophores in the binding pockets and the presence of linker proteins[43], play in efficient energy transfer in PBSs.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 31 May; accepted 11 September 2017.
Published online 18 October 2017.

1. Sidler, W. A. in *The Molecular Biology of Cyanobacteria* (ed. Bryant, D. A.) 139–216 (Kluwer, 1994).

2. Adir, N. Elucidation of the molecular structures of components of the phycobilisome: reconstructing a giant. *Photosynth. Res.* **85,** 15–32 (2005).
3. Watanabe, M. & Ikeuchi, M. Phycobilisome: architecture of a light-harvesting supercomplex. *Photosynth. Res.* **116,** 265–276 (2013).
4. Croce, R. & van Amerongen, H. Natural strategies for photosynthetic light harvesting. *Nat. Chem. Biol.* **10,** 492–501 (2014).
5. Schirmer, T. *et al.* Crystal structure analysis and refinement at 2.5 Å of hexameric C-phycocyanin from the cyanobacterium *Agmenellum quadruplicatum*. The molecular model and its implications for light-harvesting. *J. Mol. Biol.* **188,** 651–676 (1986).
6. de Marsac, N. T. & Cohen-bazire, G. Molecular composition of cyanobacterial phycobilisomes. *Proc. Natl Acad. Sci. USA* **74,** 1635–1639 (1977).
7. Arteni, A. A., Ajlani, G. & Boekema, E. J. Structural organisation of phycobilisomes from *Synechocystis* sp. strain PCC6803 and their interaction with the membrane. *Biochim. Biophys. Acta* **1787,** 272–279 (2009).
8. Arteni, A. A. *et al.* Structure and organization of phycobilisomes on membranes of the red alga *Porphyridium cruentum*. *Photosynth. Res.* **95,** 169–174 (2008).
9. Gantt, E. & Lipschultz, C. A. Structure and phycobiliprotein composition of phycobilisomes from *Griffithsia pacifica* (Rhodophyceae). *J. Phycol.* **16,** 394–398 (1980).
10. Guglielmi, G., Cohen-Bazire, G. & Bryant, D. A. The structure of *Gloeobacter violaceus* and its phycobilisomes. *Arch. Microbiol.* **129,** 181–189 (1981).
11. Williams, R. C., Gingrich, J. C. & Glazer, A. N. Cyanobacterial phycobilisomes. Particles from *Synechocystis* 6701 and two pigment mutants. *J. Cell Biol.* **85,** 558–566 (1980).
12. Ducret, A., Sidler, W., Wehrli, E., Frank, G. & Zuber, H. Isolation, characterization and electron microscopy analysis of a hemidiscoidal phycobilisome type from the cyanobacterium *Anabaena* sp. PCC 7120. *FEBS* **236,** 1010–1024 (1996).
13. MacColl, R. Allophycocyanin and energy transfer. *Biochim. Biophys. Acta* **1657,** 73–81 (2004).
14. Gao, X. *et al.* Molecular insights into the terminal energy acceptor in cyanobacterial phycobilisome. *Mol. Microbiol*. **85,** 907–915 (2012).
15. 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).
16. 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).
17. Glazer, A. N. & Bryant, D. A. Allophycocyanin B (λ_{max} 671, 618 nm): a new cyanobacterial phycobiliprotein. *Arch. Microbiol.* **104,** 15–22 (1975).
18. Peng, P. P. *et al.* The structure of allophycocyanin B from *Synechocystis* PCC 6803 reveals the structural basis for the extreme redshift of the terminal emitter in phycobilisomes. *Acta Crystallogr. D Biol. Crystallogr*. **70,** 2558–2569 (2014).
19. Gantt, E. Structure and function of phycobilisomes: light harvesting pigment complexes in red and blue-green algae. *Int. Rev. Cytol*. **66,** 45–80 (1980).
20. Adir, N., Dines, M., Klartag, M., McGregor, A. & Melamed-Frank, M. in *Complex Intracellular Structures in Prokaryotes* (ed. Schwender, J.) 1–31 (Plant Metabolic Networks, 2009).
21. 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).
22. Ashby, M. K. & Mullineaux, C. W. The role of ApcD and ApcF in energy transfer from phycobilisomes to PSI and PSII in a cyanobacterium. *Photosynth. Res*. **61,** 169–179 (1999).
23. Chang, L. *et al.* Structural organization of an intact phycobilisome and its association with photosystem II. *Cell Res.* **25,** 726–737 (2015).
24. Yamanaka, G., Lundell, D. J. & Glazer, A. N. Molecular architecture of a light-harvesting antenna. Isolation and characterization of phycobilisome subassembly particles. *J. Biol. Chem*. **257,** 4077–4086 (1982).
25. Scholes, G. D., Mirkovic, T., Turner, D. B., Fassioli, F. & Buchleitner, A. Solar light harvesting by energy transfer: from ecology to coherence. *Energy Environ. Sci.* **5,** 9374–9393 (2012).
26. Liu, L.-N., Chen, X.-L., Zhang, Y.-Z. & Zhou, B.-C. Characterization, structure and function of linker polypeptides in phycobilisomes of cyanobacteria and red algae: an overview. *Biochim. Biophys. Acta* **1708,** 133–142 (2005).
27. Reuter, W., Wiegand, G., Huber, R. & Than, M. E. Structural analysis at 2.2 Å of orthorhombic crystals presents the asymmetry of the allophycocyanin-linker complex, AP.LC7.8, from phycobilisomes of *Mastigocladus laminosus*. *Proc. Natl Acad. Sci. USA* **96,** 1363–1368 (1999).
28. Gao, X. *et al.* Crystal structure of the N-terminal domain of linker L(R) and the assembly of cyanobacterial phycobilisome rods. *Mol. Microbiol*. **82,** 698–705 (2011).
29. Clout, N. J., Tisi, D. & Hohenester, E. Novel fold revealed by the structure of a FAS1 domain pair from the insect cell adhesion molecule fasciclin I. *Structure* **11,** 197–203 (2003).
30. Klotz, A. V. & Glazer, A. N. Characterization of the bilin attachment sites in R-phycoerythrin. *J. Biol. Chem.* **260,** 4856–4863 (1985).
31. Nagy, J. O., Bishop, J. E., Klotz, A. V., Glazer, A. N. & Rapoport, H. Bilin attachment sites in the alpha, beta, and gamma subunits of R-phycoerythrin. Structural studies on singly and doubly linked phycourobilins. *J. Biol. Chem*. **260,** 4864–4868 (1985).
32. Blot, N. *et al.* Phycourobilin in trichromatic phycocyanin from oceanic cyanobacteria is formed post-translationally by a phycoerythrobilin lyase-isomerase. *J. Biol. Chem.* **284,** 9290–9298 (2009).
33. Underhaug, J. *et al.* Mutation in transforming growth factor beta induced protein associated with granular corneal dystrophy type 1 reduces the proteolytic susceptibility through local structural stabilization. *Biochim. Biophys. Acta* **1834,** 2812–2822 (2013).
34. 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).
35. Bonaventura, C. & Myers, J. Fluorescence and oxygen evolution from *Chlorella pyrenoidosa*. *Biochim. Biophys. Acta* **189,** 366–383 (1969).
36. 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).
37. Kaňa, R. *et al.* Phycobilisome mobility and its role in the regulation of light harvesting in red algae. *Plant Physiol.* **165,** 1618–1631 (2014).
38. Liu, H. *et al.* Phycobilisomes supply excitations to both photosystems in a megacomplex in cyanobacteria. *Science* **342,** 1104–1107 (2013).
39. Ren, Y. *et al.* Clarifying and illustrating the electronic energy transfer pathways in trimeric and hexameric aggregation state of cyanobacteria allophycocyanin within the framework of Förster theory. *J. Comput. Chem.* **36,** 137–145 (2015).
40. Contreras-Martel, C. *et al.* The structure at 2 Å resolution of Phycocyanin from *Gracilaria chilensis* and the energy transfer network in a PC–PC complex. *Biophys. Chem.* **125,** 388–396 (2007).
41. Knox, R. S. Ultrashort processes and biology. *J. Photochem. Photobiol. B* **49,** 81–88 (1999).
42. Glazer, A. N. Light guides. Directional energy transfer in a photosynthetic antenna. *J. Biol. Chem*. **264,** 1–4 (1989).
43. Nganou, C., David, L., Adir, N. & Mkandawire, M. Linker proteins enable ultrafast excitation energy transfer in the phycobilisome antenna system of *Thermosynechococcus vulcanus*. *Photochem. Photobiol. Sci.* **15,** 31–44 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Lei, L. Cheng and the staff at the Tsinghua University Branch of the National Center for Protein Sciences Beijing for providing facility support; J. Wang for model building, structure refinement and model validation; J. Dai for recommending RNA sequencing; L. Li and T. Zhu for transcriptome annotation; P. Meng and F. Chen for genomic DNA sequencing and data mining; V. Ragunath, O. Kovalevskiy and G. Murshudov (MRC Laboratory of Molecular Biology) for structure refinement; N. Gao for model building; H. Lin, L. Chang and Y. Liu for discussion; H.-W. Wang and J. Wu for commenting on the manuscript; and the members of the Tsinghua National Laboratory for Information Science and Technology for providing the "Explorer 100" cluster system as a computational resource. S.-F.S. and J.Zhao were supported by the National Basic Research Program (2016YFA0501101, 2017YFA0504600 and 2015CB150100). S.-F.S. and S.S. were supported by the National Natural Science Foundation of China (31230016, 31370717 and 31670745 and 31400632 and 31670746). J. Zhao also received support from the CAS fund (QYZDY-SSW-SMC004).

Author Contributions S.-F.S. initiated and supervised the project; J.Zhang and J.M. prepared the samples; J.Zhang collected the EM data and performed the EM analysis and the model building; J.M. performed the biochemical and biophysical analyses; J.Zhang and D.L. refined the structure; S.S. and S.-F.S. analysed the structure; S.Q. contributed to the sample selection; J.Zhao participated in the analysis of biochemical data; and S.S., J.Zhao and S.-F.S. wrote the manuscript with help from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to S.-F.S. (suisf@mail.tsinghua.edu.cn) or S.S. (shansun@mail.tsinghua.edu.cn).

Reviewer Information *Nature* thanks H. van Amerongen, Z. Liu and C. Mullineaux for their contribution to the peer review of this work.

METHODS

PBS preparation. To obtain the high-resolution structure, we screened PBSs from several algal species for homogeneity. The PBS from *G. pacifica* exhibited a very compact and uniform structure, as indicated by the raw micrographs of the cryo-EM PBS samples and the 2D class averages of the cryo-EM particles (Extended Data Figs 1g, h, 2a, c).

Griffithsia pacifica (From UTEX Culture Collection of Algae) was cultured in Enrichment Seawater Medium (UTEX, https://utex.org/products/enrichment-solution-for-seawater-medium-recipe) bubbled with sterilizing filtered air at 22 °C, under a light–dark period of 16 h: 8 h, with illumination from cool-white fluorescent lamps at a light intensity of approximately 50 μmol photons m^{-2} h^{-1}. Algal tissue was suspended in Buffer A (0.65 M Na/K-PO_4 buffer with 0.5 M sucrose and 10 mM EDTA, pH 7.0) at 0.3 g of wet weight ml^{-1}. The tissue was homogenized at room temperature using a homogenizer and pestle, and was further ultrasonicated for 10 min in the presence of 1 mM phenylmethylsulfonyl fluoride (PMSF). Triton X-100 was added to the suspension at a final concentration of 2% v/v and incubated for 30–40 min at room temperature followed by centrifugation at 20,000*g* for 30 min at 18 °C. The middle aqueous violet solution was loaded onto a discontinuous sucrose gradient (2 ml of 0.5 M, 2 ml of 0.75 M, 2 ml of 1.0 M, 2 ml of 1.5 M, 1 ml of 2.0 M, all in Buffer B: 0.75 M K/$NaPO_4$ buffer with 10 mM EDTA, pH 7.0). The samples were centrifuged at 120,000*g* for 4 h at 18 °C using a SW41 rotor on a Beckman Coulter Optima XPN-100 centrifuge. Two visible bands were obtained after centrifugation, with band 1 as the main layer of intact PBSs (Extended Data Fig. 1a).

Absorption and fluorescence spectra. The absorption spectra between 300 and 800 nm were measured using an Ultrospec 2100 Pro ultraviolet-visible spectrophotometer (Biochrom Ltd). Fluorescence emission spectra between 500 and 800 nm (excitation 450 nm) were recorded with a Hitachi FL-4500 fluorescence spectrophotometer (Hitachi) at room temperature.

RNA isolation and sequencing, *de novo* assembly and gene annotation. Total RNA was extracted from *G. pacifica* using TRIzol Reagent (Invitrogen) in accordance with the manufacturer's protocol. The isolated RNA was treated with DNase I to degrade DNA in the sample, and the mRNA was enriched using oligo (dT) magnetic beads (NEBNext Poly(A) mRNA Magnetic Isolation Module, NEB). A library of total mRNA was constructed using the NEBNext Ultra™ Directional RNA Library Prep Kit for Illumina (NEB) and sequenced by Illumina HiSeq2500 using a PE125 strategy at the Bionova (Beijing) Biotechnology Co., Ltd. After sequencing, raw data were filtered to remove PCR adaptor reads and low quality reads, resulting in 6 GB clean data. *De novo* assembly was conducted by Trinity[44] software using the clean reads. Open reading frames (ORFs) were predicted using TransDecoder v2.0.1 based on the assembled transcripts and annotated according to Gene Ontology terms with the Blast2Go software v3.1[45] and through Blastp against the NCBI non-redundant (nr) protein database, Pfam database and Swiss-Prot database.

Genomic DNA isolation, sequencing and data mining. To obtain the full-length sequences of the protein components from PBS, we performed genomic DNA sequencing. *G. pacifica* genomic DNA was isolated using the Plant Genomic DNA Kit (TIANGEN). DNA samples were then sheared into smaller fragments. Sequencing libraries were generated using NEBNext Ultra DNA Library Prep Kit for Illumina (NEB) following the manufacturer's recommendations and index codes were added. Sequencing was then performed on an Illumina MiSeq platform (Illumina Inc.) to obtain pair-end 300-bp reads. After sequencing, the raw data were filtered to remove PCR adaptor reads and low quality reads by FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), and a total of 116,112,075 pair-end reads and 8 GB clean data were obtained. *De novo* assembly was conducted using SOAPdenovo2 software (https://sourceforge.net/projects/soapdenovo2/) using the clean reads, and 46,187 contigs were obtained with an N50 of 9,541 bp. ORFs were searched using GeneMark-ES software[46] based on the assembled contigs, and a total of 61,216 ORF sequences were predicted. The identified 25 PBS protein sequences from RNA sequencing (see below) were then searched against the ORF database to define the full-length sequences of proteins (Supplementary Table 2).

Mass spectrometric (MS) analysis. Gel bands of PBS separated by 12% Bis-Tris SDS–PAGE in MES buffer were excised for in-gel digestion and proteins were identified by mass spectrometry. Briefly, disulfide bonds of the proteins were reduced with 5 mM dithiothreitol (DTT) and alkylated with 10 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 before being redissolved in 0.1% trifluoroacetic acid. The peptides were separated using a 120 min gradient elution, at a flow rate of 0.300 μl min^{-1}, with a Thermo-Dionex Ultimate 3000 HPLC system, which was directly interfaced with a Thermo Orbitrap Fusion Lumos 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. An LTQ-Orbitrap mass spectrometer was operated in data-dependent acquisition mode using Xcalibur 4.0.27.10 software and there was a single full-scan mass spectrum in the Orbitrap (350–1500 *m/z*, 60,000 resolution) followed by 3 s data-dependent MS/MS scans in an Ion Routing Multipole at 40% normalized collision energy (HCD). MS/MS spectra from each LC–MS/MS run were searched against the transcriptome database of *G. pacifica* using Proteome Discoverer v1.4 searching algorithm. The search criteria were as follows: full tryptic specificity was required; two missed cleavages were allowed; carbamidomethylation (C) was set as fixed modification; oxidations (M) were set as variable modifications; precursor ion mass tolerance was 10 parts per million (p.p.m.) for all MS acquired in the Orbitrap mass analyser; and fragment ion mass tolerance was 0.8 Da for all MS2 spectra acquired in the LTQ. High confidence score filter (FDR <1%) was used to select the 'hit' peptides and their corresponding MS/MS spectra were manually inspected. For intact PBS complexes in solution, sample was precipitated with 15% trichloroacetic acid (TCA) and lyophilized. The pellet was dissolved in 8 M urea, 100 mM Tris, pH 8.5, followed by DTT reduction, iodoacetamide alkylation, and trypsin digestion. The subsequent analysis was performed using the same method as described above. Finally, all 25 protein components of PBS were identified in the samples (Extended Data Fig. 1d, f).

Identification of protein components in PBS. Two strategies were used to find the potential protein components in the PBS of *G. pacifica*. First, we performed RNA sequencing and searched the obtained transcriptome database of *G. pacifica* against the known protein sequences of cyanobacterial PBS components manually collected from published databases. This method yielded 40 potential proteins that had high sequence similarity to the PBS proteins from the cyanobacteria. Second, the MS data from the gel-extracted samples were searched against the transcriptome database of *G. pacifica* and the proteins with the 10 highest scores were selected as candidates. Then all of the candidates were subjected to 2D and 3D structure prediction, using PSIPRED and I-TASSER, respectively. The results were compared with the density maps and the atomic models built with poly-alanine sequences, and proteins with high agreement between them were selected and fitted into the density maps. The sequence assignments were then further improved using side chain information; details are given below in the description of model building and refinement. Finally, 25 protein sequences (Supplementary Table 2) were identified and confirmed by high agreement of secondary structural features and side chain information between the predicted models and the density maps in COOT[47]. The presence of these proteins in PBS was also confirmed using the MS data obtained from the intact PBS complex in solution (Extended Data Fig. 1f).

Cryo-EM sample preparation and data collection. Cryo-EM grids were prepared with Vitrobot Mark IV (FEI Company) at 16 °C and 100% humidity. We found that the PBS molecules had severe preferential orientation on grids that were glow-discharged by normal methods, probably owing to the hydrophobic nature of a specific side of the PBS. In a modified protocol, we prepared holey carbon grids (Quatifoil R2/2) by glow discharging the grids for about 40 s and keeping them in an air-dry space for about 12 h. This decreased the charge on the grid surface. We were able to detect multiple views of the complex on grids treated in this way. We applied 3.5-μl aliquot samples of the PBS complex with a concentration of 1 mg ml^{-1} to the holey carbon grids and waited for 60 s. We then added 1.5 μl 50 mM Tris, pH 7.0 to the grids and mixed the sample quickly twice to reduce the salt concentration. The grids were then blotted for 2.5 s and plunged into liquid ethane cooled by liquid nitrogen in the Vitrobot.

The vitrified grids were loaded on to a Tecnai Arctica microscope operated at a voltage of 200 kV and equipped with a Falcon II direct electron detector at a nominal magnification of 55,000×, providing a pixel size of 1.92 Å. Electron micrographs were collected manually with a defocus ranging from −1.5 μm to −2.5 μm in movie mode with each micrograph stack containing seven frames. Grids with good quality particle distribution and ice thickness were subsequently transferred to a Titan Krios microscope operated at a voltage of 300 kV and equipped with a K2 Summit direct electron detector (Gatan). Electron micrographs were collected in the counting mode of the detector using UCSF-Image4 software[48] with a nominal magnification of 22,500×. The final pixel size of the micrographs was 1.32 Å and the data set had defocuses ranging from −1.4 μm to −2.3 μm. The detector dose rate was set at about 8.2 counts per physical pixel per s with a total exposure time of 8 s, and each micrograph stack contained 32 frames.

Image processing. A total of 2,346 micrographs were collected with the Arctica microscope, and a total of 7,477 micrographs were collected using the Titan Krios. In each movie stack, all the frames were aligned and summed to correct specimen drift and beam-induced motion with an overall image motion correction program MOTIONCORR (written by X. Li)[48]. Micrograph screening, particle picking and normalization were done using EMAN2[49] and RELION[50,51]. The contrast transfer

function parameters of each micrograph were estimated using CTFFIND3[52] and all the 2D and 3D classification and refinement were performed with RELION.

The Arctica data set was collected and processed to inspect the sample quality and composition (Extended Data Fig. 1g–l). Approximately 5,000 particles with different views were manually picked to produce templates for the reference-based particle picking procedure in RELION. Particles were then semi-automatically picked by RELION and two rounds of reference-free 2D classification were performed to remove ice contaminants, aggregated and damaged particles. On the basis of the analysis, additional rounds of manual particle picking and screening were performed to improve the overall quality of the final data set, which included 352,718 particles. On the basis of the 2D average classes, we determined that there were about five major distinct views and that the shape of this complex was cylindrical. We produced a cylinder (700 Å high, 400 Å diameter) by SPIDER and used it as an initial model for 3D classification. The particles were divided into eight classes during the first round of 3D classification by global search and three classes of the particles, representing 28.8% of the total particles, displayed a regular and normal appearance. All 80,857 particles from these three classes were grouped together and subjected to another round of 3D classification, generating a final class containing 68,647 particles. These particles produced a reconstruction with an average resolution of 5.0 Å after auto-refinement with C2 symmetry. We also performed another round of auto-refinement without imposing symmetry and obtained a reconstruction with an average resolution of 5.8 Å. The two maps, with and without C2 symmetry imposition, were very similar and the angular distribution of the particles used for the final reconstruction was rational, verifying the correctness of the reconstruction and the presence of C2 symmetry in the complex. The final density map with C2 symmetry shows clear features of the secondary structural elements so that the crystal structure of phycoerythrin (PDB 2VJH) fit well into the density map.

The Titan Krios data set was processed using the same procedure as described above (Extended Data Fig. 2f). In total, 471,482 particles were picked by semi-automatic picking, manual picking and screening with RELION and EMAN2. Two rounds of reference-free 2D classification were then performed to further remove bad particles or ice contaminants, yielding 396,572 particles to be used for 3D classification. Two rounds of 3D classification were used and in each round all particles were split into eight classes. For the first round of 3D classification, the density map produced from the Arctica data set was used as the initial model and this classification allowed us to generate 71,383 particles (18% of the total). In the second round of 3D classification, four classes with good qualities were separated out and produced 53,403 particles. These four classes were combined and used for final reconstruction with an average resolution of 3.5 Å. The core region of this density map shows clear secondary structural elements and amino acid side chains. The results of the 3D classification indicate that some subunits in side regions may dissociate owing to the decreased salt concentration. To deal with the missing and flexible nature of these regions, several local masks were applied to the auto-refine procedure. First we applied an overall mask of this complex at the 17-iteration step, at which local refinement did not start. To improve the resolution of the core region, we used three masks with gradually reducing sizes as the last step of the auto-refine procedure. To improve the resolution of side regions, three different overlap local masks were individually applied to selected regions and then several local masks were gradually applied to target regions (Extended Data Fig. 2g). These procedures finally resulted in improved quality of local maps with resolutions ranging between 3.4 Å and 4.3 Å for most regions (Extended Data Fig. 2g). The maps for the target regions were extracted from the overall map by CHIMERA[53], and the masks were created by RELION.

All of the resolutions reported are estimated by the gold-standard $FSC_{0.143}$ criterion. The modulation transfer function (MTF) of the detector and a negative *B*-factor were applied during the post-process procedure. All the local resolution maps were calculated by using ResMap[54].

Model building and refinement. Owing to the wide range of resolution limits for various regions of the PBS complex, we combined *de novo* model building and homologous structure docking to generate an atomic model. Local maps generated by the different masking strategies described above were used to facilitate the model building process.

The majority of the PBS complex was constituted by PBPs, so we first built the models of PBPs. Crystal structures of the *Mastigocladus laminosus* APC-linker complex (PDB 1B33)[27] were used as the initial homology template for core region modelling, and crystal structures of phycocyanin from *Synechococcus elongatus* (PDB 1JBO)[55] and phycoerythrin from *Gloeobacter violaceus* phycocyanin (PDB 2VJH) were used as the initial homology templates for side region modelling. All of these crystal structures were rigid body docked into the density map, with repeats, manually using CHIMERA. The types of different PBPs were confirmed by high agreement of side chain information between the sequences and the density maps in COOT. Owing to the differing side chain information, four types of PBPs (α^{APC}, β^{APC}, ApcD and ApcF) in the core region, and four types of PBPs (α^{PC}, β^{PC}, α^{PE} and β^{PE}) in the side region can be unambiguously distinguished. The number of bilins in PBPs is highly conserved between cyanobacteria and red alga, which provided more information for us to localize PBPs. After docking, the entire chains were manually checked and adjusted with COOT. Model building using this procedure is summarized in Supplementary Table 1 under 'Homology modeling'.

In addition to PBPs, there are many densities that belong to the linker proteins. The reported crystal structures of the *M. laminosus* core–linker protein fragment (Pfam01383, PDB 1B33)[27] and *Synechocystis sp.* PCC 6803 rod–linker L_R protein fragment (Pfam00427, PDB 3NPH)[28] provide us with some information about linker proteins. These two crystal structures were fitted into the density maps based on their secondary structure information in CHIMERA, and we found that they mainly localized to the core region and that the rod region connected to the core. For modelling of these linker proteins, the 2D and 3D predictions of potential PBS proteins were taken into consideration and eight linker proteins ($L_{RC}1$, $L_{RC}2$, $L_{RC}3$, L_R1, L_R2, L_R3, L_C and L_{CM},) were confirmed by high agreement of secondary structural features and side chain information between the predicted models and the density maps. For building the regions outside of these two crystal structures in both the N and C directions, poly-alanine models were first built manually in COOT with the help of secondary structure features to trace the main chain, and sequence assignments were further improved by well-resolved bulky residues such as Phe, Tyr, Trp and Arg. Taking L_{CM} as an example, the three Pfam00427 domains and one α-subunit domain were initially fitted into the density map by CHIMERA, and the main chains and side chains were then manually adjusted in COOT. The remaining part of L_{CM} was built by poly-alanine tracing and amino acid replacement in COOT. Model building using this procedure is summarized in Supplementary Table 1 under 'Homology modeling' and '*De novo* building'.

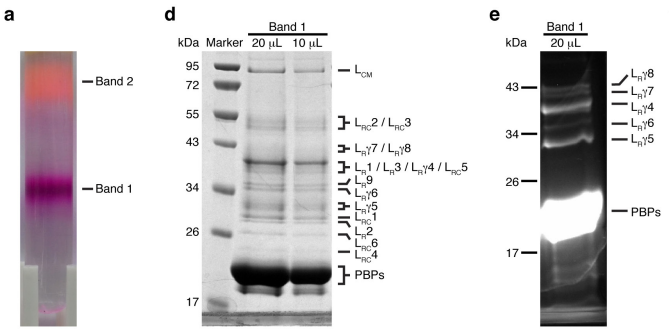
For those densities that could not be fitted by crystal structures, *de novo* model building was performed, as summarized in Supplementary Table 1 under '*De novo* building'. Based on the density map, nine linker proteins ($L_{RC}4$, $L_{RC}5$, $L_{RC}6$, $L_R\gamma4$, $L_R\gamma5$, $L_R\gamma6$, $L_R\gamma7$, $L_R\gamma8$ and L_R9) were initially modelled with poly-alanine sequences and then the amino acids were replaced in COOT. During modelling, the chemical properties of amino acids were considered, in order to facilitate model building, and sequence assignments were guided mainly by bulky residues such as Phe, Tyr, Trp and Arg. Unique patterns of sequences were exploited for validation of residue assignment.

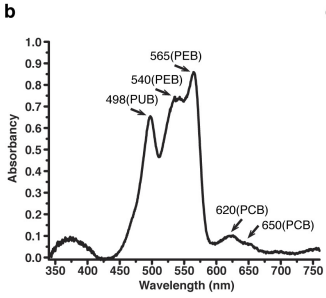
Initial structure refinement was carried out using phenix.real_space_refine[56] with secondary structure and geometry restraints to prevent over-fitting. The local maps for different parts generated during reconstruction were applied during real space refinement. The overall model was then refined using REFMAC[57,58] in reciprocal space using secondary structure restraints that were generated by ProSMART[59]. During this process, to prevent geometry clash among all proteins, each protein and its interacting proteins were grouped together and refined against the map of the best resolution for this protein using a purpose-written script in REFMAC. Finally, the overall model was refined again against the overall 3.5 Å map using phenix.real_space_refine. All protein structures in the PBS complex were validated through examination of their Molprobity scores and the statistics of Ramachandran plots; the results are shown in Extended Data Tables 1, 2.

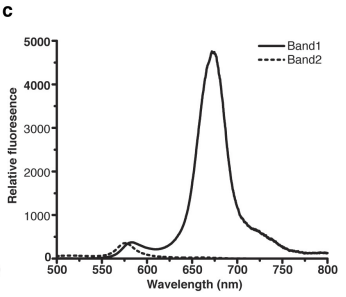
Data availability. The atomic coordinates have been deposited in the Protein Data Bank with the accession code 5Y6P. The EM maps have been deposited in the Electron Microscopy Data Bank with the accession codes 6769 for the overall map and 6758–6768 for the eleven local maps. The 25 DNA sequences have been deposited in GenBank under accession numbers MF523540–MF523564. All other data are available from the corresponding authors upon reasonable request.

44. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protocols* **8,** 1494–1512 (2013).
45. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21,** 3674–3676 (2005).
46. Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O. & Borodovsky, M. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res.* **33,** 6494–6506 (2005).
47. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132 (2004).
48. 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).
49. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157,** 38–46 (2007).
50. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180,** 519–530 (2012).

51. Scheres, S. H. W. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415,** 406–418 (2012).
52. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142,** 334–347 (2003).
53. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
54. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11,** 63–65 (2014).
55. Nield, J., Rizkallah, P. J., Barber, J. & Chayen, N. E. The 1.45 Å three-dimensional structure of C-phycocyanin from the thermophilic cyanobacterium *Synechococcus elongatus*. *J. Struct. Biol.* **141,** 149–155 (2003).
56. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221 (2010).
57. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255 (1997).
58. Brown, A. *et al.* Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D* **71,** 136–153 (2015).
59. Nicholls, R. A., Fischer, M., McNicholas, S. & Murshudov, G. N. Conformation-independent structural comparison of macromolecules with ProSMART. *Acta Crystallogr. D* **70,** 2487–2499 (2014).

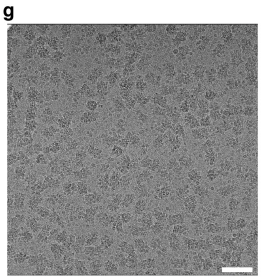


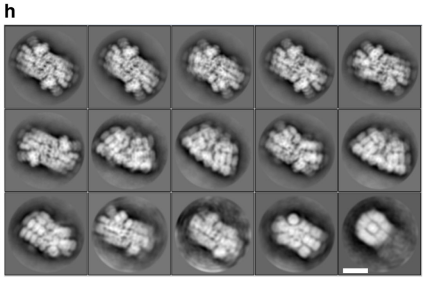


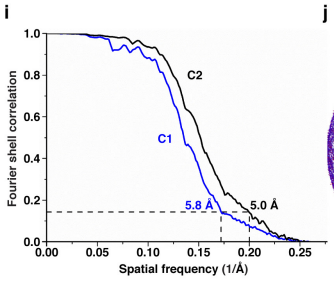


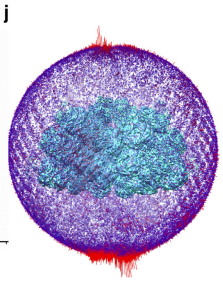
f

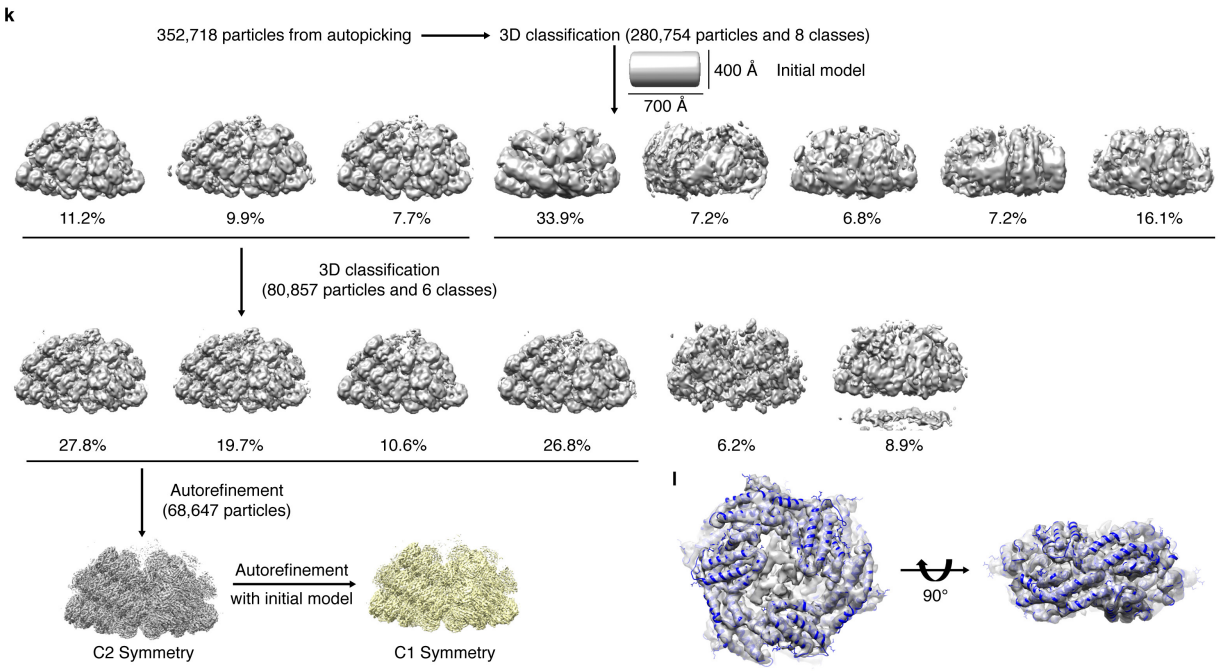
Name	# Peptides		# PSMs		MW [kDa]
	batch 1	batch 2	batch 1	batch 2	
α^{APC}	16	18	147	153	17.6
β^{APC}	15	17	167	180	17.5
ApcD	7	11	16	29	18.1
ApcF	10	10	66	67	19.9
α^{PC}	16	15	318	325	17.4
β^{PC}	14	17	178	197	18.4
α^{PE}	18	20	536	552	17.7
β^{PE}	20	20	362	419	18.5
L_C	4	5	21	23	10.5
L_{CM}	43	50	313	334	100.7
$L_{RC}1$	11	12	59	54	27.1
$L_{RC}2$	19	22	153	142	52.6
$L_{RC}3$	28	24	84	80	57.1
$L_{RC}4$	9	9	28	29	20.5
$L_{RC}5$	11	13	45	49	33.4
$L_{RC}6$	16	17	61	75	27.6
L_R1	13	17	99	116	40.6
L_R2	11	12	40	45	30.4
L_R3	18	18	67	69	38.9
$L_R\gamma4$	12	12	29	23	34.4
$L_R\gamma5$	11	7	45	52	30.9
$L_R\gamma6$	5	5	6	7	32.7
$L_R\gamma7$	9	12	26	23	37.6
$L_R\gamma8$	5	7	10	14	38.0
L_R9	13	12	56	40	38.7





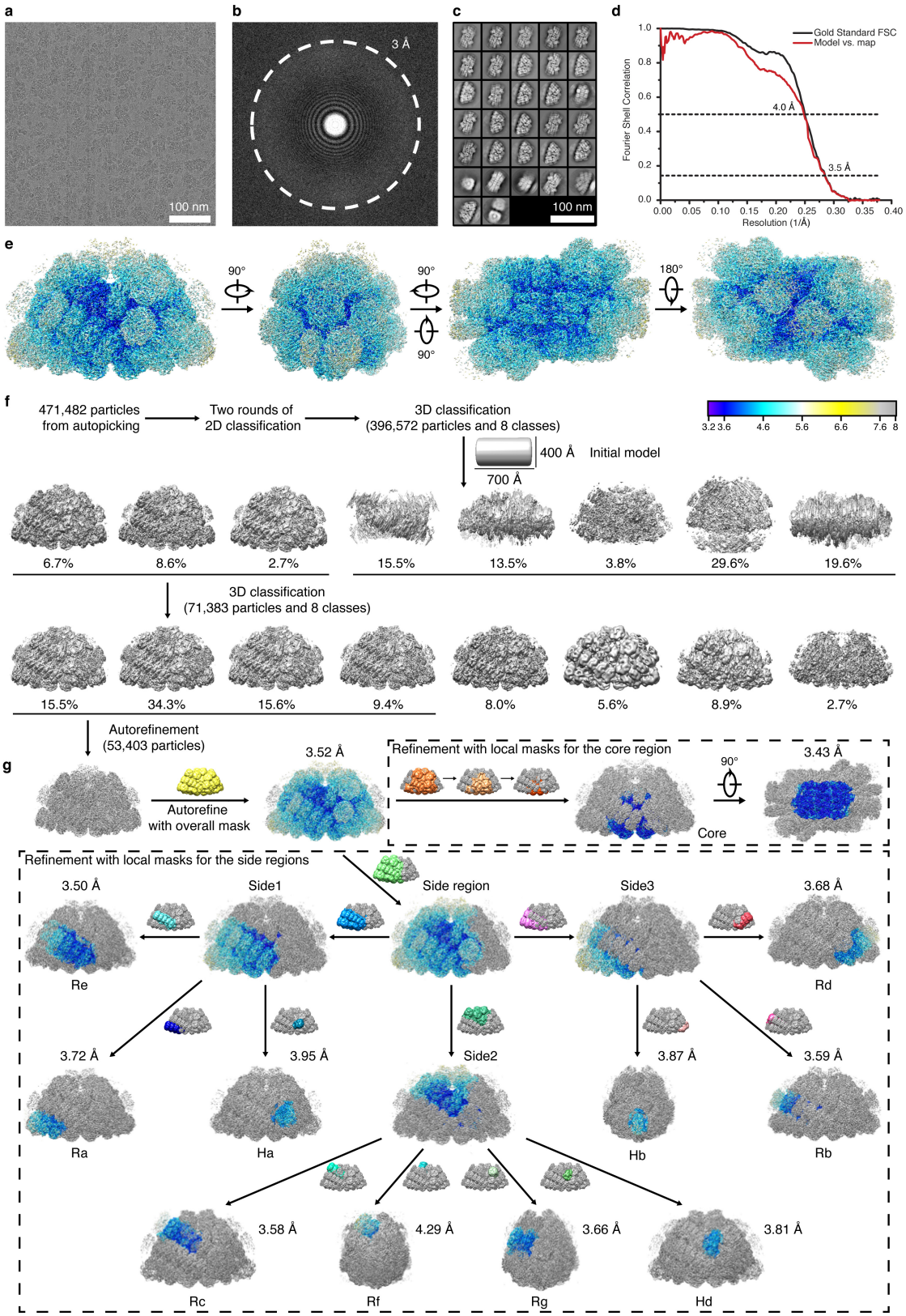




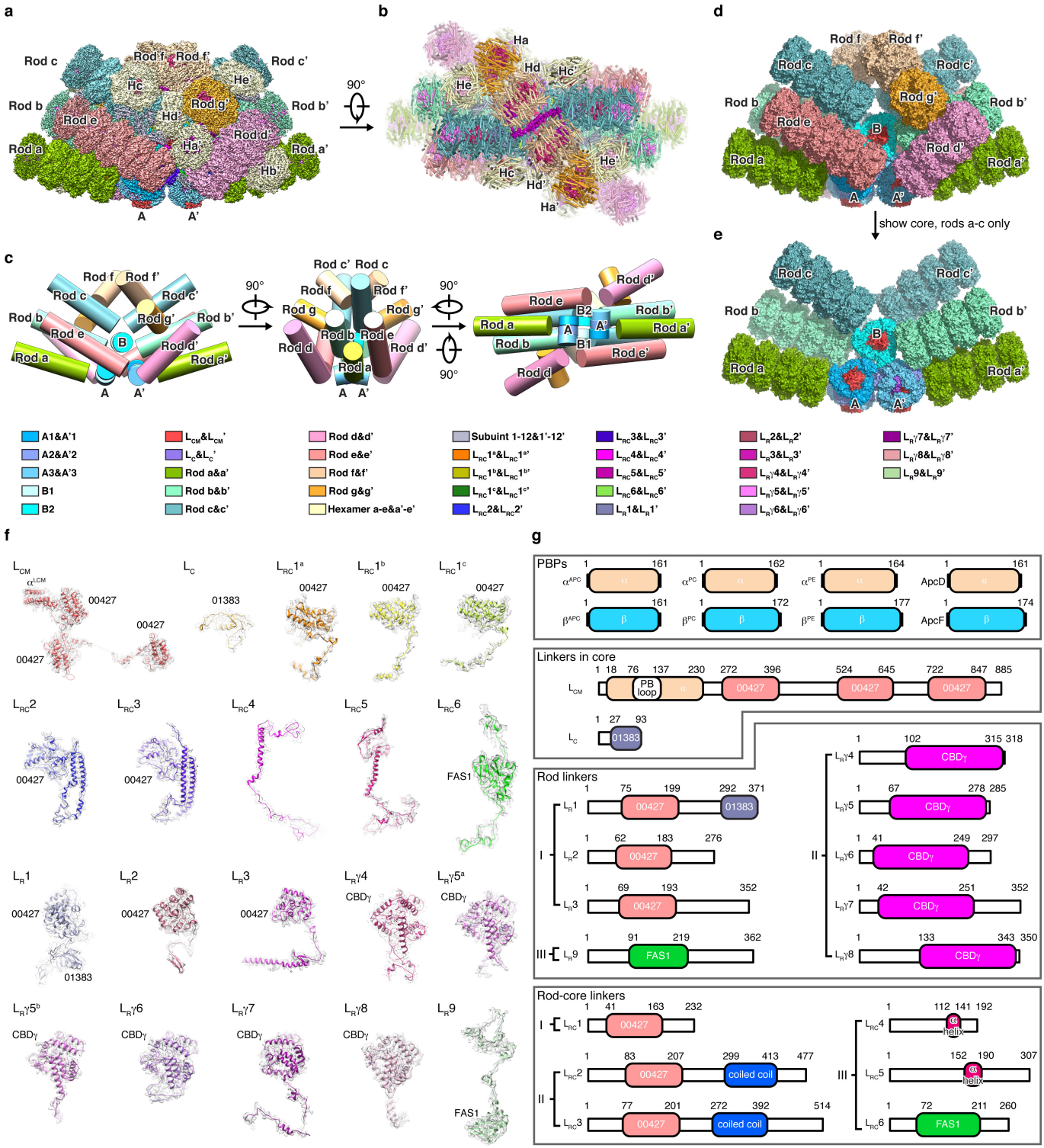


Extended Data Figure 1 | See next page for caption.

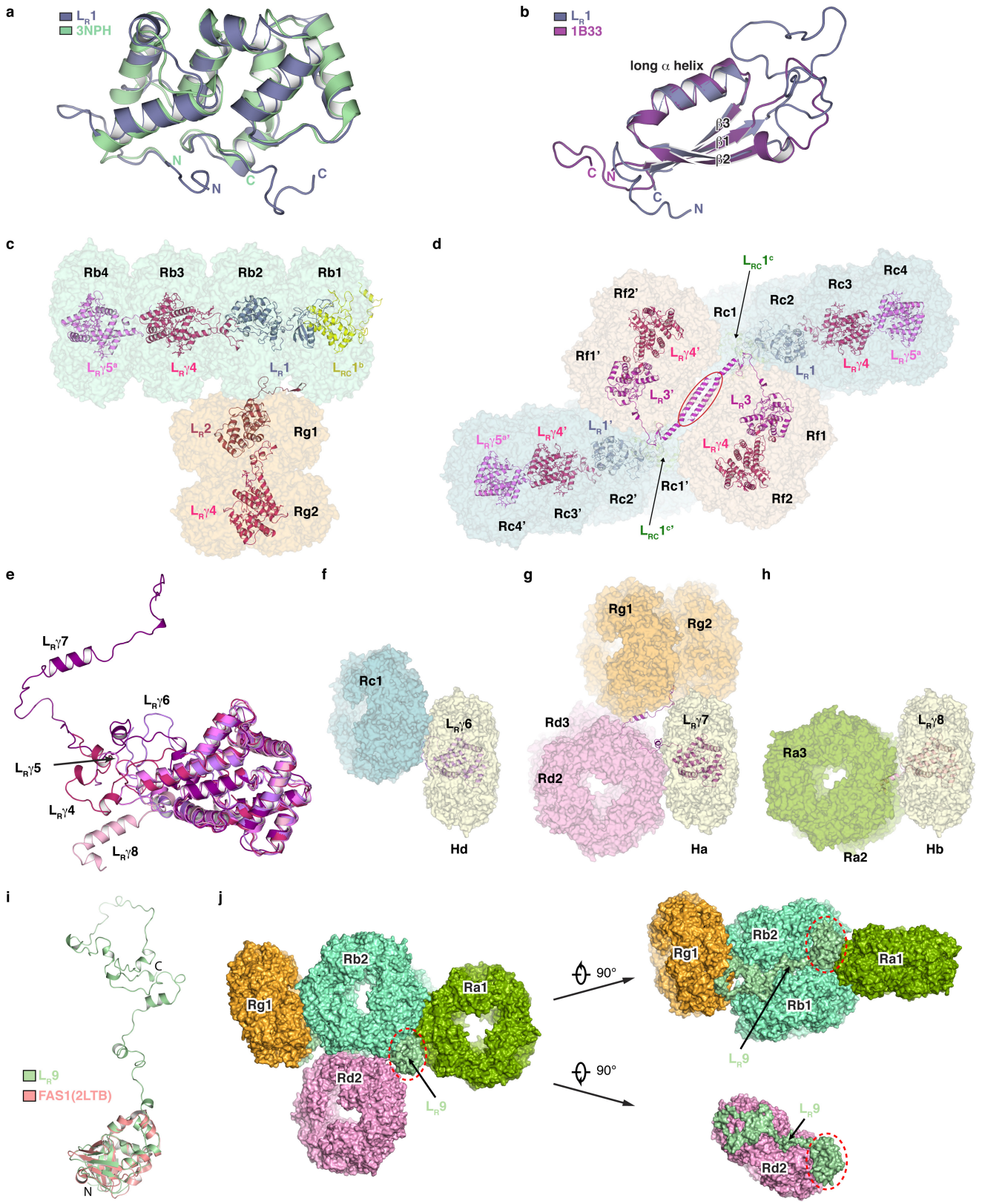
Extended Data Figure 1 | Preparation and characterization of the PBS from *Griffithsia pacifica* and cryo-EM analysis using Tecnai Arctica microscope and FEI Falcon II direct electron detector. a, Isolation of PBSs using sucrose density gradient centrifugation. Band 1 is the sample of PBSs used for single particle analysis in this study. Band 2 is disassembled PBSs as evidenced by the fluorescence spectra shown in **c**. **b**, Absorption spectrum of band 1. The peaks at 498 nm, 620 nm and 650 nm are from phycourobilins of the $L_R\gamma$ proteins, phycocyanobilins of phycocyanins and phycocyanobilins of allophycocyanins, respectively. The peaks at 540 nm and 565 nm are from phycoerythrobilins of phycoerythrin. **c**, Fluorescence emission spectra of band 1 and band 2. Band 1 has an emission peak at 674 nm and band 2 has an emission peak at 580 nm, indicating that band 1 contains intact PBSs and band 2 contains disassembled PBSs. **d**, SDS–PAGE analysis of protein components in PBS from band 1. The gel was stained with Coomassie brilliant blue. The bands of L_{CM}, L_R and L_{RC} proteins and PBPs identified by MS analysis are indicated. For gel source data, see Supplementary Fig. 1. **e**, Protein composition analysis of band 1 by SDS–PAGE stained with $ZnSO_4$ to detect bilin-containing proteins with ultraviolet light by Zn-enhanced fluorescence. The bands of $L_R\gamma4$–8 and PBPs were identified by MS analysis and are indicated. For gel source data, see Supplementary Fig. 1. **f**, MS analysis of purified PBSs. Two batches of sample were analysed. The similar results confirmed the consistency of our purification method. **g**, A representative motion-corrected electron micrograph of PBSs. Scale bar, 100 nm. **h**, Typical good, reference-free 2D class averages from single-particle images of PBSs. Scale bar, 20 nm. **i**, Gold-standard Fourier shell correlation (FSC) curves of the cryo-EM map of PBS with or without C2 symmetry imposed. **j**, Euler angle distribution of particles contributing to the final reconstruction of PBS. Each cylinder represents one view and the size of the cylinder is proportional to the number of particles for that view. **k**, The workflow for 3D classification. **l**, The crystal structure of allophycocyanin (PDB 1B33) can be fitted well with the density map of PBS in different view.



Extended Data Figure 2 | Cryo-EM analysis of PBS using Tecnai Krios microscope and Gatan K2 Summit direct electron detector. **a**, A representative motion-corrected electron micrograph of PBSs. Scale bar, 100 nm. **b**, Fourier power spectrum of the micrograph showing the Thon ring extending to 3 Å. **c**, Typical good, reference-free 2D class averages from single particle PBS images. Scale bar, 100 nm. **d**, Gold-standard FSC curve of the final cryo-EM map with a value of 0.143 at a resolution of 3.5 Å and the FSC curve between the refined model and the final map with a value of 0.5 at a resolution of 4.0 Å. **e**, Colour-coded final 3D reconstruction of a PBS showing local resolution in different views. **f**, The workflow for the 2D and 3D classifications for cryo-EM data processing. **g**, The masking strategy for dealing with flexible regions of PBS. Several local masks for different parts were applied and further, reduced-size masks were applied to some specific areas, resulting in improved resolution and local map qualities.

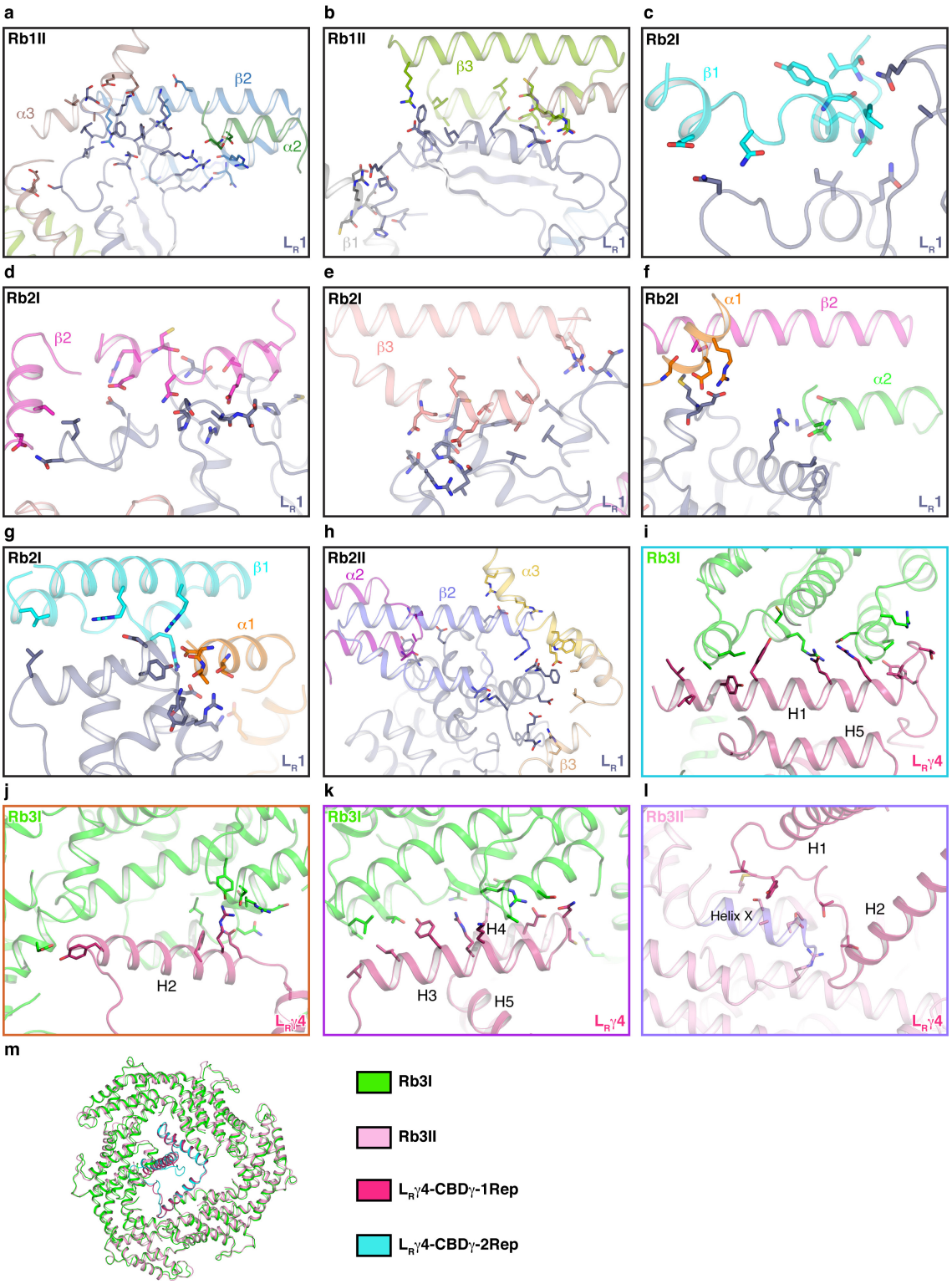


Extended Data Figure 3 | Overall structure of PBS. a, 3D cryo-EM density map of PBS. **b**, Top view of the atomic model of PBS. **c**, Schematic diagram showing the organization of the rods and core cylinders from three perpendicular views. **d**, The organization of the rods and core displayed in surface representation. **e**, The organization of the key structural elements (rods a–c and the core) of PBS displayed in surface representation. The protein components are colour-coded as shown. **f**, Structures of the 20 well-resolved linker proteins. Atomic models of these linker proteins superimposed with their segmented cryo-EM densities (transparent grey). Domains of α^{LCM}, Pfam00427 (00427), Pfam01383 (01383), CBDγ, and FAS1 are labelled. **g**, Schematic drawing of structural elements of the PBS subunits. PBS subunits are classified into four groups according to their positions and functions: phycobiliproteins (PBPs), linkers in the core, rod linkers and rod–core linkers. Different structural elements are separately coloured.

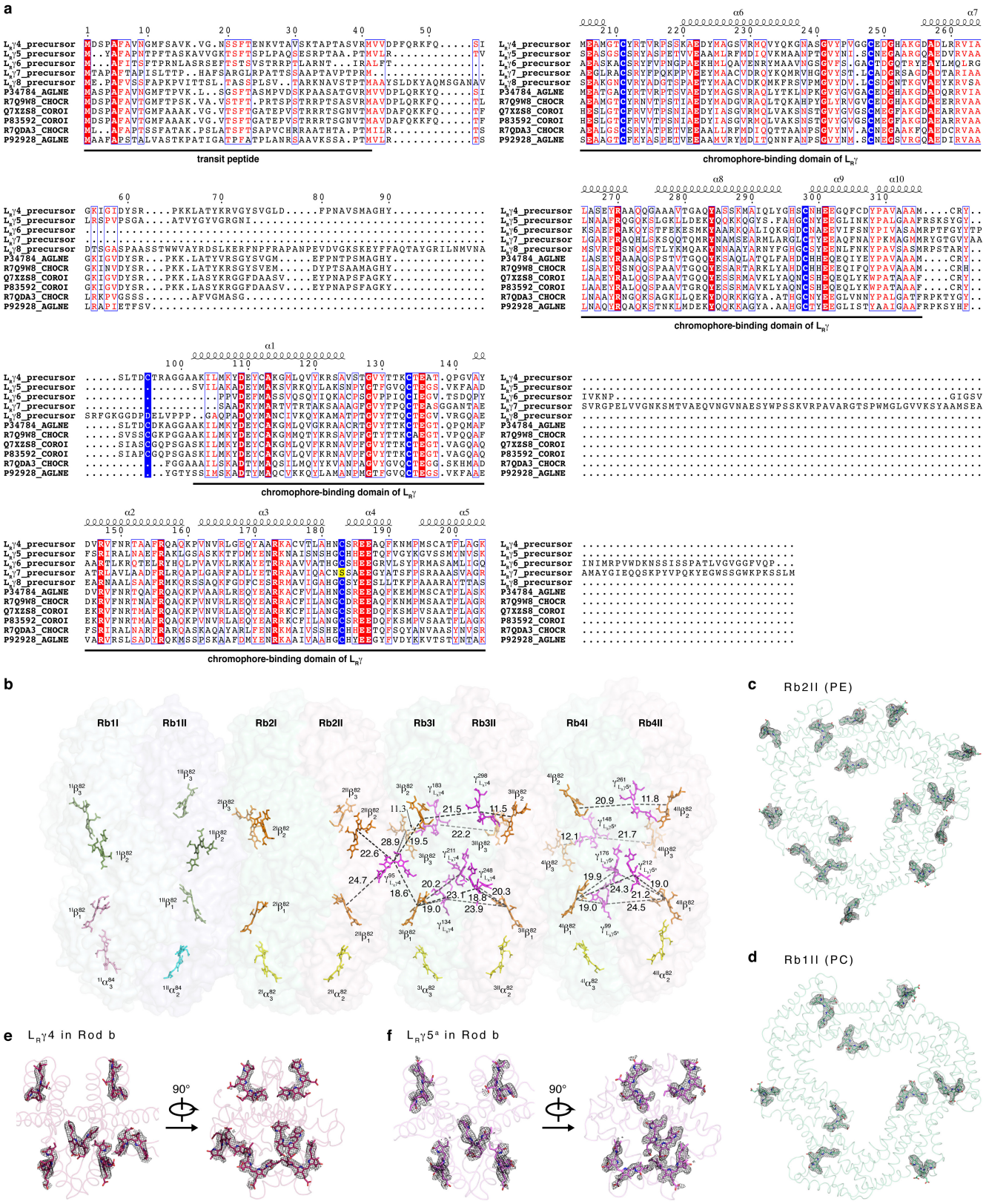


Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | Structures and binding partners of rod linker proteins. a, Comparison of the Pfam00427 domain of L_R1 with the crystal structure of the Pfam00427 domain (PDB 3NPH) from linker L_R^{30} from *Synechocystis* sp. PCC 6803. The two structures are almost identical. **b**, Comparison of the Pfam01383 domain of L_R1 with the crystal structure of the Pfam01383 domain (PDB 1B33) from the PBS of *Mastigocladus laminosus*. The long α-helix and three β-strands are labelled. **c**, L_R2 is located in the inner hexamer of rod Rg (Rg1). Its C-terminal extended loop and a small β-sheet that is composed of two antiparallel β-strands at the very C terminus contact the outer surface of hexamers Rb1 and Rb2. **d**, L_R3 is located in the inner hexamer of rod Rf (Rf1) and its C terminus ends with a long α-helix. The N-terminal part of the long helix inserts into the area between hexamers Rc1 and Rf1 and the C-terminal part of the long helix forms a coiled-coil structure (red oval) with its symmetric counterpart L_R3' located in hexamer Rf1′. These structures stabilize the rod pair Rf and Rf′ by association with each other and anchor them to rods Rc and Rc′. **e**, Structural superimposition of $L_R\gamma4$–8 proteins, showing the almost identical CBDγ structures and different conformations outside the CBDγ. **f**–**h**, $L_R\gamma6$–8 with their CBDγ domains located in hexamers Hd, Ha, and Hb, respectively. The C-terminal loop of $L_R\gamma6$ anchors on the outer side of the innermost hexamer of rod Rc (Rc1) (**f**); the C terminus of $L_R\gamma7$ contains several short α-helices and interacts with the surface of hexamers Rd2 and Rg1 (**g**); and an α-helix of the N terminus of $L_R\gamma8$ inserts into the space between the two hexamers Ra2 and Ra3 (**h**). **i**, Structural superimposition of the N-terminal domain of L_R9 and FAS1 domain (PDB 2LTB). **j**, L_R9 sits between rods Ra, Rb, Rd and Rg and interacts with them. The FAS1 domain is circled by a red dashed line.

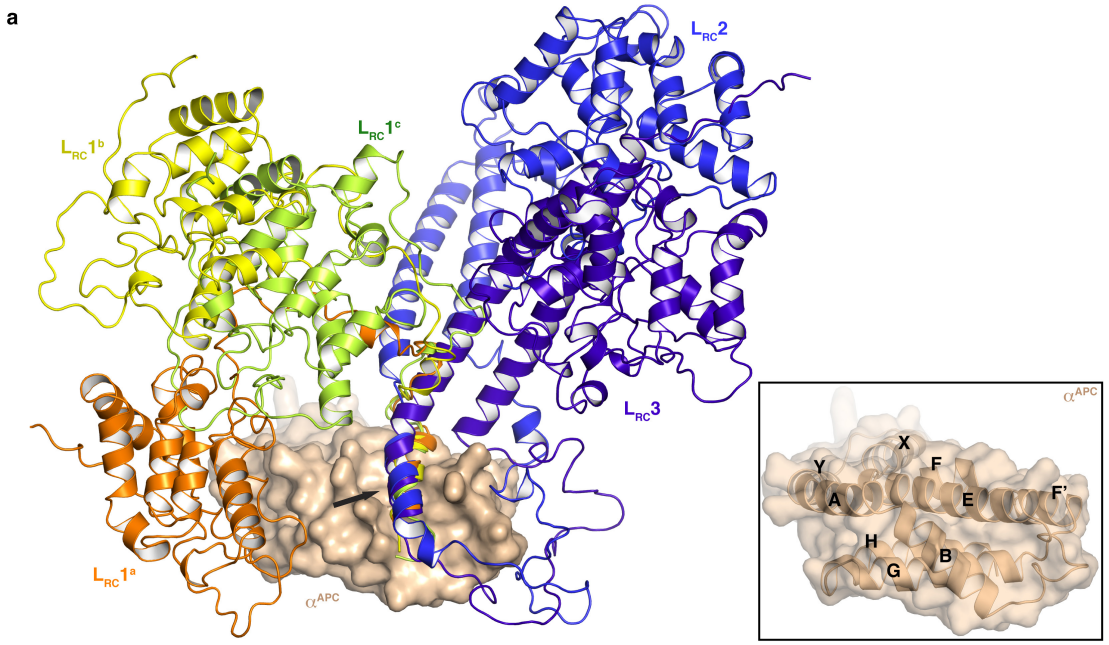


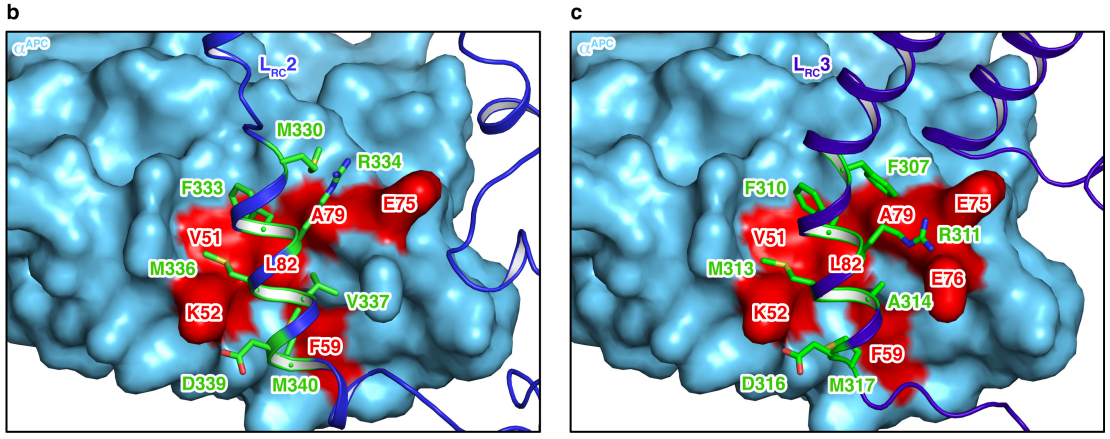
Extended Data Figure 5 | Extensive interactions between rod linker proteins and PBPs in rod Rb. a, **b**, Expanded views of interactions between the CTR of L_R1 and trimer Rb1II, corresponding to Fig. 2c. **c**–**e**, Expanded views of interactions between the intertwined NTE and CTE of the N-terminal region of L_R1 and trimer Rb2I, corresponding to Fig. 2d. **f**, **g**, Expanded views of interactions between the Pfam00427 domain of L_R1 and trimer Rb2I, corresponding to Fig. 2f. **h**, Expanded view of interactions between the Pfam00427 domain of L_R1 and trimer Rb2II, corresponding to Fig. 2g. **i**–**k**, Expanded views of interactions between the CBDγ of $L_R\gamma4$ and trimer Rb3I, corresponding to Fig. 2i. **l**, Expanded view of interactions between the CBDγ of $L_R\gamma4$ and trimer Rb3II, corresponding to Fig. 2j. Helix X is labelled and highlighted. Residues involved in atomic interaction sites are shown in the stick representation. α, β and H denote α-subunit, β-subunit and helix, respectively. **m**, CBDγ contacts the inner face of the rod hexamer in a symmetrical fashion. When the first repeat of CBDγ of $L_R\gamma4$ ($L_R\gamma4$-CBDγ-1Rep) is aligned with the second repeat of CBDγ of $L_R\gamma4$ ($L_R\gamma4$-CBDγ-2Rep), the trimers (Rb3I and Rb3II) that they interact with are also aligned well.

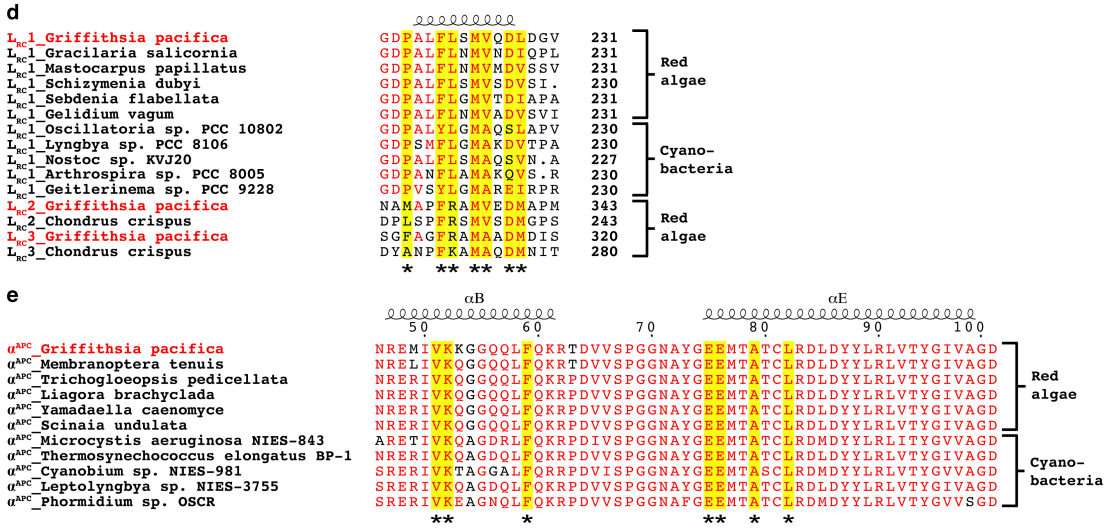


Extended Data Figure 6 | See next page for caption.

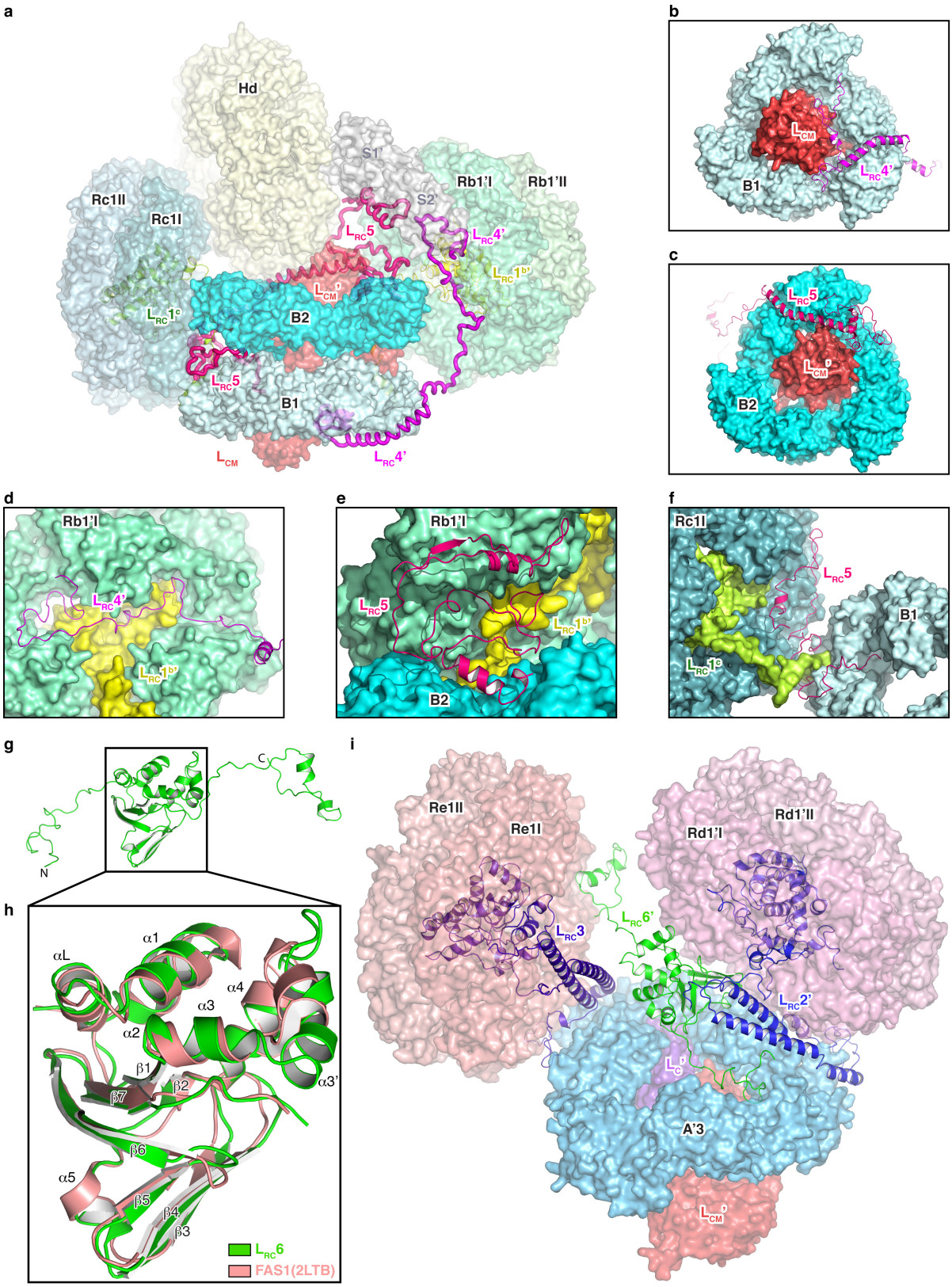
Extended Data Figure 6 | Sequence alignment of $L_R\gamma4$–8 and the γ-subunits from different red algal species and the bilins in rod b. **a**, Amino acid sequence alignment. The predicted transit peptides deduced from the sequence alignment and the CBDγ domains based on the structures are labelled. Conserved cysteine residues covalently attached to chromophores are highlighted in blue. Note that one of the cysteine residues in the CBDγ of $L_R\gamma7$ is replaced by serine (highlighted in yellow), consistent with the fact that only four chromophores are found in $L_R\gamma7$ as compared with other CBDγ domains, which have five chromophores. P34784, P92928, γ-subunits from *Aglaothamnion neglectum*; R7Q9W8, R7QDA3, γ-subunits from *Chondrus crispus*; Q7XZS8, P83592, γ-subunits from *Corallina officinalis*. **b**, Distances between some bilins in Rb. Magenta, bilins of $L_R\gamma$; orange, bilins of β-subunits in phycoerythrin hexamers; yellow, bilins of α-subunits in phycoerythrin hexamers; dark green, bilins of β-subunits in phycocyanin hexamer; cyan, bilins of α-subunits in phycocyanin hexamer; pink, bilins nearest to the core in phycocyanin hexamer. The numbers near the dashed lines indicate the distances (Å) between bilins on linker proteins and adjacent bilins on PBPs. **c**–**f**, Cryo-EM densities (mesh) of bilins (stick) in the phycoerythrin trimer Rb2II (**c**), the phycocyanin trimer Rb1II (**d**), the $L_R\gamma4$ protein (**e**) and the $L_R\gamma5^a$ protein (**f**).



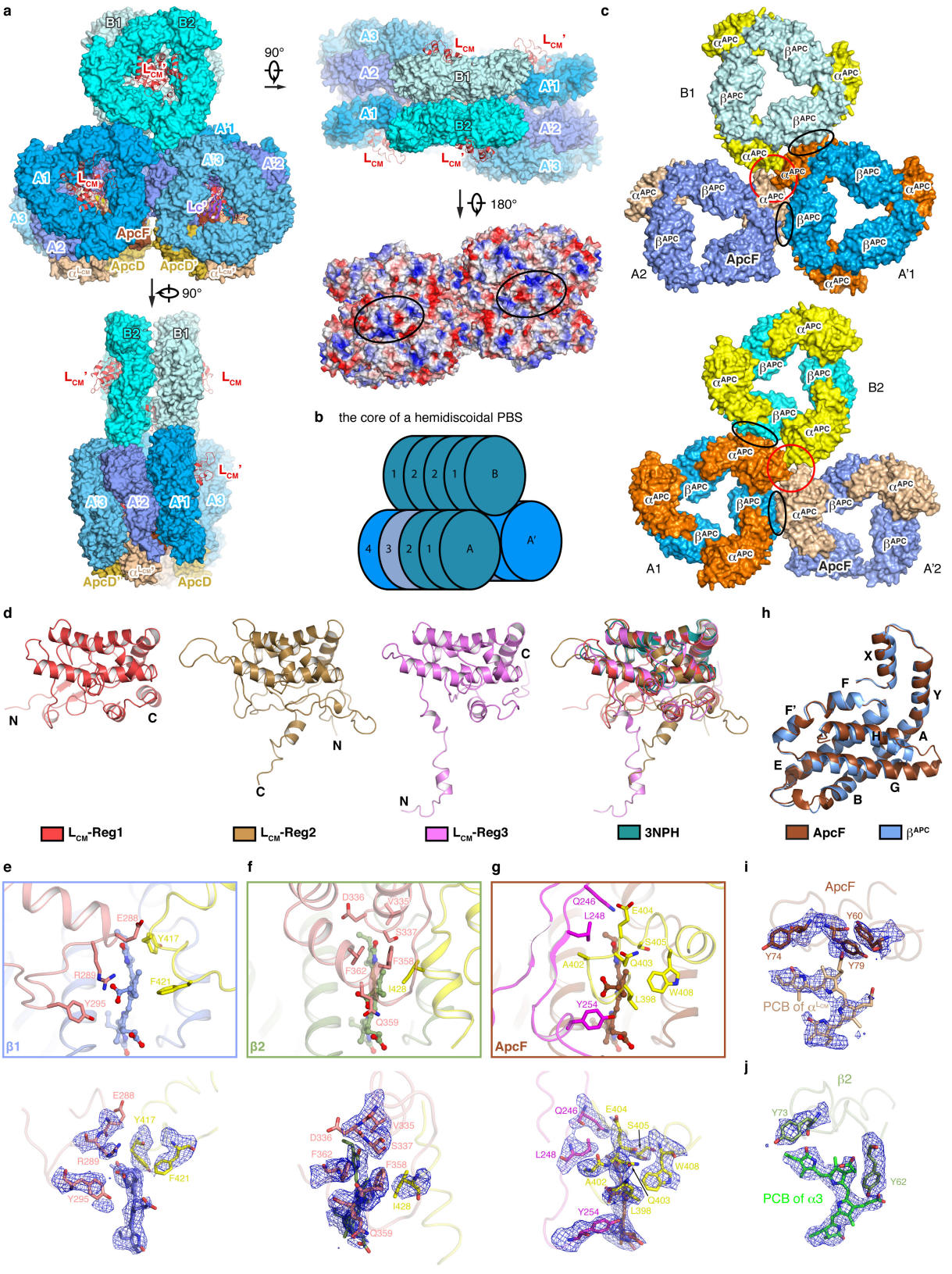




Extended Data Figure 7 | Interactions between $L_{RC}1$–3 and the core. **a**, Superimposition of the α^{APC} subunits and the spatial positions of their interacting L_{RC} proteins. The alignments were performed using the α^{APC} subunit as a reference. Note that the contacting helices of all L_{RC} proteins (black arrow) are also aligned very well. Inset shows the groove formed by helices B and E of the α^{APC} subunit. **b**, **c**, Interactions between the α^{APC} subunit and the helices of $L_{RC}2$ (**b**) and $L_{RC}3$ (**c**). The residues involved in the interaction of L_{RC} proteins are coloured green and shown in stick representation. The α^{APC} are shown in surface representation, and the residues involved in the interaction are red. **d**, **e**, Sequence alignment of $L_{RC}1$–3 (**d**) and α^{APC} (**e**) from different red algal and cyanobacterial species. The residues involved in the interaction are indicated by asterisks and highlighted yellow.

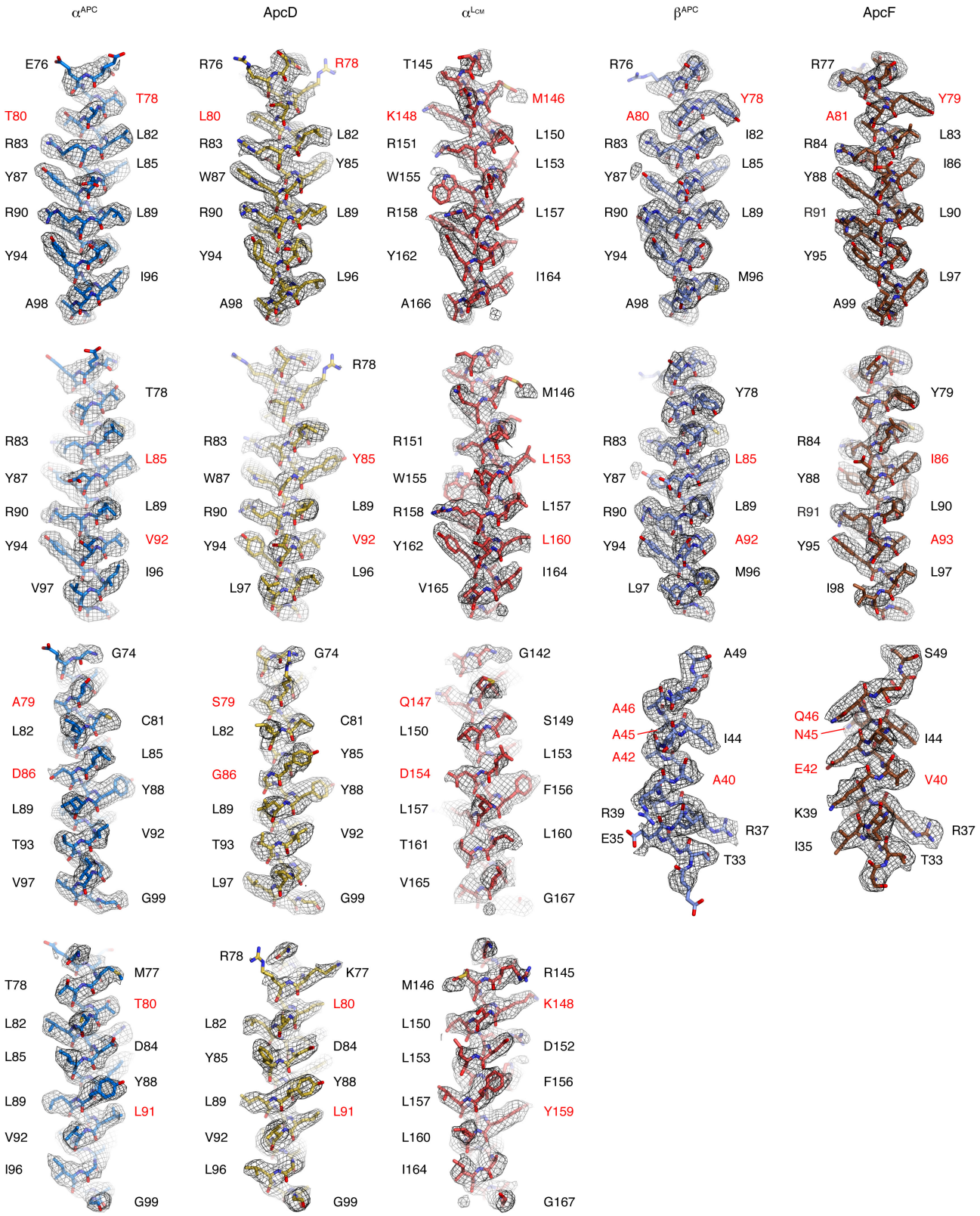


Extended Data Figure 8 | Structures and assembling roles of $L_{RC}4$, $L_{RC}5$ and $L_{RC}6$. a, Overview of $L_{RC}4$ and $L_{RC}5$ wrapping their target protein components. S1′ and S2′ are single α- and β-subunits, respectively. **b**, The long α-helix at the middle of $L_{RC}4'$ spans across one α-subunit of core trimer B1. **c**, The long α-helix at the middle of $L_{RC}5$ spans across one α-subunit of the trimer B2. **d**, The N-terminal extension of $L_{RC}4'$ meanders at the surface of trimer Rb1′I and $L_{RC}1^{b'}$. **e**, The C-terminal extension of $L_{RC}5$ interacts with core layer B2, rod trimer Rb1′I, and $L_{RC}1^{b'}$. **f**, The N-terminal extension of $L_{RC}5$ wraps around the surfaces of trimers Rc1I and B1, and also contacts $L_{RC}1^{c}$. **g**, The structure of $L_{RC}6$. **h**, Structural superimposition of the FAS1 domain of $L_{RC}6$ and the known FAS1 domain (PDB 2LTB). α-helices and β-strands are labelled. **i**, $L_{RC}6'$, like a wedge, is inserted into the area between trimers Rd1′I and Re1I and layer A′3 of the core, and makes extensive contacts with the surrounding protein components, including A′3, L_{CM}', L_{C}', $L_{RC}2'$, Rd1′I and Re1I with its extended loops on both sides of the FAS1 domain.

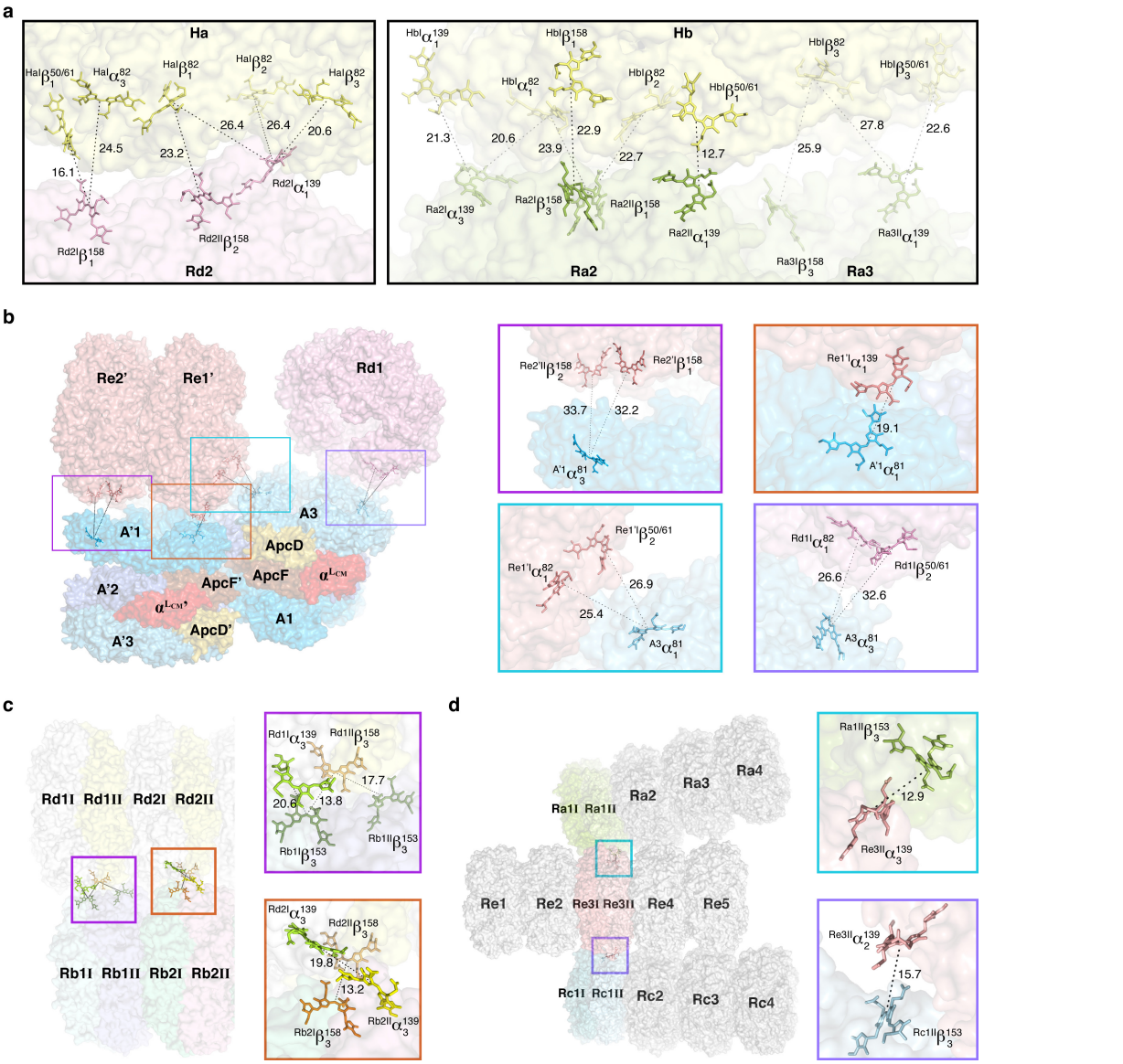


Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Structure of the core. a, Structure of the core in four perpendicular views shows the assembly and arrangement of the core layers. The bottom view (the same as the view of Fig. 4c) shows the surface electrostatic potential distributions of the core. The surfaces of α^{LCM} molecules are circled. The surface electrostatic potentials were calculated in PyMol. Red, negative electrostatic potential; blue, positive electrostatic potential; grey, non-polar. **b**, A cartoon model of the core of a hemidiscoidal PBS (adapted from ref. 1). Note that the basal cylinder has four trimers. **c**, The inter-cylinder interaction in the core is symmetrical: trimers B1, A2 and A′1 interact with one another and trimers B2, A′2 and A1 interact with one another. The association is mediated by the interaction between three α-subunits (red circle) and two pairs of α–β subunits (black circle). **d**, Comparison of the three regions of L_{CM} with the crystal structure of the Pfam00427 domain. The structures of the three regions of L_{CM} (L_{CM}-Reg1–3) are displayed as indicated and they plus the crystal structure of the Pfam00427 domain (PDB 3NPH) are aligned with one another. **e**–**g**, Close-up views of three PCB pockets of the β-subunits boxed by dashed lines in Fig. 4d: β1 (**a**), β2 (**b**) and ApcF (**c**). The bilins and their surrounding residues are shown in ball-stick representation and stick representation, respectively. Their cryo-EM densities (mesh) superimposed with the respective atomic models (stick) are shown in the lower panels. **h**, Structural superimposition of ApcF and β^{APC} shows that ApcF has an extension at the tip of the G–H helix hairpin. **i**, **j**, Cryo-EM densities (mesh) of the bilins in α^{LCM} and the α3 subunit of core layer A2, and their surrounding residues, corresponding to Fig. 4f, superimposed with their respective atomic models (stick; coloured as in Fig. 4).



Extended Data Figure 10 | Identification of the APC subunits. Cryo-EM densities (mesh) of representative regions of the APC subunits, superimposed with respective atomic models (stick). Note the different densities of the residues labelled in red.



Extended Data Figure 11 | Plausible energy transfer between the individual hexamer and rods; between rods Rd and Re and the core; and between rods. a, Distances (Å) between bilins at the interfaces of the individual hexamer Ha and rod Rd (left), and Hb and Ra (right). **b**, Plausible energy transfer paths from bilins on Rd and Re′ to the core. Inset panels indicate the distances (Å) between two adjacent bilins. **c**, Plausible energy transfer paths between bilins on Rd and Rb. Inset panels indicate the distances (Å) between two adjacent bilins. **d**, Plausible energy transfer from bilins on Re to Ra or Rc. Inset panels indicate the distances (Å) between two adjacent bilins.

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

	Titan Krios (EMD-6758---EMD-6769) (PDB 5Y6P)
Data collection and processing	
Magnification	22,500
Voltage (kV)	300
Electron exposure (e–/Å^2)	50 (32 frames) / 20 (13 frames)
Defocus range (μm)	-1.4 - -2.3
Pixel size (Å)	1.32
Exposure time (s)	8
Symmetry imposed	C2
Number of images	7,477
Number of frames/image	32
Initial particle images (no.)	471,482
Final particle images (no.)	53,403
Map resolution (Å)	3.5
FSC threshold	0.143
Map resolution range (Å)	3.4 – 7
Refinement	
Map sharpening *B* factor (Å^2)	-70
Model composition	
Non-hydrogen atoms	1,147,832
Protein residues	154,768
Ligands	2048
R.m.s. deviations	
Bond lengths (Å)	0.014
Bond angles (°)	1.908
Validation	
MolProbity score	2.33
Rotamer outlier(%)	1.6
Ramachandran plot	
Favored (%)	94.8
Allowed (%)	4.61
Disallowed (%)	0.6

Extended Data Table 2 | Summary of model validation for the phycobilisome components

Molecule*	MolProbity Scores	Ramachandran plot statistics (%)			RMS deviations	
		Preferred	Allowed	Outlier	Bonds Length (Å)	Bonds Angles (°)
Core	2.02	95.47	4.13	0.4	0.0221	2.08
Ra/Ra'	2.24	95.02	4.46	0.52	0.0152	1.80
Rb/Rb'	2.30	95.04	4.42	0.54	0.0152	1.84
Rc/Rc'	2.14	95.16	4.49	0.35	0.0145	1.67
Rd/Rd'	1.94	94.90	4.49	0.61	0.0134	1.58
Re/Re'	2.28	94.48	4.16	1.36	0.0164	1.88
Rf/Rf'	2.20	95.63	3.98	0.39	0.0098	1.36
Rg/Rg'	2.01	93.95	5.83	0.22	0.0146	1.69
Ha/Ha'	2.26	94.31	5.43	0.26	0.0124	1.53
Hb/Hb'	2.24	93.75	5.99	0.26	0.0127	1.57
Hc/Hc'	2.41	95.10	4.13	0.77	0.0069	1.17
Hd/Hd'	2.12	96.27	3.23	0.50	0.0125	1.54
He/He'	2.27	95.26	4.7	0.04	0.0059	1.06
M1	2.70	89.04	9.68	1.28	0.0175	1.86
M2	1.89	92.99	6.66	0.35	0.0112	1.47
M3	1.79	92.94	6.54	0.52	0.0107	1.41

*Core contains all α-subunits, β-subunits in core, and L_C/L_C′, L_{CM}/L_{CM}′, L_{RC}4/L_{RC}4′, L_{RC}5/L_{RC}5′ and L_{RC}6/L_{RC}6′; each rod (Ra/Ra′ through Rg/Rg′) contains all α-subunits, β-subunits and linker proteins in the rod; each individual hexamer (Ha/Ha′–He/He′) contains all α-subunits, β-subunits and linker proteins in the hexamer; M1 contains individual α-subunits S1/S1′ and individual β-subunits S2/S2′–S4/S4′; M2 contains individual β-subunits S5/S5′–S8/S8′; M3 contains individual β-subunits S9/S9′–S12/S12′.

nature research

Corresponding Author: Sen-Fang Sui

Date: Aug 10, 2017

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   Amount of cryo-EM micrographs collected was 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.

2. Data exclusions

   Describe any data exclusions.

   No data have been excluded.

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   All attempts at replication were successful. The cryo-EM data collections were consistent throughout the experiments.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   N/A

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   N/A

   Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
☒	☐	A statement indicating how many times each experiment was replicated
☒	☐	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☒	☐	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☒	☐	The test results (e.g. *p* values) given as exact values whenever possible and with confidence intervals noted
☒	☐	A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☒	☐	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Relion v1.4, EMAN2.1, MOTIONCORR v1, CTFFIND3, ResMap v1.1, CHIMERA v1.10, Coot v0.8, PHENIX v1.11, Pymol v1.8, REFMAC5, Trinity v2.2, TransDecoder v2.0, Blast2Go v3.1, FASTX-Toolkit v0.0.13, SOAPdenovo2 r240, GeneMark-ES v4.33

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	N/A

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	N/A
b. Describe the method of cell line authentication used.	N/A
c. Report whether the cell lines were tested for mycoplasma contamination.	N/A
d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	N/A

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	N/A

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	N/A