Structural basis of energy transfer in *Porphyridium purpureum* phycobilisome

https://doi.org/10.1038/s41586-020-2020-7

Received: 15 July 2019

Accepted: 19 December 2019

Published online: 19 February 2020

Check for updates

Jianfei Ma[1,3], Xin You[1,3], Shan Sun[1,3], Xiaoxiao Wang[2], Song Qin[2] & Sen-Fang Sui[1]✉

Photosynthetic organisms have developed various light-harvesting systems to adapt to their environments[1]. Phycobilisomes are large light-harvesting protein complexes found in cyanobacteria and red algae[2–4], although how the energies of the chromophores within these complexes are modulated by their environment is unclear. Here we report the cryo-electron microscopy structure of a 14.7-megadalton phycobilisome with a hemiellipsoidal shape from the red alga *Porphyridium purpureum*. Within this complex we determine the structures of 706 protein subunits, including 528 phycoerythrin, 72 phycocyanin, 46 allophycocyanin and 60 linker proteins. In addition, 1,598 chromophores are resolved comprising 1,430 phycoerythrobilin, 48 phycourobilin and 120 phycocyanobilin molecules. The markedly improved resolution of our structure compared with that of the phycobilisome of *Griffithsia pacifica*[5] enabled us to build an accurate atomic model of the *P. purpureum* phycobilisome system. The model reveals how the linker proteins affect the microenvironment of the chromophores, and suggests that interactions of the aromatic amino acids of the linker proteins with the chromophores may be a key factor in fine-tuning the energy states of the chromophores to ensure the efficient unidirectional transfer of energy.

Light absorption is the first step of photosynthesis. The membrane-extrinsic soluble phycobilisomes (PBSs) are responsible for the majority of light capture in cyanobacteria and red algae[2–4]. PBSs are composed of phycobiliproteins (PBPs) and linker proteins[2], and sunlight is absorbed and the energy transferred by open-chain tetrapyrrole chromophores that covalently bind to PBPs and some linker proteins[2,6]. A heterodimer of two different PBP subunits (α- and β subunits) assembles into a ring-shaped $(\alpha\beta)_3$ trimer, which serves as the basic unit for the PBS assembly[2]. The typical PBS consists of several peripheral rods surrounding the central core[1,7]. Solar photonic energy absorbed by the peripheral rods can be rapidly funnelled to the core[1] and eventually to the terminal emitters—chromophores of the core–membrane linker protein (L_{CM})[8–10] or allophycocyanin D (ApcD)[11,12]—then transferred to photosystems I and II[2,10,13,14].

Four morphological types of PBS have been observed: hemidiscoidal[15–20], hemiellipsoidal[21,22], block-type[23] and bundle-type[24]. We recently solved the structure of the block-shaped PBS from the red alga *G. pacifica* at 3.5 Å resolution, which provided the detailed architecture of the intact PBS[5]. However, to our knowledge there have been no high-resolution structures reported for other morphological types of PBS. Moreover, although we determined the locations of all chromophores of the *G. pacifica* PBS, owing to resolution limitations we could not reveal how the energies of the bilins are modulated by their surroundings. To address these questions, we resolved the structure of a typical hemiellipsoidal-shaped PBS[21,22,25] from *Porphyridium purpureum*, one of the few unicellular red algae and a widely used model alga[26–28]. The resulting structure, determined by cryo-electron microscopy (cryo-EM) at 2.82 Å resolution, reveals how the linker proteins affect the microenvironments of chromophores.

Overall structure

Intact PBSs were purified from *P. purpureum* and their subunit composition and spectroscopic properties are shown in Extended Data Fig. 1. The overall resolution of the resulting structure is 2.82 Å, with a higher resolution of 2.68 Å for the core region (Fig. 1a, Extended Data Fig. 2, Extended Data Table 1). Some long loops in L_{CM}, which are absent in the electron microscopy map of the *G. pacifica* PBS, are clearly resolved in this reconstruction (Fig. 1b). The different types of bilins can be unambiguously assigned on the basis of the densities and the dihedral angles in combination with the results of published biochemical analysis[29–35] that classified the phycocyanin in *P. purpureum* as R-phycocyanin, containing one phycoerythrobilin (PEB) and one phycocyanobilin (PCB) on the β subunit[34] (Fig. 1c, Extended Data Fig. 3). In total, we built 706 protein subunits comprising 528 phycoerythrin subunits, 72 phycocyanin subunits, 46 allophycocyanin subunits and 60 linker proteins, and we assigned 1,598 chromophores (Extended Data Table 2a).

The two-fold symmetric PBS resembles an opened fan from the face view and has an oval outline from the top view, with approximate dimensions of 610 Å (length), 390 Å (height) and 380 Å (thickness) (Fig. 1d). It contains 14 peripheral rods surrounding a pyramidal-shaped core (Extended Data Fig. 4a). The core contains one top cylinder (B) formed by two allophycocyanin trimers stacked back to back, and two bottom

[1]State Key Laboratory of Membrane Biology, Beijing Advanced Innovation Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing, China. [2]Yantai Institute of Coast Zone Research, Chinese Academy of Sciences, Yantai, China. [3]These authors contributed equally: Jianfei Ma, Xin You, Shan Sun. ✉e-mail: suisf@mail.tsinghua.edu.cn

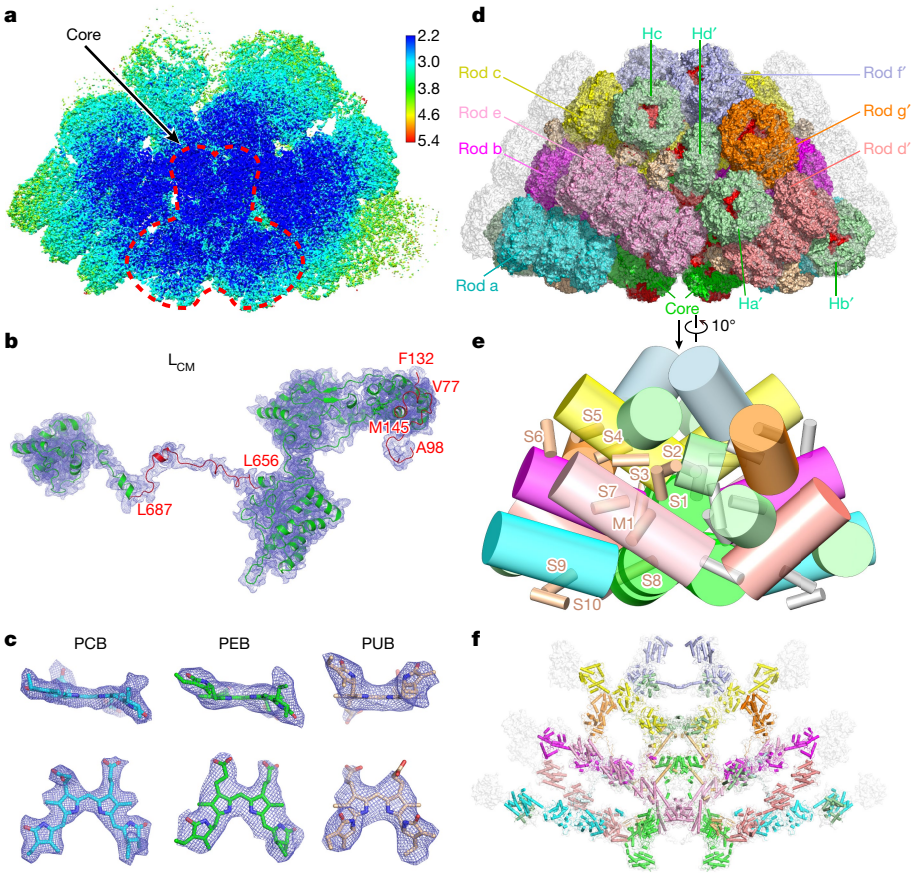


Fig. 1 | Overall architecture of the PBS from *P. purpureum*. **a**, Local resolution map of the PBS from *P. purpureum*. The map was estimated with ResMap and generated in Chimera. **b**, The density (mesh) for the linker protein L_{CM} superimposed with its atomic model (cartoon). Three loops (red)—V77 to A98, F132 to M145 and L656 to L687—are clearly resolved. **c**, The densities (mesh) of the representative PCB, PEB and phycourobilins (PUB) bilins (stick representation) show their different coplanarities. **d**, Overall structure of the PBS shown in surface representation. The rods are shown in different colours, and the core is coloured green. All extra hexamers, individual monomers and β subunits, and linker proteins are coloured light green, wheat and red, respectively. The superimposed structure of the *G. pacifica* PBS is coloured grey. **e**, Schematic showing the distribution of the individual phycoerythrin (αβ) monomers and phycoerythrin β subunits. One half (M1 and S1–S10) are coloured wheat and the other half are grey. **f**, Structures of all linker proteins shown in cartoon representation, from the same view as in **d**. They are the same colours as the hexamers in which they are located. The superimposed structures of linker proteins of the *G. pacifica* PBS are coloured grey.

cylinders (A and A′), each of which is assembled by one $(\alpha\beta)_6$ allophycocyanin hexamer and one allophycocyanin trimer (Extended Data Fig. 4b). The rods are categorized into two types according to their PBP composition: type I rods (Ra/Ra′, Rb/Rb′ and Rc/Rc′) are composed of both phycoerythrin and phycocyanin, whereas type II rods (Rd/Rd′, Re/Re′, Rf/Rf′ and Rg/Rg′) are composed entirely of phycoerythrin (Extended Data Fig. 4d). Except for rods Rf/Rf′ and Rg/Rg′, each of which contains two phycoerythrin hexamers in both PBSs, the number of phycoerythrin hexamers in each of the remaining rods is one fewer in the *P. purpureum* PBS than in the *G. pacifica* PBS (Extended Data Fig. 4c, d). The *P. purpureum* PBS also contains extra phycoerythrin hexamers; however, the hexamers He/He′, which are located near the surface of the outermost hexamer of rods Rb/Rb′ and Rc/Rc′ in the *G. pacifica* PBS, are absent owing to the short lengths of these rods in the *P. purpureum* PBS (Fig. 1d, Extended Data Fig. 4c). There are 2 individual phycoerythrin (αβ) monomers M1 (M1′) and 20 individual phycoerythrin β subunits S1–S10 (S1′–S10′) interspersed throughout the whole PBS (Fig. 1e). These components fill the empty spaces outside the rods, core and extra hexamers, and may stabilize the PBS. The *P. purpureum* PBS is aligned well with the *G. pacifica* PBS, but has a smaller size owing to the reduced number of phycoerythrin hexamers, indicating a similar overall organization of the rods and core (Fig. 1d, Extended Data Fig. 4c). The molecular mass of the *P. purpureum* PBS is 14.7 MDa, which is less than that of the *G. pacifica* PBS (18.0 MDa) after considering the molecular masses of the chromophores.

The skeleton formed by the linker proteins is very similar in the PBSs from both *P. purpureum* and *G. pacifica* (Fig. 1f), and both contain 17 types of linker protein. Superimposing the two PBSs indicate that they share very high structural conservation, except for the rod linker protein L_R6 (Extended Data Fig. 4e). L_R6 of the *P. purpureum* PBS contains the Pfam00427 domain, instead of the CBDγ domain that is present in $L_R\gamma6$ of the *G. pacifica* PBS; this is in accordance with the overall tendency for the *P. purpureum* PBS to contain fewer bilins than the *G. pacifica* PBS. The roles of linker proteins in the assembly of the PBS—such as the sequential interactions between them to form the skeleton, the extensive contacts between them and the hexamers, and the α-helix-mediated interactions between L_{RC} proteins and the core—are common between these two PBSs, highlighting their evolutionary conservation (Extended Data Fig. 5).

There are 120 PCBs, 1,430 PEBs and 48 phycourobilins in *P. purpureum* PBS (Extended Data Table 2a). The phycourobilin content in the *P. purpureum* PBS is considerably lower than that in the *G. pacifica* PBS; this is because all phycourobilins in the *P. purpureum* PBS originate solely from the $L_R\gamma$ proteins, whereas in the *G. pacifica* PBS—besides the $L_R\gamma$ proteins—all phycoerythrin β subunits also contain phycourobilin[36]. The lower phycourobilin content and the reduced number of total bilins

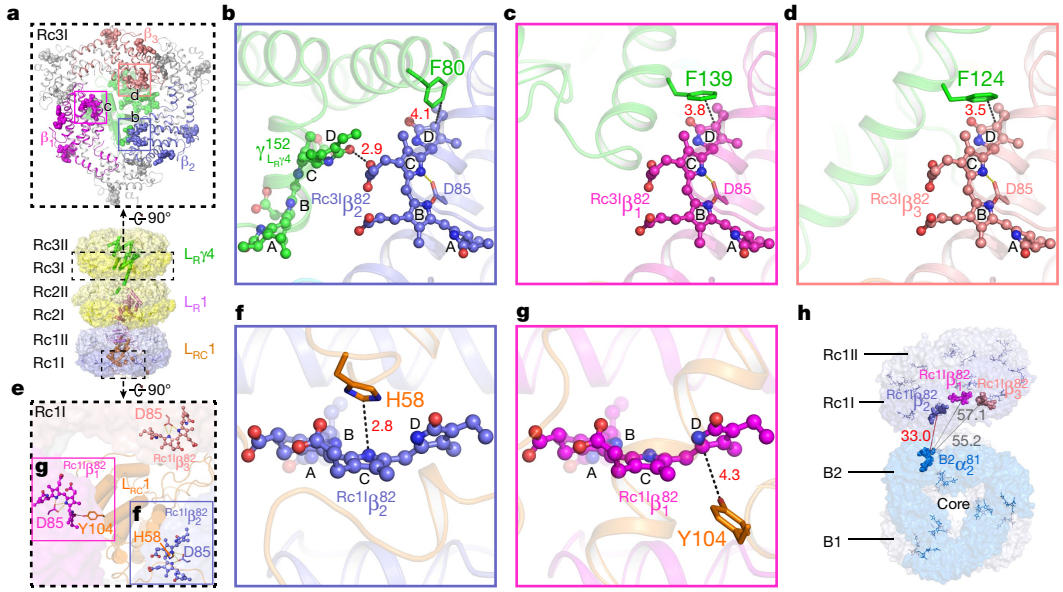


Fig. 2 | Interactions of the linker proteins $L_R\gamma$s and L_{RC}s with chromophores in the rod Rc. a, Bottom, overall structure of the rod Rc with the hexamers shown in surface representation and the linker proteins shown in cartoon representation. Top, structure of the layer Rc3I. Proteins and bilins are shown in cartoon and sphere representations, respectively. Three β subunits are coloured differently and the β82 PEBs are boxed and analysed in detail in **b**–**d**. **b**, The interactions between the residue F80 and the bilin $\gamma^{152}_{L_R\gamma4}$ from $L_R\gamma4$ with the bilin $^{Rc3I}\beta^{82}_2$. **c**, The interaction between F139 from $L_R\gamma4$ and the bilin $^{Rc3I}\beta^{82}_1$. **d**, The interaction between F124 from $L_R\gamma4$ and the bilin $^{Rc3I}\beta^{82}_3$. **e**, A focused view of the structure of the layer Rc1I showing the central triangle area. PBPs, the linker protein, bilins and residues are shown in surface, cartoon, ball-and-stick and stick representations, respectively. Two β82 PCBs are boxed and analysed in detail in **f** and **g**. **f**, The strong interaction between H58 from $L_{RC}1$ and the bilin $^{Rc1I}\beta^{82}_2$. **g**, The interaction between Y104 from $L_{RC}1$ and the bilin $^{Rc1I}\beta^{82}_1$. **h**, $^{Rc1I}\beta^{82}_2$ has the shortest distance (33.0 Å, red line) between the rod and the core compared to $^{Rc1I}\beta^{82}_1$ (57.1 Å, grey line) and $^{Rc1I}\beta^{82}_3$ (55.2 Å, grey line).

in *P. purpureum* PBS are consistent with the fact that *P. purpureum* live at the sea surface[37], where the light intensity is higher compared with that beneath the sea surface where *G. pacifica* are found[37].

Interactions of $L_R\gamma$s with chromophores

The $(\alpha\beta)_3$ trimers of the phycocyanins, phycoerythrins and allophycocyanins have very similar ring-like structures, in which the central cavity is a common feature. Three β82 chromophores are located near to the inner cavity and are subjected to interactions with linker proteins[38] (Fig. 2a). The trimers Rc3I and Rd3I—from the type I rod Rc and the type II rod Rd, respectively—are used here to illustrate how the rod linkers $L_R\gamma4$ and $L_R\gamma5$ interact with the β82 chromophores (Fig. 2). Each of the β82 PEBs of Rc3I (denoted $^{Rc3I}\beta^{82}_1$, $^{Rc3I}\beta^{82}_2$ and $^{Rc3I}\beta^{82}_3$) is bound by two hydrogen bonds formed between the nitrogen atoms of the pyrrole rings B and C and the carboxyl group of the D85 residue of the β subunit (Fig. 2b–d), in agreement with the crystal structure of R-phycoerythrin[39]. In particular, three aromatic residues of $L_R\gamma4$ (F80, F139 and F124) are located close to rings D of $^{Rc3I}\beta^{82}_1$, $^{Rc3I}\beta^{82}_2$ and $^{Rc3I}\beta^{82}_3$, respectively, which will stabilize ring D and may expand the conjugation of the system owing to π–π interactions (Fig. 2b–d). Moreover, an extra PEB from $L_R\gamma4$ ($\gamma^{152}_{L_R\gamma4}$) is adjacent—and very close—to the chromophore $^{Rc3I}\beta^{82}_2$: the distance between their nearest two atoms is only 2.9 Å (Fig. 2b). Therefore, the chromophore pair may further downgrade the energy level of $^{Rc3I}\beta^{82}_2$ owing to excited-state coupling[39,40], with the result that $^{Rc3I}\beta^{82}_2$ is probably at a lower energy level compared with that of $^{Rc3I}\beta^{82}_1$ and $^{Rc3I}\beta^{82}_3$. Notably, the trimer Rd3I in the type II rod Rd displays similar structural features: one aromatic residue is close to each of the β82 PEBs, and an additional bilin from the linker $L_R\gamma5$ ($\gamma^{135}_{L_R\gamma5}$) also resides close to the $^{Rd3I}\beta^{82}_2$ (Extended Data Fig. 6a–d). Moreover, structural superimposition reveals that the $L_R\gamma$ linker proteins in the outmost hexamers of various rods of PBSs from both *P. purpureum* and *G. pacifica* also have similar structures (Extended Data Fig. 7a, b). These key aromatic residues, and the cysteine residues that are used to link the bilins, are well conserved in red algae (Extended Data Fig. 7c). The structural features of the interaction between $L_R\gamma$ and β82 PEBs therefore suggest that $^{3I}\beta^{82}_2$PEB is in the lowest energy state among the three β82 PEBs, and that energy migration through $^{3I}\beta^{82}_2$ PEB could be the major route by which to pass energy through the rod.

Interactions of L_{RC}s with chromophores

Energy is then transferred along the rods to the triangular area of the core-proximal hexamer (Fig. 2e). Two types of rod use different rod–core linkers to associate with the core. For Rc—which uses the linker $L_{RC}1^c$—a heterocyclic residue (H58) from $L_{RC}1^c$ is located close to $^{Rc1I}\beta^{82}_2$, with a minimum distance of 2.8 Å (Fig. 2f). The pyrrole group of H58 can form a strong π–π interaction with rings B and C of $^{Rc1I}\beta^{82}_2$ (ref. [41]). However, just one aromatic residue (Y104) from $L_{RC}1^c$ forms a relatively weak π–π interaction with ring D of $^{Rc1I}\beta^{82}_1$—because of the longer distance (4.3 Å) compared with that between H58 and $^{Rc1I}\beta^{82}_2$—and no aromatic residues interact with $^{Rc1I}\beta^{82}_3$ (Fig. 2e, g). Therefore, modified by the specific surroundings, $^{Rc1I}\beta^{82}_2$ may be in the lowest-energy state among the three β82 chromophores. Moreover, $^{Rc1I}\beta^{82}_2$ has the shortest distance to the core compared with $^{Rc1I}\beta^{82}_1$ and $^{Rc1I}\beta^{82}_3$ (Fig. 2h), which further suggests that it may act as an energy-transit station, converging the energy absorbed by the rod and transferring it to the core. Similar situations are found for another two rods of type I and the three type I rods in *G. pacifica* (Extended Data Fig. 7d). This histidine residue is conserved completely across different red algal and cyanobacterial species, which is indicative of its functional importance (Extended Data Fig. 7e).

For Rd—which uses $L_{RC}2$—two aromatic residues from $L_{RC}2$ form parallel-displaced and T-shaped π–π interactions with $^{Rd1I}\beta^{82}_2$, respectively (Extended Data Fig. 6f). By contrast, each of another two β82 PEBs interacts with only one aromatic residue (Extended Data Fig. 6g, h). Superimposition of the L_{RC} proteins ($L_{RC}2$ and $L_{RC}3$) of type II rods from both *P. purpureum* and *G. pacifica* shows that two such aromatic

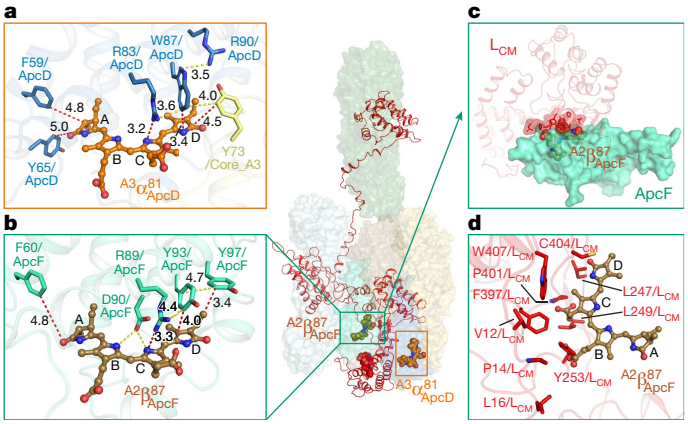


Fig. 3 | The bilins of ApcD and ApcF and their surrounding residues. a, The surrounding residues of the bilin $^{A3}\alpha^{81}_{ApcD}$. The residues and bilin are shown in stick and ball-and-stick representations, respectively. **b**, The interactions between $^{A2}\beta^{87}_{ApcF}$ and ApcF. The residues and bilin are shown in stick and ball-and-stick representations, respectively. **c**, $^{A2}\beta^{87}_{ApcF}$ is buried by a contiguous hydrophobic cap formed by the linker protein L_{CM}. L_{CM} is shown in cartoon representation in red and the cap is displayed in surface representation. $^{A2}\beta^{87}_{ApcF}$ is shown in sphere representation. **d**, The interactions between the cap and $^{A2}\beta^{87}_{ApcF}$. The hydrophobic residues in the cap are shown in stick representation in red.

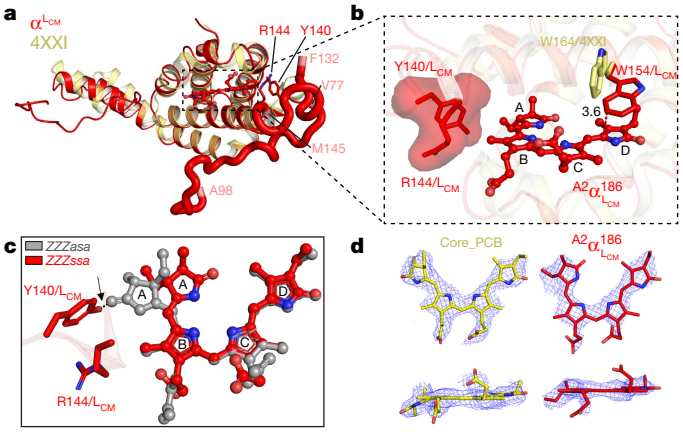


Fig. 4 | The conformation of PCB in L_{CM}. a, Structural alignment of the α subunit of L_{CM} (α^{LCM}) in the PBS and the recombinant α subunit of L_{CM} (PDB: 4XXI). α^{LCM} and 4XXI are coloured red and pale yellow, respectively. The bilin in α^{LCM} is shown in ball-and-stick representation. Two extra loops (V77–A98 and F132–M145) are displayed in sausage representation. **b**, Structural differences between α^{LCM} and 4XXI. Two different conformations of tryptophan (W154 in L_{CM} and W164 in 4XXI) are shown in stick representation in red and yellow. Y140 and R144 from one loop are shown in surface representation in red. **c**, Steric clashing is observed between Y140 from L_{CM} and the *ZZZasa* configuration of $^{A2}\alpha^{186}_{LCM}$ (grey), but is absent between Y140 and the *ZZZssa* configuration of $^{A2}\alpha^{186}_{LCM}$ (red). **d**, Cryo-EM densities (mesh) of the bilins (stick) in the α subunit of the core (Core_PCB, yellow) and α^{LCM} ($^{A2}\alpha^{186}_{LCM}$, red) show the enhanced coplanarity of rings A and B of $^{A2}\alpha^{186}_{LCM}$.

residues exist in all cases (Extended Data Fig. 7f). Sequence alignment also shows that these aromatic residues are conserved, which suggests their importance in fine-tuning the energies of the bilins (Extended Data Fig. 7g).

Key chromophores in the core

Previous studies have shown that three PCB chromophores in ApcD, ApcF and the α subunit of L_{CM} ($^{A3}\alpha^{81}_{ApcD}$, $^{A2}\beta^{87}_{ApcF}$ and $^{A2}\alpha^{186}_{LCM}$) perform critical functions in energy transfer in the core. However, why each of these chromophores have unique functions remains to be clarified. Here we use our high-resolution structure to analyse the immediate surroundings of these key core chromophores in their native states.

Functionally, ApcD is the main protein responsible for energy transfer to photosystem I[13]. In our structure, two aromatic residues—W87 from ApcD and Y73 from the β subunit of the core trimer A3—form T-shaped and parallel-displaced π–π interactions with $^{A3}\alpha^{81}_{ApcD}$, respectively; this enhances the tight fitting of ring D (Fig. 3a), which is consistent with the crystal structure of ApcD from *Synechocystis* PCC 6803 (PDB: 4PO5)[12]. Notably, we observed that W87 was surrounded by R83 and R90 from ApcD and Y73 from the core A3, which provided two cation–π interactions and one T-shaped π–π interaction to W87, respectively (Fig. 3a). It can therefore be inferred that the presence of these three residues is necessary to stabilize the orientation of W87, which is critical for the conformation of $^{A3}\alpha^{81}_{ApcD}$. In addition, more π-related interactions between residues and $^{A3}\alpha^{81}_{ApcD}$ are extracted from the high-resolution structure of the entire PBS. The cationic side chain of R83 extends to the top of ring C of $^{A3}\alpha^{81}_{ApcD}$, forming the cation–π interaction[42]. F59 and Y65 may contribute two additional π–π interactions to ring A of $^{A3}\alpha^{81}_{ApcD}$ (Fig. 3a). We then superimposed ApcDs from *P. purpureum*, *G. pacifica* and *Synechocystis* PCC 6803 with the α subunit of the core A3, and found that W87, F59 and Y65 are common to all ApcD molecules, although Y65 was replaced by V65 in the A3 α subunit (Extended Data Fig. 8a).

ApcF plays a crucial role in energy migration to the terminal chromophore of L_{CM}[14,43]. Analysis of the PCB pocket of ApcF showed that the positively charged R89 formed one cation–π interaction with ring C of $^{A2}\beta^{87}_{ApcF}$, and Y93 and Y97 formed one T-shaped π–π and one parallel-displaced π–π interaction with ring D, respectively. Moreover, R89, Y93 and Y97 interact with each other by either cation–π or π–π interactions (Fig. 3b). The superimposition of ApcF molecules from *P. purpureum* and *G. pacifica* with the β subunit from the core A2 shows that these three residues exist in the same position in all proteins, suggesting their importance for the stability of the PCB (Extended Data Fig. 8b). Except for these common features, an aromatic residue from ApcF (F60) is located above ring A of $^{A2}\beta^{87}_{ApcF}$ (Fig. 3b); an aromatic residue was also found in the ApcF of *G. pacifica* (Y60), whereas this residue is replaced by L60 in other β subunits (Extended Data Fig. 8b). This aromatic residue may therefore form additional π–π interactions with $^{A2}\beta^{87}_{ApcF}$ and hence lower its energy. Another notable feature in our structure of the complete PBS is that L_{CM} is directly involved in the interaction with $^{A2}\beta^{87}_{ApcF}$. Several hydrophobic residues of L_{CM} are located at the ApcF/L_{CM} interface and within 4 Å of $^{A2}\beta^{87}_{ApcF}$—this creates a contiguous hydrophobic 'cap' that buries $^{A2}\beta^{87}_{ApcF}$ (Fig. 3c, d, Extended Data Fig. 8c), which can enhance the stability of the conformation of $^{A2}\beta^{87}_{ApcF}$. The residues present around the chromophores from other β subunits ($^{A2}\beta^{81}_{1}$ and $^{A2}\beta^{81}_{2}$) are less hydrophobic (Extended Data Fig. 8d). Similarly, this hydrophobic cap is also found around the PCB of ApcF from the *G. pacifica* PBS (Extended Data Fig. 8e).

The terminal chromophore PCB in L_{CM} ($^{A2}\alpha^{186}_{LCM}$) exhibits fluorescence with similar emission wavelengths to those of the intact PBS, and is at a lower energy than the upstream PCBs[10]. Although the overall structure of the α-subunit domain of L_{CM} (α^{LCM}) overlapped well with the recombinant α_{LCM} (PDB: 4XXI)[10] (Fig. 4a), some differences and new structural information are revealed in this study of the native PBS. In structure 4XXI, two different conformations of W164 are found above $^{A2}\alpha^{186}_{LCM}$: one is parallel to ring D and the other is nearly perpendicular[10]. However, in our structure, W154 at the same position displays only one conformation, parallel to ring D (Fig. 4b)—this indicates that the native L_{CM} has a unique preference for how the side chain of such a residue is positioned. Moreover, compared with 4XXI, two extra loops (V77–A98 and F132–M145) were resolved in our α^{LCM} (Figs. 1c, 5a). Two residues in this loop, Y140 and R144, are in direct contact (less than 4 Å) with $^{A2}\alpha^{186}_{LCM}$

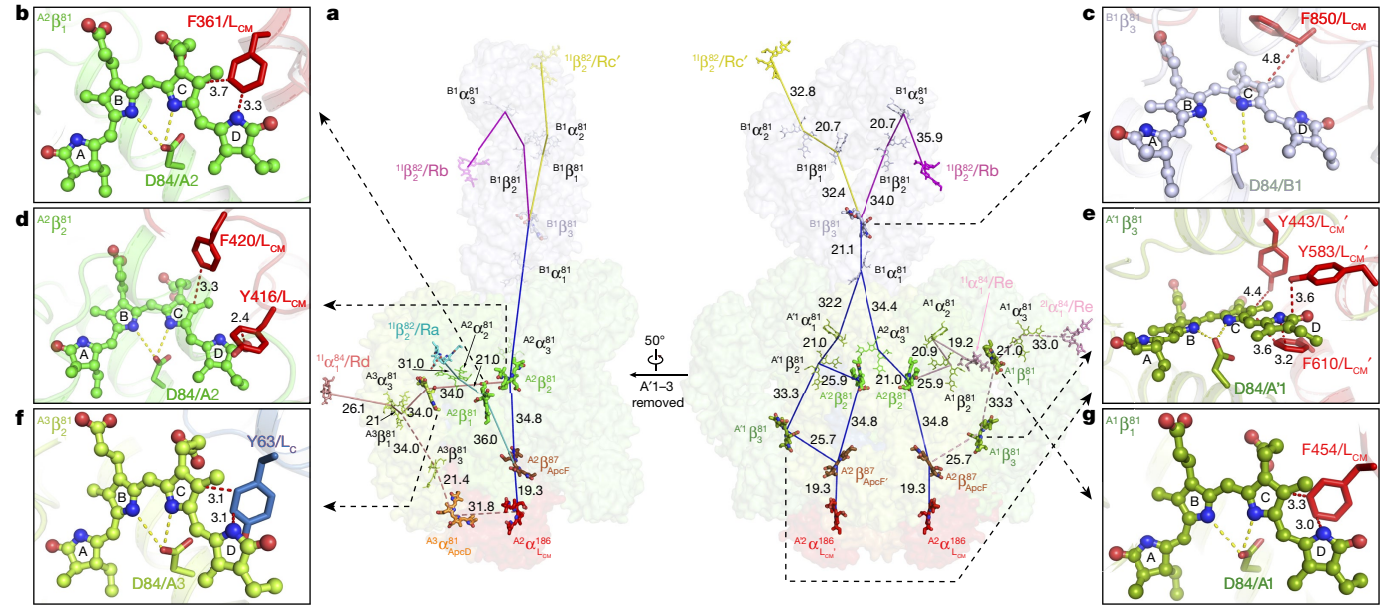


Fig. 5 | Key bilins in the core. a, Different views of the core with the bilins in different core layers shown in different colours. Bilins in $\alpha^{L_{CM}}$ ($\alpha^{L_{CM}'}$), ApcD (ApcD′) and ApcF (ApcF′) are coloured red, orange and brown, respectively. Bilins in rods are the same colours as the rods in which they are located according to the colouring scheme in Fig. 1d. Bilins are shown in stick representation, and key bilins are shown as thicker sticks. The numbers indicate the distances (Å) between the bilin pairs. **b**, Interaction between F361 from L_{CM} and the bilin $^{A2}\beta_1^{81}$. **c**, Interaction between F850 from L_{CM} and the bilin $^{B1}\beta_3^{81}$. **d**, Interactions between Y416 and F420 from L_{CM} with the bilin $^{A2}\beta_2^{81}$. **e**, Interactions between Y443, Y583 and F610 from L_{CM}' with the bilin $^{A'1}\beta_3^{81}$. **f**, Interaction between Y63 from L_C and the bilin $^{A3}\beta_2^{81}$. **g**, The interaction between F454 from L_{CM} and the bilin $^{A1}\beta_1^{81}$.

(Fig. 4b). In particular, the side chain of Y140 is oriented towards the inside of the $^{A2}\alpha_{L_{CM}}^{186}$ pocket (Fig. 4b). In such a conformation there will exist a steric clash between Y140 and the *ZZZasa* configuration of $^{A2}\alpha_{L_{CM}}^{186}$, thus providing a driving force for the formation of the *ZZZssa* configuration (Fig. 4c, Extended Data Fig. 8f). Comparison of $\alpha^{L_{CM}}$ with five other similar subunits reveals that the orientation of this tyrosine in $\alpha^{L_{CM}}$ is opposite to that in other subunits (Extended Data Fig. 8g). Therefore, Y140 of L_{CM} is another factor that causes $^{A2}\alpha_{L_{CM}}^{186}$ to uniquely adopt the *ZZZssa* geometry, which exhibits enhanced coplanarity of rings A and B compared with other PCBs in the α subunits of the core (Fig. 4d, Extended Data Fig. 3c).

In addition to $^{A3}\alpha_{ApcD}^{81}$, $^{A2}\beta_{ApcF}^{87}$ and $^{A2}\alpha_{L_{CM}}^{186}$, the energy states of some other chromophores in the core are subjected to modification by the linker proteins. The shortest distance between rod Ra and the core was found between $^{1l}\beta_2^{82}$/Ra and $^{A2}\alpha_2^{81}$ (31 Å) (Fig. 5a), which may facilitate energy transfer from Ra to the core. The bilin nearest to $^{A2}\alpha_2^{81}$ is $^{A2}\beta_1^{81}$, rings C and D of which form a parallel-displaced π–π interaction with F361 from L_{CM} (Fig. 5b), and thus may mediate energy transfer to $^{A2}\beta_{ApcF}^{87}$. The energy absorbed by rods Rb and Rc′ may travel through the core layer B1 to $^{B1}\alpha_1^{81}$, the nearest bilin to the basal cylinders (Fig. 5a). Bilin $^{B1}\beta_3^{81}$ may play an essential role in this process because it is subjected to a parallel-displaced π–π interaction with F850 from L_{CM} (Fig. 5c). In the basal cylinders, the two bilins $^{A2}\beta_2^{81}$/$^{A'2}\beta_2^{81}$ and $^{A1}\beta_3^{81}$/$^{A'1}\beta_3^{81}$—which are adjacent to the bilins on ApcF/ApcF′ and separated from them by 34.8 Å and 25.7 Å, respectively—have special microenvironments (Fig. 5a). $^{A2}\beta_2^{81}$ is affected by the π–π interactions with Y416 and F420 from L_{CM} (Fig. 5d) and $^{A'1}\beta_3^{81}$ is affected by several π–π interactions between its rings C and D with Y443, Y583 and F610 from L_{CM}' (Fig. 5e)—this suggests that these two bilins may facilitate energy flow to $^{A2}\beta_{ApcF}^{87}$/$^{A'2}\beta_{ApcF}^{87}$. The bilin pair $^{RdII}\alpha_1^{84}$ and $^{A3}\alpha_3^{81}$ shows the shortest distance (26.1 Å) between rod Rd and the core (Fig. 5a). The bilin $^{A3}\beta_2^{81}$ may mediate further energy transfer because it is subjected to π–π interactions with Y63 from L_C (Fig. 5f). For Re, both hexamers Re1 and Re2 attach to the core layer A1; as such, the energy could flow from either Re1 or Re2 to the core. The shortest distance between Re2 and the core is from $^{2l}\alpha_1^{84}$/Re to $^{A1}\alpha_3^{81}$ (33 Å) (Fig. 5a). The energy could then travel via the $^{A1}\beta_1^{81}$, because this bilin has the shortest distance to $^{A1}\alpha_3^{81}$ and is affected by F454 from L_{CM} through two parallel-displaced π–π interactions with rings C and D (Fig. 5g). Together, our results show that core linker proteins are extensively involved in the modulation of the energy states of core bilins to ensure the efficient unidirectional transfer of energy. These findings provide the framework for a detailed examination of energy transfer in future studies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2020-7.

1. Bryant, D. A. & Canniffe, D. P. How nature designs light-harvesting antenna systems: design principles and functional realization in chlorophototrophic prokaryotes. *J. Phys. B* **51**, 033001 (2018).
2. Adir, N., Dines, M., Klartag, M., McGregor, A. & Melamed-Frank, M. in *Complex Intracellular Structures in Prokaryotes* (ed. Shively, J. M.) 47–77 (Springer, 2006).
3. Glazer, A. N. Light guides. Directional energy transfer in a photosynthetic antenna. *J. Biol. Chem.* **264**, 1–4 (1989).
4. Adir, N., Bar-Zvi, S. & Harris, D. The amazing phycobilisome. *Biochim. Biophys. Acta Bioenerg.* https://doi.org/10.1016/j.bbabio.2019.07.002 (2019).
5. Zhang, J. et al. Structure of phycobilisome from the red alga *Griffithsia pacifica*. *Nature* **551**, 57–63 (2017).
6. Sidler, W. A. in *The Molecular Biology of Cyanobacteria* (ed. Bryant, D. A.) 139–216 (Springer, 1994).
7. Singh, N. K., Sonani, R. R., Rastogi, R. P. & Madamwar, D. The phycobilisomes: an early requisite for efficient photosynthesis in cyanobacteria. *EXCLI J.* **14**, 268–289 (2015).
8. Gao, X. et al. Molecular insights into the terminal energy acceptor in cyanobacterial phycobilisome. *Mol. Microbiol.* **85**, 907–915 (2012).
9. 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).
10. Tang, K. et al. The terminal phycobilisome emitter, L_{CM}: a light-harvesting pigment with a phytochrome chromophore. *Proc. Natl Acad. Sci. USA* **112**, 15880–15885 (2015).

11. Glazer, A. N. & Bryant, D. A. Allophycocyanin B (λ_{max} 671, 618 nm): a new cyanobacterial phycobiliprotein. *Arch. Microbiol.* **104**, 15–22 (1975).
12. 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* **70**, 2558–2569 (2014).
13. 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).
14. Ashby, M. K. & Mullineaux, C. W. The role of ApcD and ApcF in energy transfer from phycobilisomes to PS I and PS II in a cyanobacterium. *Photosynth. Res.* **61**, 169–179 (1999).
15. Bryant, D. A., Guglielmi, G., de Marsac, N. T., Castets, A.-M. & Cohen-Bazire, G. The structure of cyanobacterial phycobilisomes: a model. *Arch. Microbiol.* **123**, 113–127 (1979).
16. 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).
17. Yamanaka, G., Glazer, A. N. & Williams, R. C. Molecular architecture of a light-harvesting antenna. Comparison of wild type and mutant *Synechococcus* 6301 phycobilisomes. *J. Biol. Chem.* **255**, 11104–11110 (1980).
18. 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. *Eur. J. Biochem.* **236**, 1010–1024 (1996).
19. 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).
20. Chang, L. et al. Structural organization of an intact phycobilisome and its association with photosystem II. *Cell Res.* **25**, 726–737 (2015).
21. Gantt, E. & Lipschultz, C. A. Phycobilisomes of *Porphyridium cruentum*. I. Isolation. *J. Cell Biol.* **54**, 313–324 (1972).
22. Arteni, A. A. et al. Structure and organization of phycobilisomes on membranes of the red alga *Porphyridium cruentum*. *Photosynth. Res.* **95**, 169–174 (2008).
23. Gantt, E. & Lipschultz, C. A. Structure and phycobiliprotein composition of phycobilisomes from *Griffithsia pacifica* (Rhodophyceae). *J. Phycol.* **16**, 394–398 (1980).
24. Guglielmi, G., Cohen-Bazire, G. & Bryant, D. A. The structure of *Gloeobacter violaceus* and its phycobilisomes. *Arch. Microbiol.* **129**, 181–189 (1981).
25. Lange, W., Wilhelm, C., Wehrmeyer, W. & Mörschel, E. The supramolecular structure of photosystem II–phycobilisome-complexes of *Porphyridium cruentum*. *Bot. Acta* **103**, 250–257 (1990).
26. Gantt, E. & Conti, S. F. The ultrastructure of *Porphyridium cruentum*. *J. Cell Biol.* **26**, 365–381 (1965).
27. Bhattacharya, D. et al. Genome of the red alga *Porphyridium purpureum*. *Nat. Commun.* **4**, 1941 (2013).
28. Tajima, N. et al. Analysis of the complete plastid genome of the unicellular red alga *Porphyridium purpureum*. *J. Plant Res.* **127**, 389–397 (2014).
29. Gantt, E. & Lipschultz, C. A. Phycobilisomes of *Porphyridium cruentum*: pigment analysis. *Biochemistry* **13**, 2960–2966 (1974).
30. Glazer, A. N. & Hixson, C. S. Subunit structure and chromophore composition of rhodophytan phycoerythrins. *Porphyridium cruentum* B-phycoerythrin and b-phycoerythrin. *J. Biol. Chem.* **252**, 32–42 (1977).
31. Redlinger, T. & Gantt, E. Phycobilisome structure of *Porphyridium cruentum*: polypeptide composition. *Plant Physiol.* **68**, 1375–1379 (1981).
32. Ficner, R. & Huber, R. Refined crystal structure of phycoerythrin from *Porphyridium cruentum* at 0.23-nm resolution and localization of the gamma subunit. *Eur. J. Biochem.* **218**, 103–106 (1993).
33. Ducret, A., Sidler, W., Frank, G. & Zuber, H. The complete amino acid sequence of R-phycocyanin-I α and β subunits from the red alga *Porphyridium cruentum*. Structural and phylogenetic relationships of the phycocyanins within the phycobiliprotein families. *Eur. J. Biochem.* **221**, 563–580 (1994).
34. Glazer, A. N. & Hixson, C. S. Characterization of R-phycocyanin. Chromophore content of R-phycocyanin and C-phycoerythrin. *J. Biol. Chem.* **250**, 5487–5495 (1975).
35. Camara-Artigas, A. et al. pH-dependent structural conformations of B-phycoerythrin from *Porphyridium cruentum*. *FEBS J.* **279**, 3680–3691 (2012).
36. Ritter, S., Hiller, R. G., Wrench, P. M., Welte, W. & Diederichs, K. Crystal structure of a phycourobilin-containing phycoerythrin at 1.90-Å resolution. *J. Struct. Biol.* **126**, 86–97 (1999).
37. Lüning, K. in *Seaweeds: Their Environment, Biogeography, and Ecophysiology*. (ed. Yarish, C. and Kirkman, H.) (John Wiley & Sons, 1990).
38. 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).
39. Jiang, T., Zhang, J. & Liang, D. Structure and function of chromophores in R-phycoerythrin at 1.9 Å resolution. *Proteins* **34**, 224–231 (1999).
40. Cantor, C. R. & Schimmel, P. R. *Biophysical Chemistry: Part II Techniques for the Study of Biological Structure and Function* (W. H. Freeman, 1980).
41. Gervasio, F. L., Chelli, R., Marchi, M., Procacci, P. & Schettino, V. Determination of the potential of mean force of aromatic amino acid complexes in various solvents using molecular dynamics simulations: the case of the tryptophan–histidine pair. *J. Phys. Chem. B* **105**, 7835–7846 (2001).
42. Gallivan, J. P. & Dougherty, D. A. Cation–π interactions in structural biology. *Proc. Natl Acad. Sci. USA* **96**, 9459–9464 (1999).
43. 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 L_{cm}, β^{18} and $\alpha^{AP\text{-}B}$ chromophores. *Biochim. Biophys. Acta* **1186**, 153–162 (1994).

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Preparation of phycobilisomes

P. purpureum (From UTEX Culture Collection of Algae, UTEX 2757) was cultured in Bold 1NV: Erdshreiber (1:1) half-seawater medium, bubbled with sterilizing filtered air at 22 °C, under a 16 h:8 h light–dark cycle, with a white-light flux of about 37 μmol photons per m^2 per second. Algal cells were collected by centrifugation for 10 min at 6,000*g*, and resuspended in Buffer A (0.65 M Na/KPO_4 buffer with 0.5 M sucrose and 10 mM EDTA, pH 7.0) at 0.3 g of wet weight per ml. Then cells were homogenized twice at 4 °C using a French Press (EmulsiFlex-C3, Avestin) at 4,000 p.s.i., and phenylmethylsulfonyl fluoride was added to a final concentration of 1 mM. After 30 min of incubation with lauryldimethylamine *N*-oxide (Sigma) (48 mg g^{-1} wet algal cells), debris and supernatant chlorophyll were removed by centrifugation at 20,000*g* for 30 min at 18 °C. The middle aqueous violet solution was loaded in 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) and spun at 120,000*g* for 4 h at 18 °C using a SW41 rotor on Optima XPN-100 centrifuge (Beckman Coulter). Three visible bands were obtained after centrifugation and violet band 1 is the main layer of intact PBSs (Extended Data Fig. 1a).

Absorption and fluorescence spectrum measurement

Absorption of the intact PBS was measured between 300–800 nm using an Ultrospec 2100 Pro ultraviolet–visible spectrophotometer (Biochrom).

Fluorescence emission spectra were recorded using a Hitachi FL-4500 fluorescence spectrophotometer at room temperature. After exciting at 450 nm, fluorescence emission was monitored from 500 to 700 nm.

Mass spectrometry analysis

Mass spectrometry analysis was performed as previously described[5]. In brief, the purified PBS was separated by 4–12% Bis-Tris SDS–PAGE in MES buffer and the gel was stained with $ZnSO_4$ to detect bilin-containing proteins with ultraviolet light by Zn-enhanced fluorescence. Then, four fluorescence bands with molecular mass greater than 25 kDa were excised for in-gel digestion and proteins were identified by mass spectrometry (Extended Data Fig. 1b). The intact PBS complex in solution was also subjected to mass spectrometric analysis. Finally, all 25 protein components of PBS were identified in the samples (Extended Data Fig. 1e).

Cryo-EM sample preparation and data collection

We used holey-carbon copper grids (Quantifoll R2/2, 400 mesh) covered with homemade ultrathin carbon for cryo-EM sample preparation. Cryo-EM grids were prepared with Vitrobot Mark IV (FEI Company) at 16 °C and 100% humidity. The grids were glow-discharged after adding 50 μl amylamine to a glass culture dish in the plasma cleaner and vapouring into the air. We added a 1.5 μl aliquot of protein with a concentration of 1.5 mg ml^{-1} to the grids and waited for 60 s, and then added 3.5 μl of 50 mM Tris buffer (pH 8.0) to the grids and quickly mixed with the sample twice to reduce the salt concentration. The grids were then blotted for 3.5 s and plunged into liquid ethane cooled by liquid nitrogen.

The cryo-EM data were collected using a Titan Kiros Microscope (FEI) operated at a voltage of 300 kV and equipped with a Cs corrector, a GIF Quantum energy filter (Gatan) and a K2 Summit direct electron detector (Gatan). A preset defocus range of −1.2 μm to −2.2 μm was used. All cryo-EM images were recorded at a nominal magnification of 105,000× in super-resolution mode. Each stack was exposed for 5.6 s with an exposure time of 0.175 s per frame and recorded as a movie of 32 frames, resulting in the total dose rate of approximately 48 electrons per $Å^2$ for each stack. GIF was set to a slit width of 20 eV. The data were collected automatically using the software AutoEMation[44]. The stacks were motion-corrected with MotionCor2[45] and binned twofold, resulting in a pixel size of 1.091 Å per pixel.

Cryo-EM data analysis

A total of 16,218 micrographs were collected. Micrograph screening, manual particle picking and normalization were performed using EMAN2[46] and RELION3.0 beta[47–49]. The contrast transfer function parameters of each micrograph were estimated using CTFFIND4[50,51] and automatic particle picking; all the 2D, 3D classification, 3D refinement and local defocus calculation were performed with RELION3.0 beta[47–49].

The workflow of the data analysis is shown in Extended Data Fig. 2f. Two batches of data were collected and processed individually at the beginning. Particles were first manually picked from a small set of micrographs to produce templates for autopicking. Then particles were autopicked on all micrographs and manually screened to eliminate aggregation and ice contamination. Finally 322,889 and 363,480 particles were selected for the next 2D classification. After several rounds of 2D classification, 299,888 and 333,012 particles were left for the 3D classification. After 3D classification, two classes from each dataset with good quality were selected for the final reconstruction. At this point, we calculated the local defocus values for each particle and re-extracted particles from the dose-weighted micrographs[52]. Then two batches of data were merged to perform the 3D refinement. The final resolution of the 3D auto-refinement after post-processing was 2.82 Å with a final particle number of 191,825 after imposing the C2 symmetry. Application of a mask for the core region during refinement further improved the resolution of this region to 2.68 Å. We also applied local masks for each rod and extra phycoerythrin hexamer, which resulted in improved quality of local maps with resolutions ranging between 2.77 Å and 3.56 Å. The maps for the target regions were extracted from the overall map by Chimera[53], and the masks were created by RELION3.0 beta[47–49]. All resolutions were estimated with the gold-standard Fourier shell correlation 0.143 criterion with high-resolution noise substitution. All the local resolution maps were calculated using ResMap[54].

Model building and refinement

We searched the published genome and transcriptome database of *P. purpureum*[27] against the 25 protein sequences of the *G. pacifica* PBS using the Basic Local Alignment Search Tool (BLAST). A total of 24 homologues, including eight PBP proteins and 16 linker proteins were obtained by this procedure, and these proteins are used as the candidates for model building. Local maps generated by the different masks described above were used to facilitate the model building process.

Because the sequences of *P. purpureum* and *G. pacifica* have high homology with each other, we first docked the structures of the *G. pacifica* PBS proteins (PDB: 5Y6P)[5] into the electron microscopy maps using Chimera[53]. All the PBP proteins and most of the linker proteins were fitted well. Then the sequence assignments were guided by well-resolved bulky residues such as phenylalanine, tyrosine, tryptophan and arginine, and the sequences of the *G. pacifica* PBS were replaced with corresponding residues in the *P. purpureum* PBS in Coot[55] and every residue was examined and manually adjusted to better fit in the map. Some of the $L_R\gamma4$ proteins could not be fitted well at the N-terminal region. We first built the C-terminal CBDγ domain as described above, and then performed the de novo building in Coot[55] with bulky residues as land markers as most of these residues were clearly visible in our cryo-EM maps. The linker protein located at the centre cavity of the hexamer Hd is $L_R\gamma6$ in the *G. pacifica* PBS that contains the CBDγ domain; however, the density at this region in the *P. purpureum* PBS shows recognizable structural features of the Pfam00427 domain. Therefore, we named this linker protein L_R6 and first docked the structure of the Pfam00427 domain from L_R2 into the density. By carefully

examining the densities outside the Pfam00427 domain of L_R6, a YYW motif was unambiguously identified according to the clear side-chain densities. Then we obtained the full-length sequence of L_R6 by searching the published genome and transcriptome database of *P. purpureum* for the protein containing both the Pfam00427 domain and the YYW motif. The sequence of L_R2 was replaced with corresponding residues in L_R6 in Coot[55] and de novo atomic model building was conducted for the rest of the sequence in Coot[55]. Finally, 25 protein sequences were identified and confirmed by good agreement of the side-chain information between the sequences and the density maps (Supplementary Table 1).

The initial model was completed via iterative rounds of manual building with Coot[55] and refinement with phenix.real_space_refine[56,57]. During this process, each part of the whole PBS model corresponding to each local map was refined against the local map with secondary structure and geometry restraints to prevent overfitting. Then, all parts were merged into a whole PBS model and this overall model was refined again against the overall 2.8 Å map using phenix.real_space_refine[56,57]. The atomic model was cross-validated according to previously described procedures[58]. In brief, atoms in the final model were randomly shifted by up to 0.5 Å, and the new model was then refined against one of two half-maps generated during the final 3D reconstruction. FSC values were calculated between the map generated from the resulting model and the two half-maps, as well as the averaged map of two half-maps. We did not observe notable separation between FSCwork and FSCfree, indicating that our model was not over-refined (Extended Data Fig. 2e). The data collection, model refinement and validation statistics are presented in Extended Data Tables 1, 2b. The statistics of the geometries of the models were generated using MolProbity[59]. All the figures were prepared in PyMOL (http://pymol.org) or Chimera[53]. The sequence alignments were performed by ClustalX2[60] and created by ESPript[61].

Reporting summary

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

Data availability

The atomic coordinates have been deposited in the Protein Data Bank with the accession code 6KGX. The electron microscopy maps have been deposited in the Electron Microscopy Data Bank with accession codes EMD-9976 for the overall map and EMD-9977 through to EMD-9988 for the 12 local maps. The raw electron microscopy images used to build the 3D structure are available from the corresponding authors upon request.

44. Lei, J. & Frank, J. Automated acquisition of cryo-electron micrographs for single particle reconstruction on an FEI Tecnai electron microscope. *J. Struct. Biol.* **150**, 69–80 (2005).
45. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
46. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
47. Scheres, S. H. W. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012).
48. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
49. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, 163 (2018).
50. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
51. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
52. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).
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. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
56. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
57. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).
58. Amunts, A. et al. Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014).
59. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
60. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
61. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).

Acknowledgements We thank the staff at the Tsinghua University Branch of the National Protein Science Facility (Beijing) for technical support on the Cryo-EM and High-Performance Computation platforms; J. Wang for model validation; D. Liu for model building and structure refinement; H. Lin and X. Pan for discussion; and X. Li and H.-W. Wang for recommendations for the computation. This work was supported by the National Basic Research Program (grants 2016YFA0501101 and 2017YFA0504600 to S.-F.S.) and the National Natural Science Foundation of China (grants 31670745 and 31861143048 to S.-F.S.).

Author contributions S.-F.S. supervised the project; J.M. prepared the samples, collected and analysed the electron microscopy data, performed the initial model building and the biochemical and biophysical analyses; X.Y. performed the model building and the structure refinement; J.M., X.Y., S.S. and S.-F.S. analysed the structure; X.W. helped with the electron microscopy data collection and the biochemical and biophysical analyses; S.Q. contributed to the sample selection; J.M. and X.Y. wrote the initial draft; and S.S. and S.-F.S. edited the manuscript.

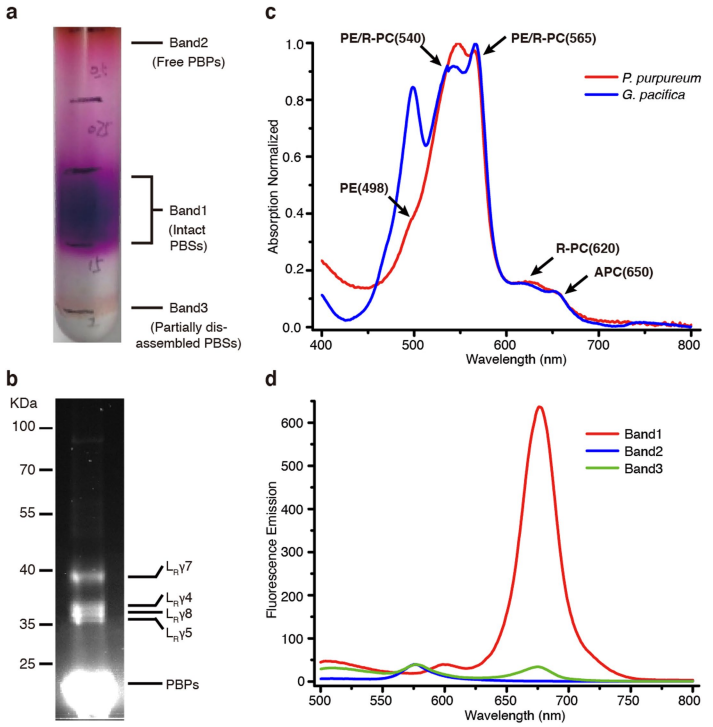
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2020-7.

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

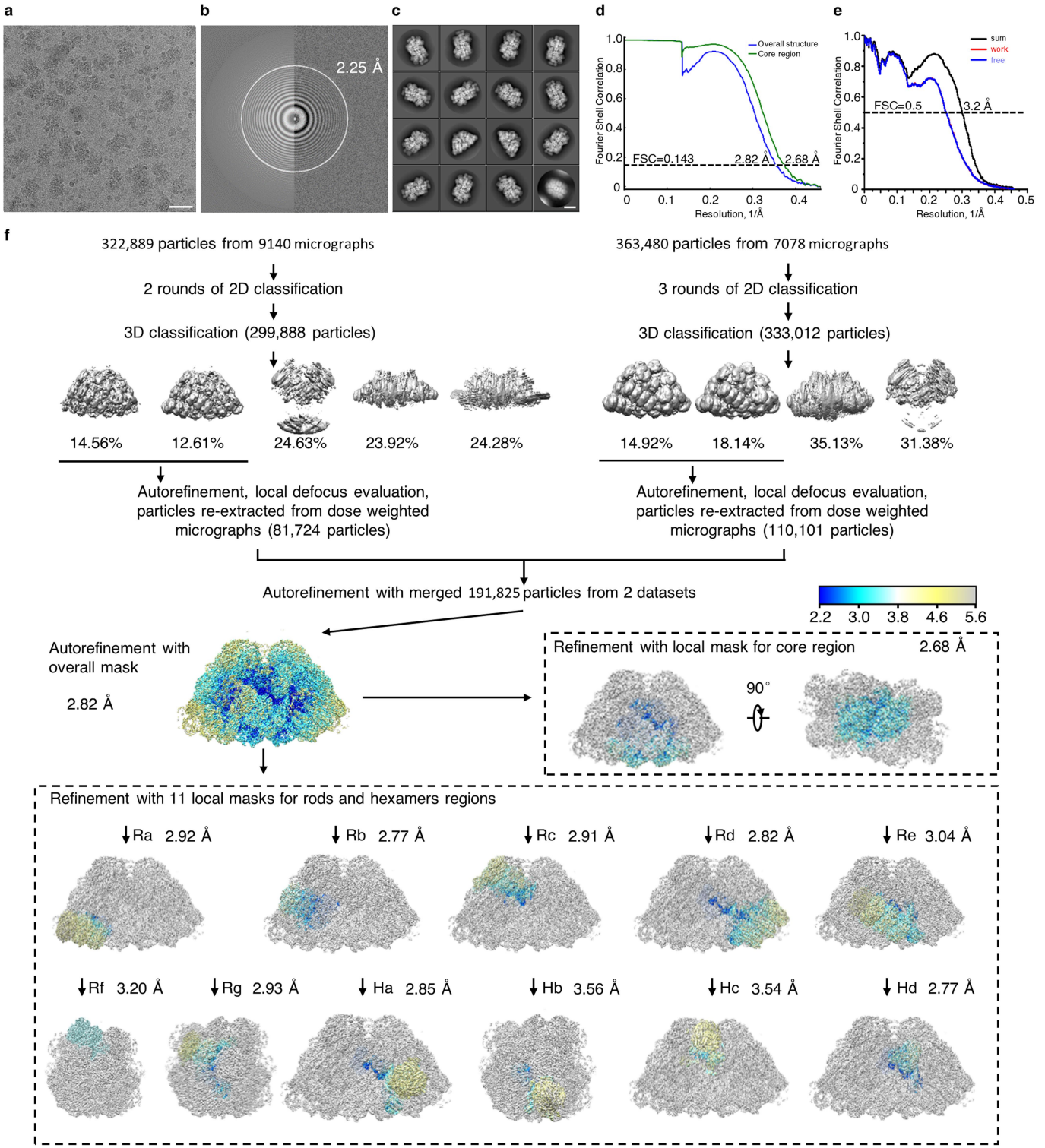
Reprints and permissions information is available at http://www.nature.com/reprints.



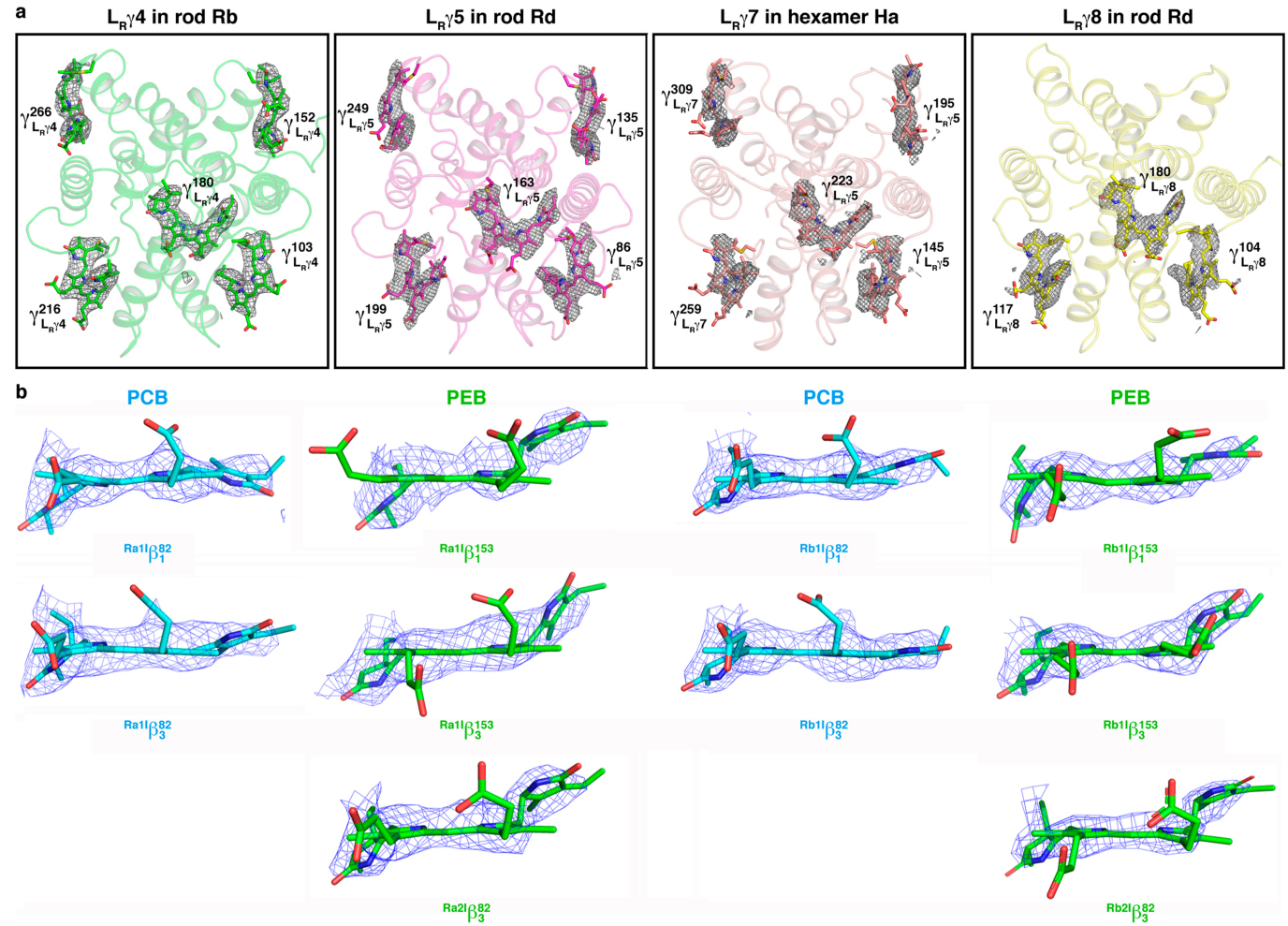
e

Name	Peptides		PSMs		MW[KDa]
	batch1	batch2	batch1	batch2	
α^{APC}	6	8	143	122	17.6
β^{APC}	11	10	521	388	17.4
ApcD	7	4	53	47	18.1
ApcF	9	10	76	54	19.8
α^{PC}	17	6	323	113	17.5
β^{PC}	16	10	521	259	18.2
α^{PE}	19	11	849	598	17.8
β^{PE}	23	13	1289	923	18.6
L_C	3	7	21	91	10.4
L_{CM}	42	38	324	471	99.4
$L_{RC}1$	16	14	159	170	27.0
$L_{RC}2$	15	25	57	238	53.5
$L_{RC}3$	17	27	77	224	55.1
$L_{RC}4$	2	5	6	75	16.9
$L_{RC}5$	7	11	59	354	30.2
$L_{RC}6$	3	5	25	54	26.6
L_R1	16	21	122	261	39.7
L_R2	3	10	26	85	32.8
L_R3	14	12	175	152	37.1
$L_R\gamma4$	13	17	68	237	31.7
$L_R\gamma5$	7	8	20	107	29.4
L_R6	7	4	39	36	43.9
$L_R\gamma7$	13	22	58	184	36.9
$L_R\gamma8$	19	8	129	54	31.3
L_R9	15	18	118	191	31.3

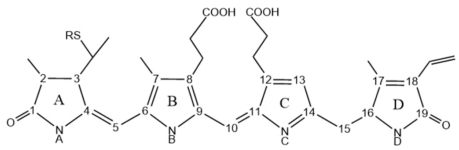
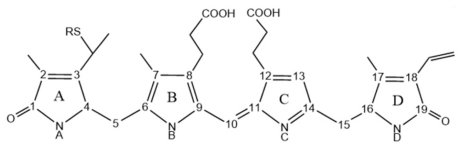
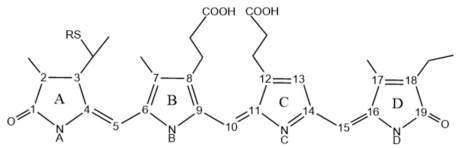
Extended Data Fig. 1 | Preparation and characterization of the PBS from *P. purpureum*. **a**, Isolation of PBSs using sucrose density gradient centrifugation. Three visible bands were observed. Band 1 is the sample of PBSs used for single-particle analysis in this study. The purification of PBS was repeated independently at least three times with similar results. **b**, Analysis of the protein composition of band 1 by SDS–PAGE stained with $ZnSO_4$ to enable the detection of bilin-containing proteins with ultraviolet light by Zn-enhanced fluorescence. The bands of $L_R\gamma4$, 5, 7, 8 and PBPs identified by mass spectrometric analysis are indicated. For gel source data, see Supplementary Fig. 1. The purification and characterization of the protein composition was repeated independently at least three times with similar results. **c**, Absorption spectrum of band 1 and the PBS from *G. pacifica*. The peaks at 498 nm, 620 nm and 650 nm are from phycourobilins, PCBs of phycocyanins and PCBs of allophycocyanins, respectively. The peaks at 540 nm and 565 nm are from PEBs. The reduced absorption of the *P. purpureum* PBS compared with the *G. pacifica* PBS at 498 nm indicates that the phycourobilin content of *P. purpureum* is much lower than that of *G. pacifica*. **d**, Fluorescence emission spectra of the three bands. Emission maxima at 580 nm and 676 nm represent the disassembled phycoerythrin hexamer and the terminal emitter in the intact PBS, respectively. Band 1 has an emission peak at 676 nm, band 2 at 580 nm and band 3 has two emission peaks at 676 nm and 580 nm, indicating that band 1 contains intact PBSs, band 2 contains free PBPs and band 3 contains partially disassembled PBSs. **e**, Results of the mass spectrometric analysis of purified PBSs. Two batches of sample were analysed. The similar results confirmed the consistency of our purification method.



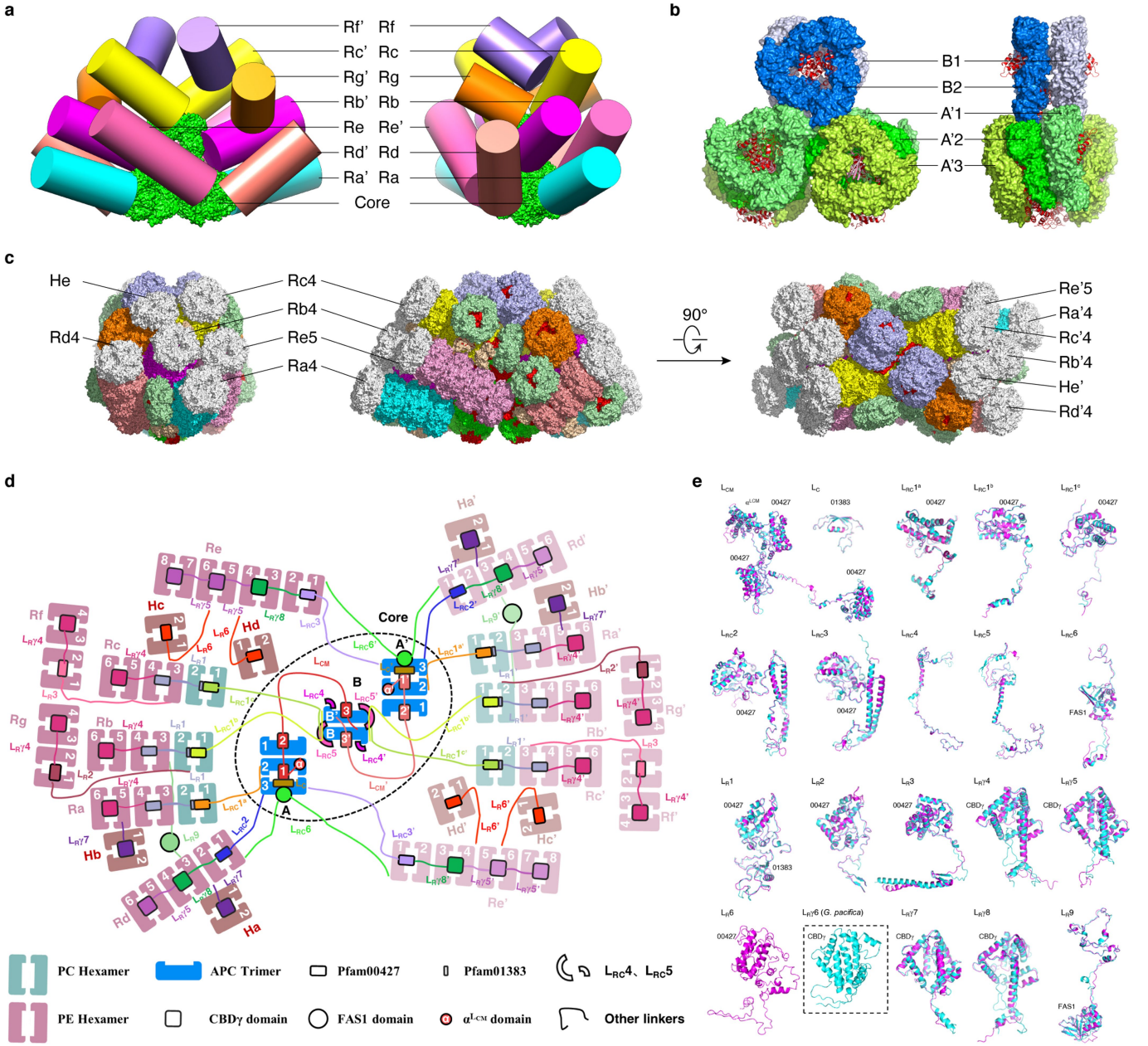
Extended Data Fig. 2 | Cryo-EM analysis of the PBS from *P. purpureum*. **a**, A representative motion-corrected electron micrograph of PBSs. Scale bar, 50 nm. Tens of thousands of micrographs were collected with similar results. **b**, Fourier power spectrum of the micrograph showing the Thon ring extending to 2.25 Å. Tens of thousands of micrographs were collected with similar results. **c**, Typical good, reference-free 2D class averages from single-particle PBS images. Scale bar, 20 nm. More than three rounds of 2D class average were performed with similar results. **d**, Gold-standard Fourier shell correlation (FSC) curves for the 3D electron microscopy reconstructions of the PBS. Blue curve, FSC curve for the overall structure; green curve, FSC curve for the core region that was masked during refinement. **e**, FSC curves for the cross-validation of the atomic model. The small difference between work and free FSC curves suggested that the model was not overfitted. **f**, The workflow for the 2D and 3D classifications for cryo-EM data processing. The masking strategy for dealing with sub-regions of PBS is enclosed within dashed lines. For details, see 'Cryo-EM data analysis' in Methods.



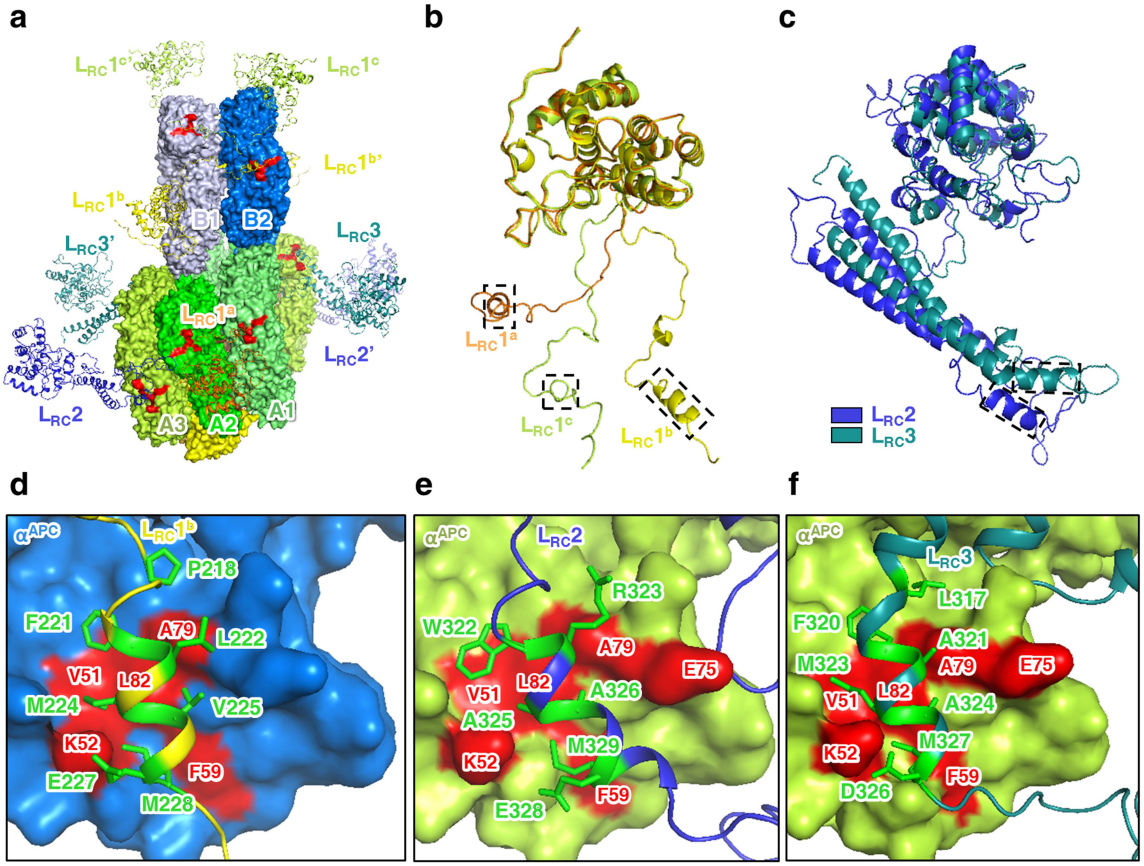
c

	Φ_1	Ψ_1	Φ_2	Ψ_2	Φ_3	Ψ_3	Chemical structure
PEB	0.1	132.4	-1.1	4.0	85.1	45.1	
PUB	-38.4	-84.1	17.3	0.2	52.9	75.4	
PCB (core)	5.8	157.9	0.6	-1.4	167.3	5.5	
PCB (L_{CM})	0.1	0	0.2	-0.1	179.8	0.1	

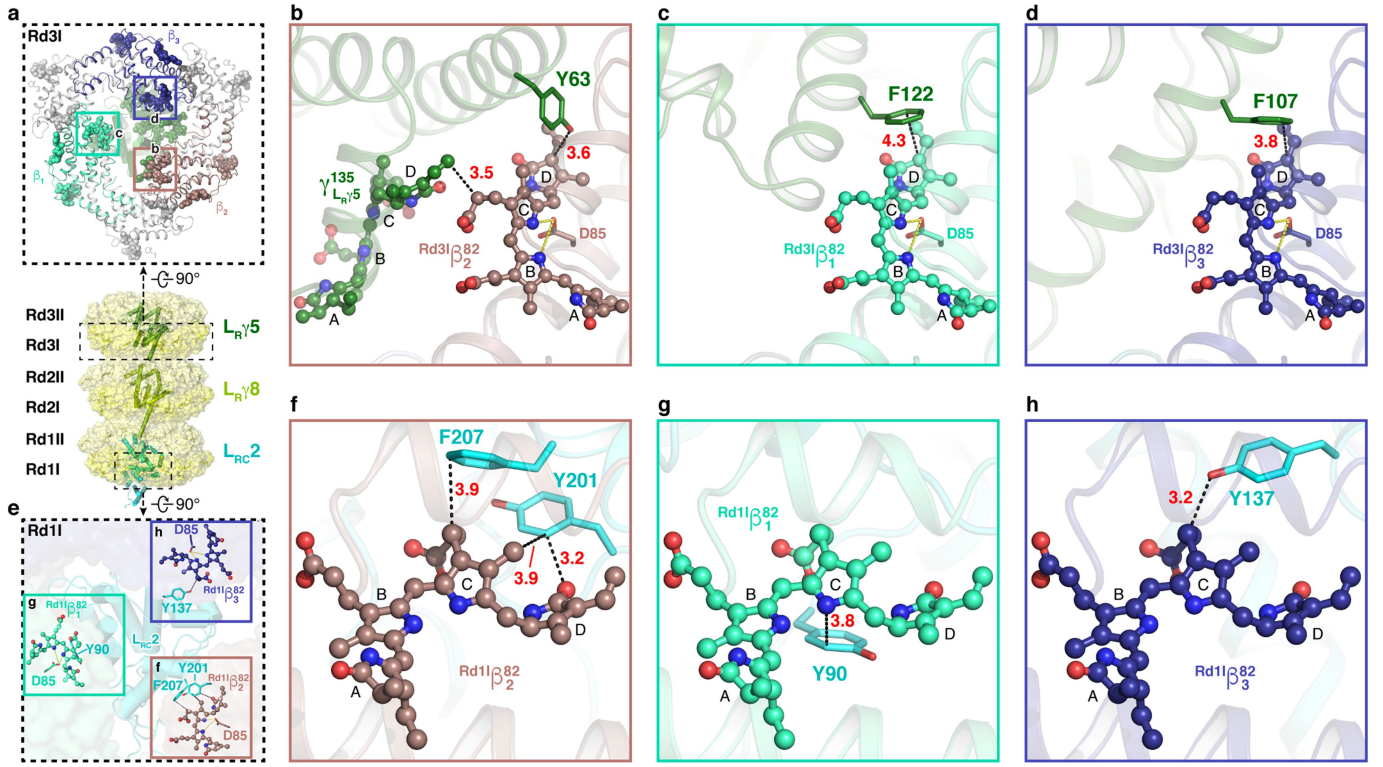
Extended Data Fig. 3 | Characterization of different types of chromophore. **a**, Cryo-EM densities (mesh) of bilins (stick) bound to $L_R\gamma4$ in the rod Rb, $L_R\gamma5$ in rod Rd, $L_R\gamma7$ in the hexamer Ha and $L_R\gamma8$ in the rod Rd. **b**, The densities (mesh) of some PCB and PEB bilins (stick) in R-phycocyanins and phycoerythrins from rods Ra and Rb to show their different coplanarities. All of the density maps of PCB bilins showed a very flat conformation of rings B, C and D, consistent with the carbon–carbon double bond between rings C and D in PCB that constraints the movement of ring D, so that ring D is coplanar with the B–C plane. Conversely, most of the density maps of PEB displayed a curved conformation of rings B, C and D owing to the single carbon–carbon bond between rings C and D in PEB that allows the rotation of ring D, so that ring D deviates from the B–C plane. However, some PEBs in R-phycocyanin also showed a planar conformation—such as $^{Ra1I}\beta^{153}_3$ and $^{Rb1I}\beta^{153}_1$—although to a lesser extent than that for a typical PCB molecule. **c**, Dihedral angles of three kinds of chromophore. The dihedral angles Φ_1, Ψ_1, Φ_2 … are defined by the atoms NA–C(4)–C(5)–C(6), C(4)–C(5)–C(6)–NB, NB–C(9)–C(10)–C(11) … etc.



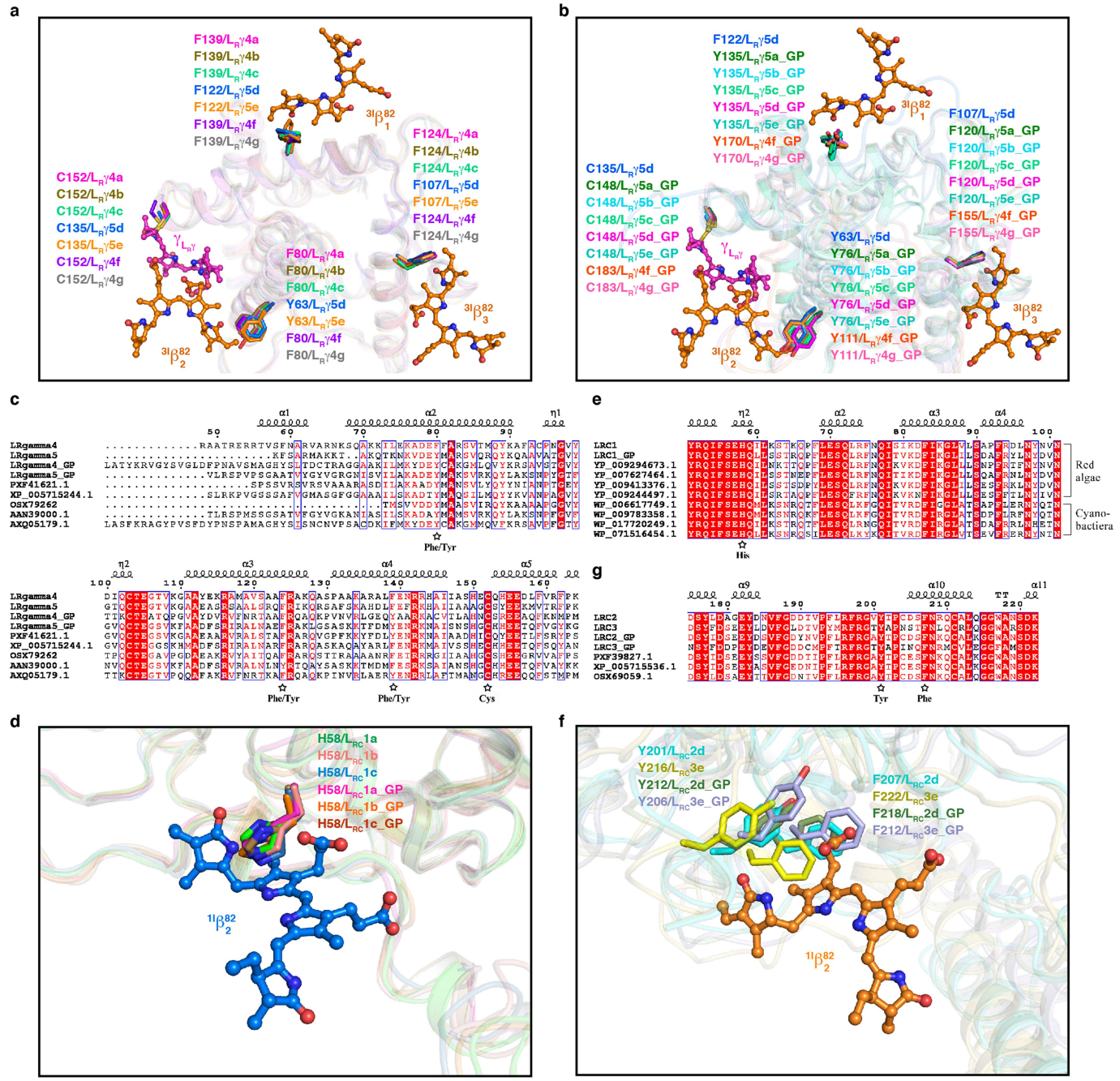
Extended Data Fig. 4 | Overall structure of the PBS from *P. purpureum* and comparison with that from *G. pacifica*. **a**, Schematic diagram showing the organization of the rods and the core from two perpendicular views. The colouring scheme is the same as in Fig. 1e. **b**, Structure of the core from two perpendicular views shows the assembly and arrangement of the core layers. **c**, Overall structure of the PBS overlapped with the *G. pacifica* PBS displayed in surface representation from three perpendicular views. The additional hexamers in the *G. pacifica* PBS are coloured white and labelled. **d**, Schematic model of the PBS architecture. The connections between PBS components are shown. Dark and light colours show C2 symmetric parts of rods. Dark and light salmon, phycoerythrin hexamers in rod; dark and light brown, extra phycoerythrin hexamers; dark and light forest green, phycocyanin hexamers; blue, allophycocyanin trimer; large rectangular box, Pfam00427 domains; small rectangular box, Pfam01383 domains; square box, CBDγ. **e**, Comparison of linker proteins from *P. purpureum* with those from *G. pacifica*. Structures of the 19 well-resolved linker proteins (magenta) are superimposed with those from the *G. pacifica* PBS (cyan). The linker proteins share very high structural conservation—such as the Pfam00427 domain in the rod–core linker (L_{RC})1–3/L_{RC}1′–3′, the rod linker (L_R)1–3/L_R1′–3′ and L_{CM}/L_{CM}′, the Pfam01383 domain in the core linker (L_C)/L_C′ and L_R1/L_R1′, the FAS1 domain in L_{RC}6/L_{RC}6′ and L_R9/L_R9′, the CBDγ domain in L_Rγ4–5/L_Rγ4′–5′ and L_Rγ7–8/L_Rγ7′–8′, the coiled-coil motif at the C termini of L_{RC}2–3/L_{RC}2′–3′, and the long α-helix in the middle of the L_{RC}4–5/L_{RC}4′–5′. Note that L_R6 from the *P. purpureum* PBS is different from L_Rγ6 from the *G. pacifica* PBS, therefore they are not aligned. Domains of $\alpha^{L\text{-}CM}$, Pfam00427 (00427), Pfam01383 (01383), CBDγ, and FAS1 are labelled.



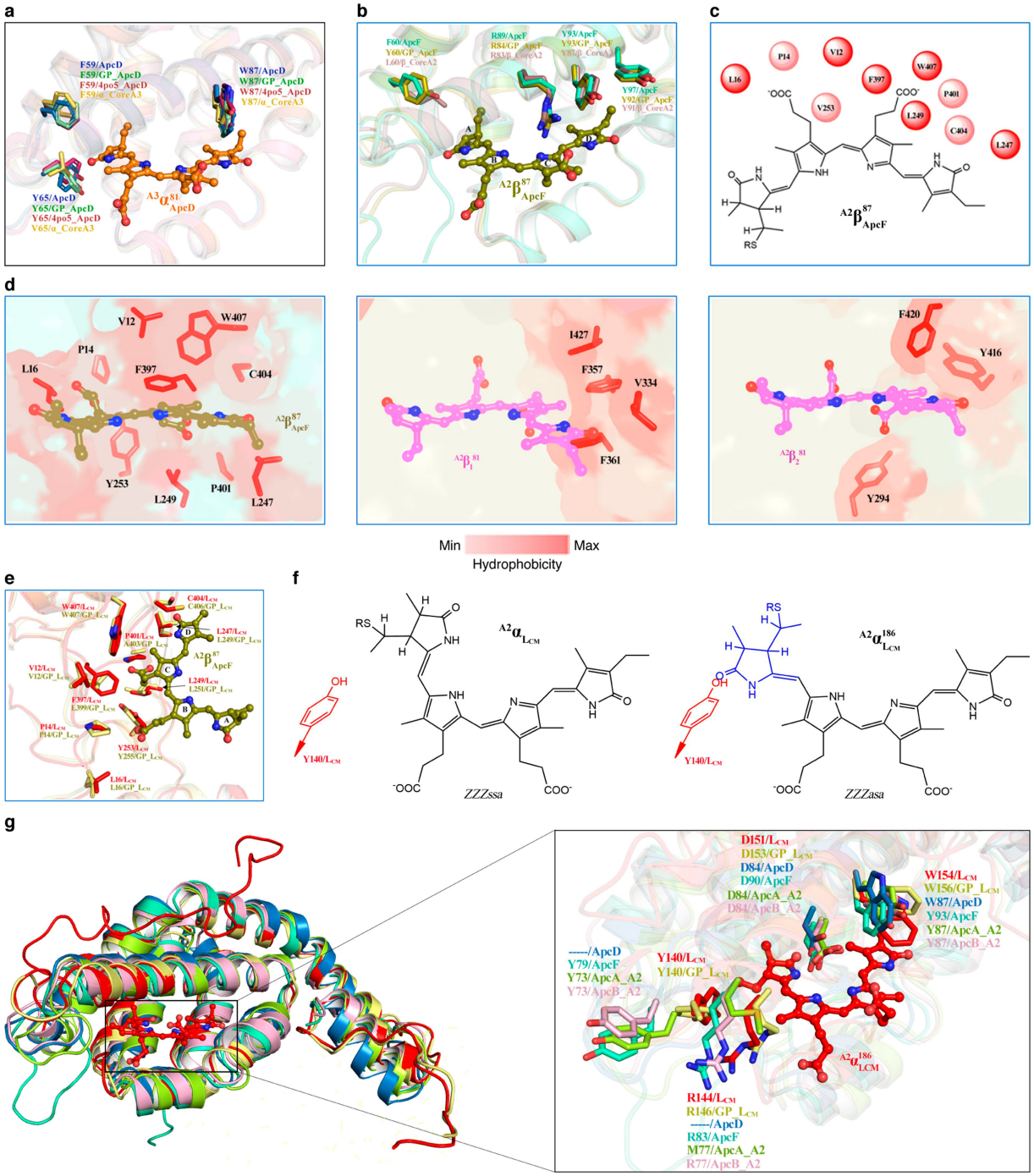
Extended Data Fig. 5 | Interactions between L_{RC} proteins and the core. **a**, Organization of L_{RC} proteins $L_{RC}1$–3/$L_{RC}1'$–$3'$ and the core. The grooves on the α subunits that contact the linker helices are shown in red. **b**, Structural similarity and differences among $L_{RC}1^{a}$, $L_{RC}1^{b}$ and $L_{RC}1^{c}$. These rod–core linkers are superimposed relative to the Pfam00427 domain. The helices that interact with the core are boxed. **c**, 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. **d**–**f**, Interactions between the α^{APC} subunit and the helices of $L_{RC}1^{b}$ (**d**), $L_{RC}2$ (**e**) and $L_{RC}3$ (**f**). 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.



Extended Data Fig. 6 | Interactions of the linker proteins $L_R\gamma$s and L_{RC}s with chromophores in the rod Rd. a, Bottom, overall structure of the rod Rd with the hexamers shown in surface representation and the linker proteins shown in cartoon representation. Top, structure of the layer Rd3I. Proteins and bilins are shown in cartoon and sphere representations, respectively. Three β subunits are coloured differently and the β82 PEBs are boxed and analysed in detail in **b**–**d**. **b**, The interactions between the residue Y63 and the bilin $\gamma^{135}_{L_R\gamma5}$ from $L_R\gamma5$ with the bilin $^{Rd3I}\beta^{82}_2$. **c**, The interaction between F122 from $L_R\gamma5$ and the bilin $^{Rd3I}\beta^{82}_1$. **d**, The interaction between F107 from $L_R\gamma5$ and the bilin $^{Rd3I}\beta^{82}_3$. **e**, A focused view of the structure of the layer Rd1I showing the central triangle area. PBPs, the linker protein, bilins and residues are shown in surface, cartoon, ball-and-stick and stick representations, respectively. Three β82 PCBs are boxed and analysed in detail in **f**–**h**. **f**, The interactions between Y201 and F207 from $L_{RC}2$ and the bilin $^{Rd1I}\beta^{82}_2$. **g**, The interaction between Y90 from $L_{RC}2$ and the bilin $^{Rd1I}\beta^{82}_1$. **h**, The interaction between Y137 from $L_{RC}2$ and the bilin $^{Rd1I}\beta^{82}_3$.



Extended Data Fig. 7 | Comparisons of linker proteins from both *P. purpureum* and *G. pacifica*. **a**, **b**, Structural alignment of $L_R\gamma$ linker proteins in the outmost hexamers of various rods from the *P. purpureum* PBS (**a**) and the *G. pacifica* PBS (**b**). β82 PEBs and residues of $L_R\gamma$ linker proteins are shown in ball-and-stick and stick representations, respectively. Note that an aromatic residue from the $L_R\gamma$ linker is present near to each β82 PEB to form π–π interactions, and one bilin from the $L_R\gamma$ linker ($\gamma_{L_R\gamma}$) always provides additional π electrons to the conjugation system of the β^{82}_2 PEB. These aromatic residues and the bilins from $L_R\gamma$ linker proteins are conserved in both *P. purpureum* and *G. pacifica*. **c**, Sequence alignment of $L_R\gamma4$–5 from *P. purpureum* and other red algae. Three aromatic residues interacting with the β82 PEBs and the cysteine residues used to link the bilins close to the β82 PEBs are marked by stars. LRgamma4_GP and LRgamma5_GP, $L_R\gamma4$–5 from *G. pacifica*; PXF41621.1, γ-subunit from *Gracilariopsis chorda*; XP_005715244.1, γ-subunit from *Chondrus crispus*; OSX79262, γ-subunit from *Porphyra umbilicalis*; AAN39000.1, γ-subunit from *Griffithsia japonica*; AXQ05179.1, γ-subunit from *Agarophyton chilense*. **d**, Structural alignment of $L_{RC}1$ linker proteins from *P. purpureum* and *G. pacifica* in the phycocyanin hexamer showing the bilin β^{82}_2 and the surroundings. The key histidine residue close to the β^{82}_2 PCB is conserved. **e**, Sequence alignment of $L_{RC}1$ from *P. purpureum* and other red algal and cyanobacterial species. The key histidine residue close to the β82 PCBs is marked with a star. LRC1_GP, $L_{RC}1$ from *G. pacifica*; YP_009294673.1, $L_{RC}1$ from red algal *G. chorda*; YP_007627464.1, $L_{RC}1$ from red algal *C. crispus*; YP_009413376.1, $L_{RC}1$ from red algal *P. umbilicalis*; YP_009244497.1, $L_{RC}1$ from red algal *A. chilense*; WP_006617749.1, $L_{RC}1$ from cyanobacteria *Arthrospira platensis*; WP_009783358.1, $L_{RC}1$ from cyanobacteria *Lyngbya* sp. PCC 8106; WP_017720249.1, $L_{RC}1$ from cyanobacteria *Oscillatoria* sp. PCC 10802; WP_071516454.1, $L_{RC}1$ from cyanobacteria *Geitlerinema* sp. PCC 9228. **f**, Structural alignment of the $L_{RC}2$ and $L_{RC}3$ linker proteins from *P. purpureum* and *G. pacifica* in the phycoerythrin hexamer proximal to the core showing the bilin β^{82}_2 and the surroundings. Two aromatic residues near to the β^{82}_2 PEB are conserved in both *P. purpureum* and *G. pacifica*. **g**, Sequence alignment of $L_{RC}2$–3 from *P. purpureum* and other red algae. Two aromatic residues close to the β82 PEBs are marked with stars. $L_{RC}2\_GP$ and $L_{RC}3\_GP$ are from *G. pacifica*. PXF39827.1, XP_005715536.1 and OSX69059.1 are from *G. chorda*, *C. crispus* and *P. umbilicalis*, respectively.



Extended Data Fig. 8 | Characterization of ApcD, ApcF and the α subunit domain of L_{CM}. **a**, Magnified view of the superimposition of ApcD proteins from *P. purpureum*, *G. pacifica* (GP_ApcD), *Synechocystis* PCC 6803 (4PO5_ApcD) and the α subunit of the core layer A3 (α_CoreA3). Bilins and residues are shown in ball-and-stick and stick representations, respectively. Three aromatic residues near to the PCB are conserved in all ApcD proteins, but not in the α subunit of the core layer A3. **b**, Magnified view of the superimposition of ApcF proteins from *P. purpureum* and *G. pacifica* (GP_ApcF), and the β subunit of the core A2 (β_CoreA2). $^{A2}\beta^{87}_{ApcF}$ is shown in ball-and-stick representation in sand. **c**, A schematic of interactions between $^{A2}\beta^{87}_{ApcF}$ and the hydrophobic cap. **d**, Magnified view of the PCB pocket of ApcF (left), $^{A2}\beta^{81}_{1}$ (middle) and $^{A2}\beta^{81}_{2}$ (right). The protein is shown in surface representation and coloured on the basis of amino acid hydrophobicity (see colour bar). The side chains of hydrophobic residues within 5 Å of the PCB are shown in stick representation. **e**, Magnified view of the structural alignment of the hydrophobic caps formed by L_{CM} proteins from *P. purpureum* and *G. pacifica*. **f**, Schematic of the steric hindrance experienced by Y140/L_{CM} and the *ZZZasa* configuration of $^{A2}\alpha^{186}_{L_{CM}}$. **g**, Structural alignment of $\alpha^{L_{CM}}$, ApcD, ApcF, the α subunit (ApcA_A2) and the β subunit (ApcB_A2) in the core. The PCB pockets are indicated in the magnified view on the right.

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

	Phycobilisome from *P. purpureum* (EMDB-9976) (PDB 6KGX)
Data collection and processing	
Magnification	105,000
Voltage (kV)	300
Electron exposure (e–/Å^2)	48
Defocus range (μm)	-1.2 ~ -2.2
Pixel size (Å)	1.091
Symmetry imposed	C2
Initial particle images (no.)	686,369
Final particle images (no.)	191,825
Map resolution (Å)	2.82
FSC threshold	0.143
Map resolution range (Å)	2.4~7.4
Refinement	
Initial model used (PDB code)	5Y6P
Model resolution (Å)	2.82
FSC threshold	0.143
Model resolution range (Å)	2.4~7.4
Map sharpening *B* factor (Å^2)	-54.74
Model composition	
Non-hydrogen atoms	1014714
Protein residues	125577
Ligands	1598
B factors (Å^2)	
Protein	57.95
Ligand	71.06
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	2.064
Validation	
MolProbity score	1.79
Clashscore	8.07
Poor rotamers (%)	0.53
Ramachandran plot	
Favored (%)	94.88
Allowed (%)	5.11
Disallowed (%)	0.01

Extended Data Table 2 | Summary of proteins, chromophores and model validation

a

Subunit	Numbers in PBS	Chromophore						Total
		PCB		PEB		PUB		
		Per subunit	Total	Per subunit	Total	Per subunit	Total	
α^{APC}	20	1	20					20
β^{APC}	22	1	22					22
ApcD	2	1	2					2
ApcF	2	1	2					2
α^{PC}	36	1	36					36
β^{PC}	36	1	36	1	36			72
α^{PE}	254			2	508			508
β^{PE}	274			3	822			822
L_C	2							
L_{CM}	2	1	2					2
$L_{RC}1$	6							
$L_{RC}2$	2							
$L_{RC}3$	2							
$L_{RC}4$	2							
$L_{RC}5$	2							
$L_{RC}6$	2							
L_R1	6							
L_R2	2							
L_R3	2							
$L_R\gamma4$	10			3	30	2	20	50
$L_R\gamma5$	6			3	18	2	12	30
L_R6	4							
$L_R\gamma7$	4			3	12	2	8	20
$L_R\gamma8$	4			1	4	2	8	12
L_R9	2							
Total	706		120		1430		48	1598

b

Molecule*	MolProbity Scores	Ramachandran plot statistics (%)			RMS deviations	
		Favored	Allowed	Outliers	Bonds Length (Å)	Bonds Angles (°)
Core	1.52	96.53	3.47	0.00	0.004	1.518
Ra/Ra′	1.75	94.43	5.57	0.00	0.009	2.149
Rb/Rb′	1.39	96.32	3.68	0.01	0.007	1.919
Rc/Rc′	1.53	95.96	4.04	0.00	0.009	1.952
Rd/Rd′	1.64	95.77	4.23	0.00	0.012	2.076
Re/Re′	1.62	94.76	5.23	0.01	0.008	2.287
Rf/Rf′	1.78	93.33	6.67	0.00	0.007	1.919
Rg/Rg′	1.44	96.16	3.84	0.00	0.010	2.175
H/H′	1.65	94.64	5.31	0.05	0.013	2.247
M/M′	1.46	95.42	4.58	0.00	0.006	2.239

a, Numbers of proteins and chromophores in the PBS. **b**, Summary of model validation for the PBS components. *Core contains all α subunits, β subunits in core, and L_C1/L_C1', 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′–Rg/Rg′) contains all α subunits, β subunits and linker proteins in the rod; H/H′ contains all extra hexamers (Ha/Ha′, Hb/Hb′, Hc/Hc′ and Hd/Hd′) including all α subunits, β subunits and linker proteins in the hexamers, and L_R9/L_R9'; M/M′ contains individual (αβ) monomers and all individual β subunits.

nature research

Corresponding author(s): Sen-Fang Sui

Last updated by author(s): Dec 6, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

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

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	AutoEMation 1.0
Data analysis	Relion 2.0, Relion 3.0 beta, MotionCor2 1.1.0 , EMAN2.1, CTFFIND4, Phenix 1.14-3260, Coot 0.8.9.1, Pymol 1.8.2.1, Chimera 1.12, CLUSTAL 2.0, ResMap v1.1, MolProbity 4.4, ENDscript 2.0

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The atomic coordinates have been deposited in the Protein Data Bank with the accession code 6KGX. The EM maps have been deposited in the Electron Microscopy Data Bank with the accession codes EMD-9976 for the overall map and EMD-9977 to EMD-9988 for the twelve local maps. All other data and materials are available from the corresponding authors upon reasonable request.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	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.
Data exclusions	The exclusion criteria were not pre-established. 2D and 3D classification yielded multiple classes. Only the particles in the classes that showed clear structural signals and intact structures were selected, combined and used in the final reconstruction and refinement. Details are described in the flowchart of Extended Data Figure 2e and Methods.
Replication	Multiple rounds of structural refinement have been performed and all resulted in same density maps (with different resolutions though). The purification and characterization of PBS (sucrose density gradient centrifugation, SDS-PAGE, absorption spectrum, fluorescence emission spectra) have been repeated for at least three times with similar results. Two batches of sample were analyzed by MS and the similar results confirmed the consistency of our purification method.
Randomization	Randomization of samples is not relevant for a single particle electron microscopy study because the study focused on a specific protein complex.
Blinding	Binding is not relevant because we are studying a specific protein complex.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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