Algal photosystem I dimer and high-resolution model of PSI-plastocyanin complex

Andreas Naschberger[1,7], Laura Mosebach[2,7], Victor Tobiasson[1], Sebastian Kuhlgert[2], Martin Scholz[2], Annemarie Perez-Boerema[1], Thi Thu Hoai Ho[2,6], André Vidal-Meireles[2], Yuichiro Takahashi[3,4], Michael Hippler[2,5] ✉ and Alexey Amunts [1] ✉

Photosystem I (PSI) enables photo-electron transfer and regulates photosynthesis in the bioenergetic membranes of cyanobacteria and chloroplasts. Being a multi-subunit complex, its macromolecular organization affects the dynamics of photosynthetic membranes. Here we reveal a chloroplast PSI from the green alga *Chlamydomonas reinhardtii* that is organized as a homodimer, comprising 40 protein subunits with 118 transmembrane helices that provide scaffold for 568 pigments. Cryogenic electron microscopy identified that the absence of PsaH and Lhca2 gives rise to a head-to-head relative orientation of the PSI–light-harvesting complex I monomers in a way that is essentially different from the oligomer formation in cyanobacteria. The light-harvesting protein Lhca9 is the key element for mediating this dimerization. The interface between the monomers is lacking PsaH and thus partially overlaps with the surface area that would bind one of the light-harvesting complex II complexes in state transitions. We also define the most accurate available PSI–light-harvesting complex I model at 2.3 Å resolution, including a flexibly bound electron donor plastocyanin, and assign correct identities and orientations to all the pigments, as well as 621 water molecules that affect energy transfer pathways.

A chloroplast photosystem I (PSI) of green algae consists of a core complex and three antenna modules: inner belt, outer belt and Lhca2-Lhca9 heterodimer, which together comprise 24 subunits[1–4]. As a short-term light acclimation mechanism in response to fluctuating illumination and anoxia, the algal PSI additionally associates with two light-harvesting complex II (LHCII) trimers[5,6]. Structural studies have shown that the oligomeric state of a chloroplast PSI is a monomer, due to the presence of the subunit PsaH, whereas in cyanobacteria,

structures of dimers[7–9] and trimers[10,11] were also reported. Cyanobacterial PSI oligomerizes via direct contacts between subunits PsaI and PsaL; however, such an association has been ruled out for chloroplast PSI owing to structural constraints of PsaH presence that impose an apparent rigidity[12,13]. Yet, recent structural studies of PSI from a chloroplast of a salt-tolerant alga suggested that its functional core may vary more than previously believed[14]. Functional PsaH-free particles were found, thus showing a potential architectural plasticity of PSI in

[1]Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Solna, Sweden. [2]Institute of Plant Biology and Biotechnology, University of Münster, Münster, Germany. [3]Research Institute for Interdisciplinary Science, Okayama University, Okayama, Japan. [4]Japan Science and Technology Agency-CREST, Saitama, Japan. [5]Institute of Plant Science and Resources, Okayama University, Kurashiki, Japan. [6]Present address: Faculty of Fisheries, University of Agriculture and Forestry, Hue University, Hue, Vietnam. [7]These authors contributed equally: Andreas Naschberger, Laura Mosebach. ✉e-mail: mhippler@uni-muenster.de; amunts@scilifelab.se

response to the ecological environment. On the macromolecular level, an atomic force microscopy analysis of a plant thylakoid membrane showed that, when its architecture is altered upon transition from darkness to light, larger inter-membrane contacts are formed, leading to a reduced diffusion distance for the mobile electron carriers[15]. The membrane architecture in dark- and light-adapted membranes consists of ordered rows of closely packed PSI dimers, which are more abundant in the dark state[15]. Similarly, closely associated PSI–light-harvesting complex I (LHCI) complexes were detected in plants by negative-stain electron microscopy[16], and dimers were found in a subpopulation of PSI from a temperature-sensitive photosystem II mutant alga[17]. This suggests that reversible PSI dimer formation may have a physiological role in thylakoid membrane structure maintenance in chloroplasts. However, very little is known about PSI–LHCI dimers, and information on their structures is lacking. In the absence of high-resolution data, no evidence is available on composition, elements regulating and mediating dimerization, and how the arrangement would differ from the cyanobacterial counterparts. In this article, we present the structure of a chloroplast PSI-LHCI dimer that suggests a structural mechanism for the regulation of dimerisation.

Structure determination

We grew *Chlamydomonas reinhardtii* cells containing a His-tag at the N-terminus of PsaB in low light and under anoxic conditions (Methods). The thylakoid membranes were solubilized with *n*-dodecyl-α-D-maltoside (α-DDM), followed by affinity purification, crosslinking via the chemically activated electron donor plastocyanin (Pc) and sucrose density gradient centrifugation (Extended Data Fig. 1a–f). Two PSI fractions were detected on the sucrose gradient, and 2D polyacrylamide gel electrophoresis (native/reducing 2D-PAGE) of isolated thylakoids indicated the presence of PSI dimers (Extended Data Fig. 1i,j). The heavier green band on the gradient was subjected to single-particle cryogenic electron microscopy (cryo-EM) analysis (Extended Data Table 1). We used 2D classification to separate PSI dimers from monomers in a reference-free manner, followed by 3D classification leading to a subset of 14,173 particles, which were refined to an overall resolution of 2.97 Å by applying C2 symmetry (Extended Data Fig. 2). PSI dimers were also found in 2D class averages in a dataset recorded from a sample without the use of crosslinker. Upon symmetry expansion, the resolution was further improved to 2.74 Å (Extended Data Fig. 2). The remaining 74,209 particles containing the monomer were refined to 2.31 Å resolution, representing a considerable improvement on the previously reported maps[1–3]. A density corresponding to the bound electron donor Pc was found at the lumenal side of both PSI forms, dimer and monomer.

Overall structure

To derive a structure of the chloroplast PSI dimer, we first built an accurate model of one monomer using the 2.74 Å resolution map and then fitted it into the cryo-EM density of the C2 refined dimer. Compared with the monomer, all but two core subunits (PsaH and PsaO) and one light-harvesting protein (Lhca2) are found in the dimer (Fig. 1 and Supplementary Video 1). The structure contains 40 protein subunits, 398 chlorophyll *a* (Chl *a*), 60 chlorophyll *b* (Chl *b*), 56 β-carotenes, 54 luteins, 2 violaxanthins, 2 neoxanthins, 4 phylloquinones, 6 iron–sulfur clusters and 32 lipids (Fig. 1). In addition, two unaccounted densities corresponding to two loops on the stromal side of Lhca9 and PsaG could be interpreted in the dimer, due to the stabilization by the adjacent monomer (Extended Data Fig. 3). The first better-defined density is the Lhca9 loop region 132–153, which is stabilized owing to a direct interaction with PsaL of the second monomer within the dimer. As a result, the area is closely packed with PsaG, and therefore also the PsaG loop region 63–77 is better resolved in the dimer (Extended Data Fig. 3). Finally, co-factors of Lhca9 could be modelled at the interface between the monomers. No density for the His-tag on PsaB could be detected. Overall, taking into account the challenges of modelling
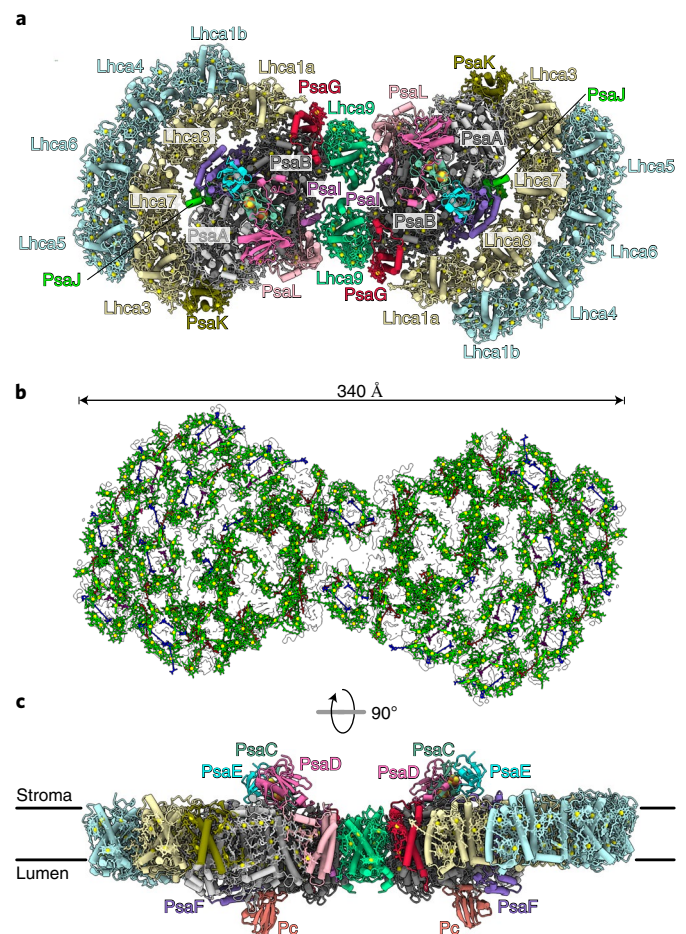

Fig. 1 | Overall structure of the PSI dimer. a, View of individual proteins from stroma. **b**, Arrangement of the pigments in the outline of the map: chlorophylls green (Mg yellow), luteins blue, β-carotenes red, violaxanthin purple and neoxanthin pink. **c**, Overall view along the membrane.

photosynthetic complexes[18], we were able to further improve the quality of the structure, resulting in better validation statistics, compared with the most recent cryo-EM[17] and X-ray crystallography studies[19,20] (Extended Data Table 1). Thus, the current study represents the most complete reference model of PSI.

Structural basis for PSI dimerization

The structural basis for the algal chloroplast PSI dimerization is fundamentally different from cyanobacteria (Fig. 2a). In cyanobacteria, PSI dimerizes via the stromal region of PsaL[7–9] and trimerizes via the lumenal C-terminus of PsaL, assisted by PsaI[11]. In our structure of the chloroplast PSI–LHCI dimer, neither PsaL nor PsaI interacts with each other between the neighbouring units. Instead, PsaH, which normally preserves a monomer, is not present, and Lhca9 with its associated co-factors acts as a symmetrical linker between the monomers, highlighting the importance of the light-harvesting antenna proteins for regulation of the macro-organization. Lhca9 is distinct among the light-harvesting proteins in our structure owing to a truncated loop between helices A and C and lack of the associated chlorophyll[6]. As a result, it contains the fewest chlorophylls among Lhcas (Supplementary Table 1). On the basis of this difference, we rationalized how Lhca9 allows for dimerization, as a longer AC loop would clash with the neighbouring PsaB (Extended Data Fig. 4).

The two Lhca9 copies tether the PSI monomers in a head-to-head fashion, resulting in a 340-Å-long structure (Figs. 1 and 2b, and Supplementary Video 1). They form interactions of four types covering the
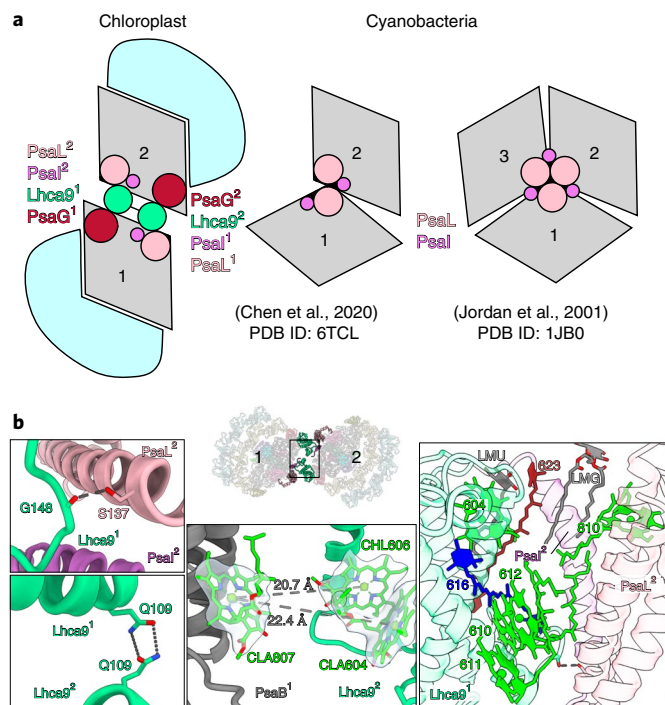
Fig. 2 | Dimerization of PSI. a, Schematic representation: chloroplast PSI–LHCI dimer associated via Lhca9, cyanobacterial PSI dimer (PDB ID: 6TCL, ref. [7]) and trimer (PDB ID: 1JB0, ref. [11]) associated via PsaI/L. The PSI core is grey and LHCI light blue. **b**, The dimer interface is formed by: hydrogen bonds between PsaL and Lhca9, and between two Lhca9 copies (left); potential energy transfer paths between the two monomers (centre); pigments and lipids (right).

entire membrane span: (1) a hydrogen bond of the backbone carbonyl of G148 with S137 of PsaL in the stroma; (2) a hydrogen bond between the two Q109 of the Lhca9 copies; (3) hydrophobic contacts via coordinated co-factors in the membrane that include a newly modelled β-carotene 623 (N2 in nomenclature according to ref. [6]) and five chlorophylls (604, 610, 611, 612 and 810); (4) lipid-mediated hydrophobic interactions via monogalactosyl diglyceride LMG852 and LMU624. One acyl chain of lipid 852 associates with chlorophyll 810 from monomer 2, while the other acyl chain associates with β-carotene BCR623 from monomer 1 (Fig. 2b). Thus, lipids contribute to the oligomerization of PSI, meaning that the membrane itself plays a role in the association. The finding that specific carotenes and lipids enable inter-molecular contacts that bridge the PSI monomers is of a particular interest, as it can be detected only by high-resolution structural studies. Similarly, a recently determined structure of the reaction centre dimer from *Rhodobacter sphaeroides* revealed a bound sulfoquinovosyldiacylglycerol that brings together each monomer forming an S-shaped array[21]. The involvement of lipids in the oligomerization is consistent with the formation of supercomplexes in other bioenergetic membranes[22–25].

To solidify the structural observations, we engineered an *lhca9* insertional mutant having the His-tag at the N-terminus of PsaB and repeated the purification procedure in the same way as for the wild type. This time, no PSI dimer band could be found in the sucrose density gradient (Extended Data Fig. 5). Notably, Lhca9 is present in *Δlhca2*, while Lhca2 is absent from *Δlhca9* (ref. [26]). Moreover, Lhca9 stably associates with PSI–LHCI after sucrose density gradient centrifugation of solubilized thylakoids isolated from *lhca2* insertional mutant (Extended Data Fig. 5).

Implications of PSI dimerization

The specific interactions between the monomers are enabled owing to unoccupied positions of PsaH and Lhca2. As PsaH is also required for the lateral binding of LHCII to the PSI core in state transitions[5,6],

we next compared the structure of PSI–LHCI dimer to the state transition complex (Fig. 3). The superposition shows that Lhca9 from the neighbouring monomer is positioned in the membrane, where Lhca2 resides in PSI–LHCI–LHCII, and their three transmembrane helices would overlap with each other (Fig. 3b). The presence of the PsaH transmembrane helix is not compatible with the Lhca9²-associated co-factors CLA610-612, LMG852 and BCR623 that extend from the neighbouring monomer into the dimer. In addition, the superposition shows that there would be a clash between PsaG and Lhca1 of the inner belt with one of the LHCII trimers, but not the other (Fig. 3b). As Lhca2 and PsaH are absent, the structure of the algal PSI dimer would not facilitate LHCII binding at this position. However, our 2D-PAGE indicated a co-migration of LHCII polypeptides with the dimer fraction, and therefore a structural adaptation cannot be excluded (Extended Data Fig. 1). The antagonistic relationship of Lhca9² and Lhca2 and the assembly state of PsaH might further reflect a regulation of PSI dimerization (Fig. 3a).

PsaH is an 11 kDa transmembrane protein that is imported into chloroplasts and peripherally associates with the PSI reaction centre on the opposite side to the LHCI belt[27]. A recent study of PSI biogenesis apparatus showed that PsaH is assembled in a separate protein module that comes before Lhca2 assembly[28]. *Arabidopsis* mutants of PsaH can grow photoautotrophically[29], and structures of functional algal PSI particles that lack PsaH have been determined[2,14]. In addition, RNA sequencing analysis in algae further confirmed that expression of *psaH* is regulated upon physiological stimuli[30]. It is also of note that PsaH, unlike other PSI subunits, was found to be specifically enriched within the pyrenoid tubules[31]. Together, these data suggest that PsaH-lacking complexes represent a previously overlooked functional form of PSI, where PsaH is either downregulated or could not be assembled. Building on these data, our structure of PSI–LHCI dimer further shows that PsaH/Lhca2-lacking particles can associate with each other to form larger complexes (Figs. 1 and 4). This feature is likely to be conserved in plants, as larger-than-monomer fractions of a plant PSI have been reported[32], and 2D projections from negative stain images of *Arabidopsis* PSI suggest the presence of a putative Lhca1/Lhca4 heterodimer at the PsaL pole that is analogous to alga[33]. The dense organization would be beneficial for membrane crowding and compartmentalization of PSI, even if the two monomers are not energetically coupled.

For other model organisms, such as *Saccharomyces cerevisiae*, whole-cell modelling revealed that metabolic strategies are driven by specific protein expression profiles and compartment-specific proteome constraints[34]. This allows eukaryotic cells to adapt metabolically to nutrient and proteome limitations. Crowding of large complexes in the membrane systems of thylakoid[35,36] and mitochondrial cristae[37], as well as in the cytosol, for example, ribosome dimers[38], is an established and conserved mechanism for a more efficient differentiation of functions, ecological adaptation and material storage during less active periods. As the basic structure of the PSI core is rigid and it operates according to conserved mechanisms, the regulation by PsaH/Lhca2 provides a degree of flexibility on the macromolecular level that would allow the photosynthetic apparatus to adapt to stresses and tolerate changes in the range of light intensities that might involve membrane re-organization[15,39]. Thus, PsaH/Lhca2 appears to be a regulator that defines a late assembly pathway and coordinates the macromolecular organization of PSI in chloroplasts (Fig. 3 and Extended Data Fig. 6). This regulatory role could also be mediated through post-translational modifications such as phosphorylation in *C. reinhardtii*[40] or acetylation in *Arabidopsis*[41]. In *C. reinhardtii*, a consequence of PsaH regulation would be the differential binding of Lhca2. Reduction of Lhca2 alters regulation of photosynthetic electron transfer and hydrogen production, suggesting further potential functional consequences of membrane re-organization and PSI remodelling[26].
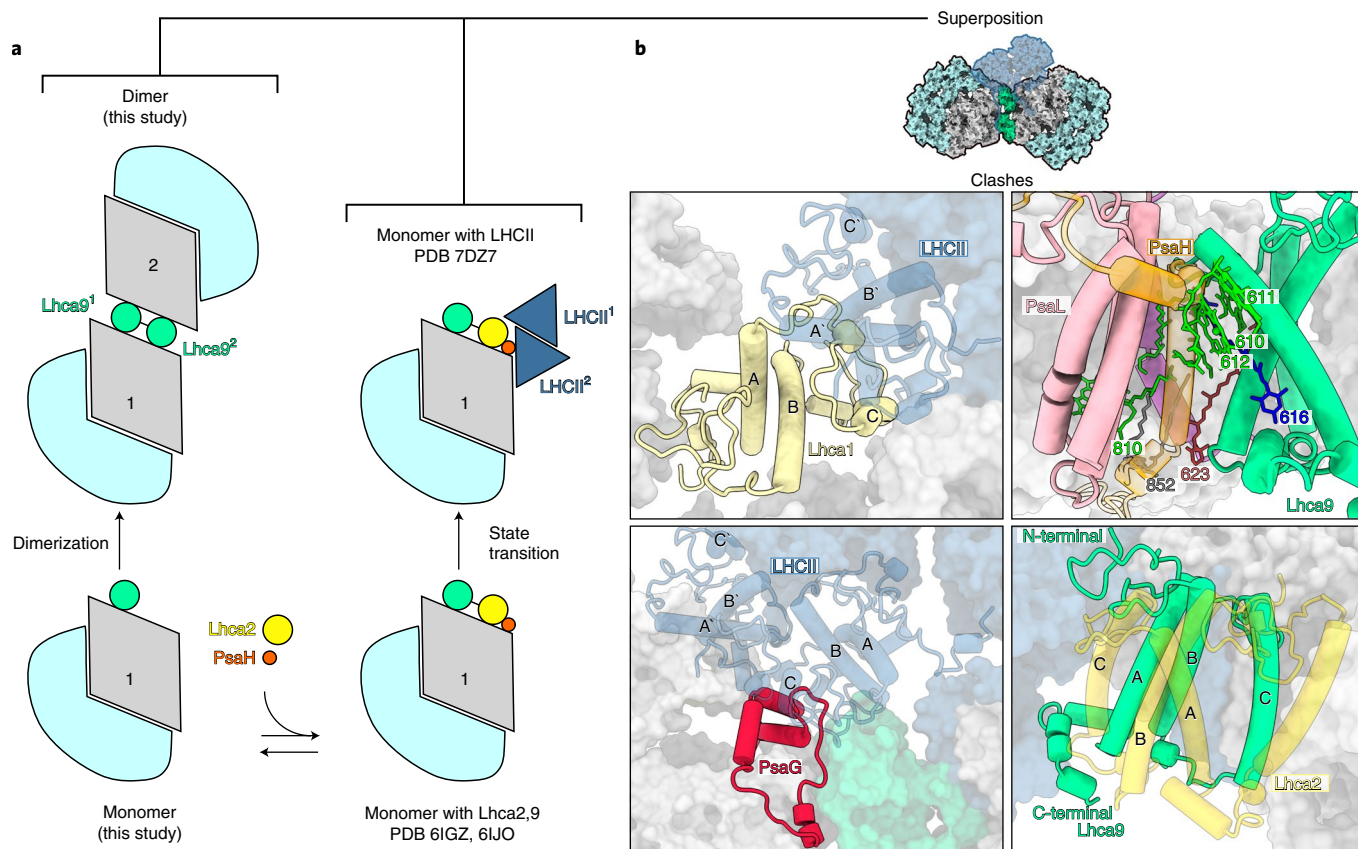
Fig. 3 | PSI dimer and LHCII association in *C. reinhardtii*. a, Schematic view of PSI divarication. The pathway towards state transition or dimer is dependent on presence/absence of PsaH and Lhca2. **b**, Superposition of PSI dimer with PSI–LHCII shows that state transition (PDB ID: 7DZ8, transparent) would result in clashes of Lhca1 (top left) and PsaG (bottom left) with LHCII. PsaH (top right) and Lhca2 (bottom right) would clash with the Lhca9 from the neighbouring PSI monomer.

Coupling of PSI monomers in the dimer

To further extrapolate potential conformational changes during the dimerization of PSI, we applied multi-body refinement analysis of the PSI dimer using the two monomers as bodies (Extended Data Fig. 7). The analysis indicated no distinct conformational states, but instead revealed continuous motions in the three eigenvectors describing a relative movement of the monomers in relation to each other (Extended Data Fig. 7a). The intrinsic flexibility is dominated by combinations of all three rotations of one monomer with respect to the other up to 13° (Extended Data Fig. 7b–d). Therefore, excitation energy transfer between the PSI monomers in the dimeric scaffold would also depend on degrees of rotation around the identified pivot points. Specifically, three chlorophylls are found within a potential cross-monomer excitation sharing: CLA807 (PsaB), CLA604 (Lhca9²) and CHL606 (Lhca9²), and the distance between them is ~20 Å in the consensus map (Fig. 2). While such a positioning might suggest direct coupling, the multi-body analysis indicates considerable variability (Extended Data Fig. 7e,f). Therefore, similarly to the cyanobacterial PSI dimer, an excitation coupling between the two monomers is less favourable in vitro, and this is consistent with measurements of room-temperature and 77 K fluorescence spectra that showed only a minor shift between monomer and dimer (Extended Data Fig. 1g,h). However, in vivo the observed PSI–LHCI dimer conformation, and therefore the distance between the chlorophylls at the interface, could also be affected by a local membrane curvature.

High-resolution features and hydration of PSI

In our PSI monomer reconstruction, the resolution in the core is ~2.1 Å, and in the LHCI inner belt 2.1–2.5 Å, revealing additional structural details of chlorophylls, carotenoids and 621 water molecules (Fig. 4, Extended Data Fig. 8 and Extended Data Table 1). The map can improve the level of detail not only compared with the previous cryo-EM studies of algal PSI[1–3,17], but also the plant PSI maps obtained by X-ray crystallography[19,20] (Extended Data Fig. 8a,b). The quality of the data aided in improving the previous models in functionally important regions. This includes the identification of nine Chl *b* molecules, two newly modelled luteins, a β-carotene and more accurate estimation of the coordination of 53 chlorophylls (Extended Data Figs. 8c and 9, Supplementary Table 1). Particularly, Chl *b* molecules are identified at positions 601 and 606 in Lhca4, Lhca5, Lhca6, Lhca7 and Lhca8. The two newly modelled luteins 720 and 626 are in the N-terminus of Lhca3 next to Chl *a* 614, and in Lhca5 next to Chl *a* 617, respectively (Extended Data Figs. 8c). The newly modelled β-carotene 622 is in Lhca9 and could be identified owing to structural stabilization of the interface region in the dimer (Extended Data Figs. 8c). As Chl *b* limits free diffusion of excitation energy[42], some of the new assignments affect the energy pathways between the antenna proteins. Together with the new structural data, this allowed us to produce a more accurate map of the energy channelling in PSI based on the new model (Fig. 4a).

Another striking feature of the high-resolution cryo-EM map is resolvable density for multiple newly detected water molecules, which particularly aided in modelling the coordination of chlorophylls (Fig. 4b and Supplementary Video 1). Thus, we report the most complete available experimental picture of a chemical environment for chlorophyll binding (Supplementary Table 1). Particularly, it allows one to distinguish between mono- and di-hydrated forms, which largely escaped detection by X-ray crystallography (Extended Data Fig. 8d,e). This is mechanically important because
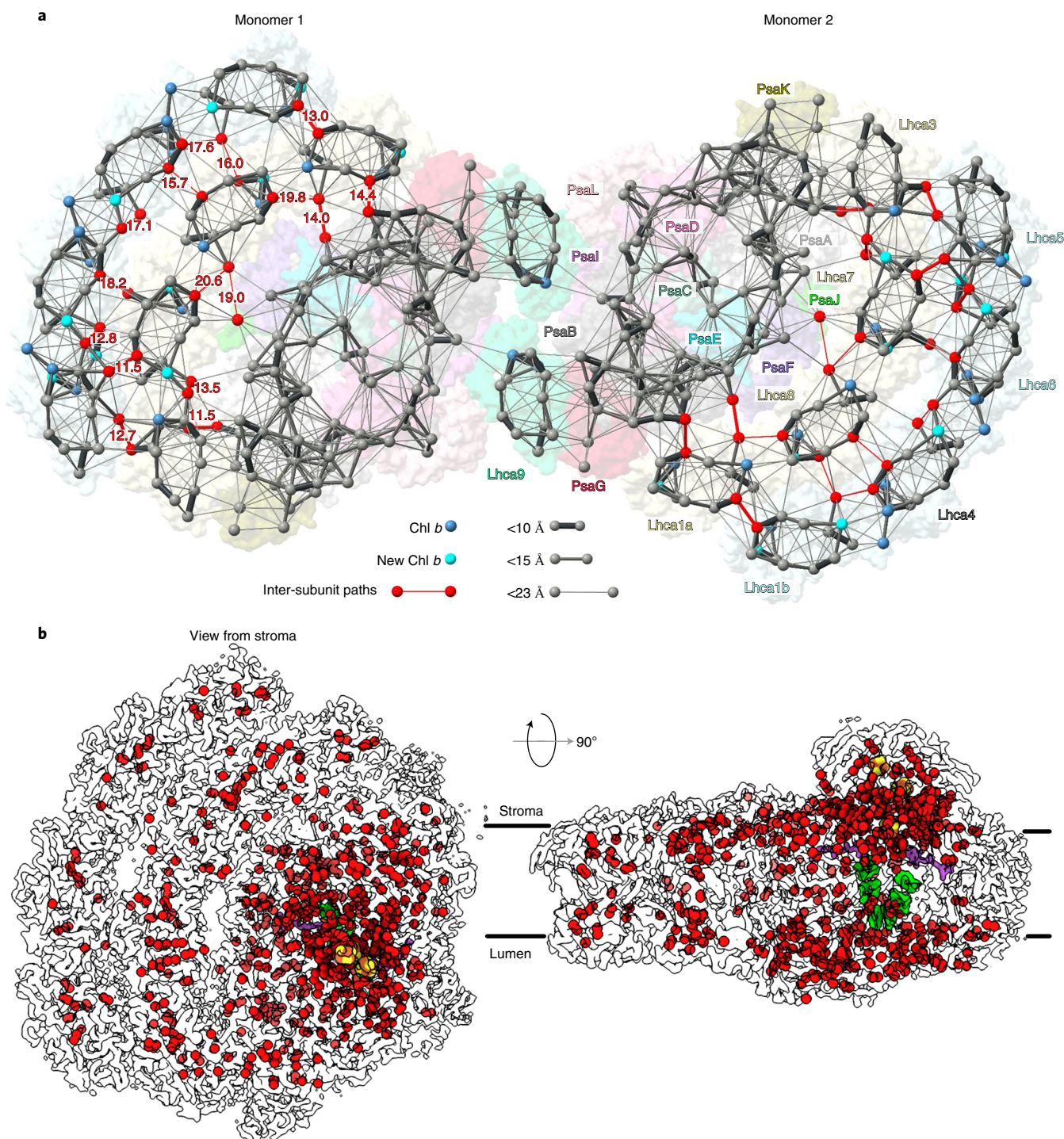
Fig. 4 | High-resolution features of co-factors and hydration. a, Energy transfer pathways. Distribution of chlorophylls indicated by Mg atoms. Chl *a* is grey, Chl *b* blue and newly identified Chl *b* cyan. Pathways within 23 Å are connected by lines, and the line width reflects the distance. The most likely inter-subunit pathways are red. **b**, Water molecules modelled in PSI are shown as red spheres (oxygen atoms) in the outline of the map, chlorophylls green, phylloquinones purple and iron–sulfur clusters yellow-red.

the di-hydrated derivative is chemically more stable, as illustrated by quantum chemical calculations[43]. We observe that, other than the previously reported CLA824 (ref. [20]), only two water molecules can be involved in penta-coordinated Mg for all the chlorophylls. Remarkably, water molecules play a coordinative role for most of Chl *b*, for which the relative ratio of water coordination is four times higher than for Chl *a* (Supplementary Table 1). The difference between Chl *a* and Chl *b* is a methyl versus a formyl group; thus, water serves

as a hydrogen bond donor to the latter, while it also interacts with charged/polar protein residues or lipids. Therefore, the immediate surrounding of Chl *b* molecules is more enriched with non-protein material than previously thought, which plays a role in tuning the photophysics and the transport properties of excitation energy in PSI. Together, the presented model now allows for comparison of PSI phylogenetic conservation also on the level of chlorophyll coordination and solvent positioning.

Structure of PSI-Pc complex

On the lumenal side of PSI, we observed a density corresponding to the associated electron donor Pc, whose binding has been stabilized by crosslinking (Methods). The bound Pc is found on PSI monomers and dimers. Signal subtraction, followed by focused 3D classification, allowed us to rigid body fit a model for Pc into the density at a local resolution of ~3.5 Å (Extended Data Fig. 2). We then performed flexible fitting using self-similarity restraints in Coot[44,45]. With respect to the PSI–Pc interactions, comparison between our model with a plant counterpart[46,47] revealed two main differences (Fig. 5). In *C. reinhardtii*, the negatively charged residues of the Pc acidic patch are shifted by ~5 Å owing to the missing residues P58–E59, and therefore, the interaction with K101 of PsaF is weakened (Fig. 5a,b). Instead, the binding strength is compensated with the PsaF region K78–K92, which has six lysines (78, 81, 82, 85, 89 and 92) increasing a positively charged concentration at a distal site (Fig. 5c–e), thus supporting additional electrostatic interactions with the acidic residues of Pc. The importance of the distal lysine residues is supported by site-directed mutagenesis of the *C. reinhardtii* N-terminal PsaF domain and functional analyses of electron transfer between Pc and mutant as well as wild-type PSI[48]. Thus, algal and plant Pc have adapted their slightly different interfaces for optimal interactions with PSI.

In summary, this study illustrates that PSI–LHCI can dimerize and explains how the process is structurally regulated by subunits PsaH, Lhca2 and Lhca9 (Fig. 3a and Extended Data Fig. 6). It remains to be seen how thylakoid regulatory networks manage to implement PsaH allocation strategies. The data provide the most complete description so far of the structure of PSI, including newly identified co-factors that play specific roles in accommodating the light harvesting and excitation transfer functions, as well as water molecules involved in chlorophyll coordination. Finally, the binding of electron donor Pc is resolved. Together, the data explain how the PSI is modulated to perform its functional and structural roles in a chloroplast.

Methods

Strains and growth conditions

Experiments were performed using a strain expressing PsaB fused with His$_{20}$-tag after the third residue from the N-terminus[2] and an *Δlhca9* insertional mutant[49] transformed with the corresponding PsaB His-tag plasmid. Chloroplast transformation was performed using a particle gun[50], and transformants were selected on Tris-acetate-phosphate (TAP) medium containing 150 µg spectinomycin ml^{-1} and recloned three to four times until they were homoplasmic[2]. Further experiments were performed with the *C. reinhardtii* wild-type strain 137c as well as *Δlhca2* insertional mutant[49] back-crossed into wild-type 137c. All strains were maintained on TAP medium, solidified with 1.5% w/v agar at 25 °C in the presence of ~20 µmol photons m^{-2} s^{-1} photosynthetically active, continuous illumination. For experiments, strains were cultured in TAP medium on a rotary shaker (120 rpm) at 25 °C in the presence of ~20 µmol photons m^{-2} s^{-1} photosynthetically active, continuous illumination.

Purification of PSI

C. reinhardtii cells were incubated in anoxic conditions (~10^8 cells ml^{-1} in TAP medium + 10 mM glucose, 40 U ml^{-1} glucose oxidase and 50 U ml^{-1} catalase) and dim light (~20 µmol photons m^{-2} s^{-1}) for 60 min. All following steps were performed at 4 °C and dim light. Cells were disrupted in 0.33 M sucrose, 25 mM HEPES–KOH pH 7.5, 5 mM MgCl$_2$, 1 mM PMSF, 1 mM benzamidine and 5 mM aminocaproic acid with a nebulizer (2 bar, two passages). Broken cells were centrifuged at 32,800g for 10 min (Beckman Coulter JA-25.50 rotor, 20,000 rpm). The pellet was carefully resuspended in 0.5 M sucrose, 5 mM HEPES–KOH pH 7.5, 10 mM EDTA, 1 mM benzamidine and 5 mM aminocaproic acid with a potter homogenizer. The resuspended material was layered on top of a sucrose density step gradient (1.8 M and 1.3 M sucrose, 5 mM HEPES–KOH pH 7.5,

10 mM EDTA, 1 mM benzamidine and 5 mM aminocaproic acid). Thylakoid membranes were extracted via ultracentrifugation at 70,800g for 1 h and 20 min (Beckman Coulter SW 32 Ti rotor, 24,000 rpm). Thylakoids were collected from the step gradient interphases with a Pasteur pipet, diluted four times with 5 mM HEPES–KOH pH 7.5 and centrifuged at 37,900g for 20 min (Beckman Coulter JA 25.50 rotor, 21,500 rpm).

Isolated thylakoids were set to 1 mg chlorophyll ml^{-1} in 5 mM HEPES–KOH pH 7.5 and solubilized by addition of an equal volume of 2% α-DDM for 10 min. Unsolubilized material was separated by centrifugation. The supernatant was diluted four times to 125 µg chlorophyll ml^{-1} and 0.25% α-DDM. The sample was loaded onto a TALON metal affinity column (1 ml resin mg chlorophyll^{-1}) in 5 mM HEPES–KOH pH 7.5, 100 mM NaCl, 5 mM MgSO$_4$ and 10% glycerol at a flow rate of ~0.5 ml min^{-1}. The column was washed with ten times the bed volume of 5 mM HEPES–KOH pH 7.5, 100 mM NaCl, 5 mM MgSO$_4$, 10% glycerol and 0.02% α-DDM at a flow rate of ~1 ml min^{-1}. A second wash was performed with ten times the bed volume of 5 mM HEPES–KOH pH 7.5, 100 mM NaCl, 5 mM MgSO$_4$, 10% glycerol, 0.02% α-DDM and 5 mM imidazole at a flow rate of ~1 ml min^{-1}. The PSI was eluted with 5 mM HEPES–KOH pH 7.5, 100 mM NaCl, 5 mM MgSO$_4$, 10% glycerol, 0.02% α-DDM and 150 mM imidazole. The PSI was concentrated with a spin column (regenerated cellulose: 100,000 molecular weight cut-off (MWCO)) to ~1.5 mg chlorophyll ml^{-1}, diluted five times with 30 mM HEPES–KOH pH 7.5 and 0.02% α-DDM and reconcentrated twice.

PSI-Pc crosslinking was performed in 30 mM HEPES–KOH pH 7.5, 1 mM ascorbate, 0.1 mM diaminodurene (a redox mediator between ascorbate and Pc), 3 mM MgCl$_2$ with 0.1 mg chlorophyll ml^{-1} PSI particles and 20 µM activated Pc for 45 min at room temperature. Pc activation was performed in 10 mM MOPS–KOH pH 6.5 with 100 µM recombinant Pc, 5 mM 1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride and 10 mM sulfo-*N*-hydroxysuccinimide for 20 min at room temperature. The crosslinker was removed and the buffer exchanged to 30 mM HEPES–KOH pH 7.5 via a PD G25 desalting column followed by ultrafiltration with a centricon (regenerated cellulose: 10,000 MWCO).

The crosslinked PSI-Pc particles (~60 µg chlorophyll per gradient) or solubilized thylakoids of non-tagged strains (~250 µg chlorophyll per gradient) were loaded onto a 1.3 M to 0.1 M sucrose density gradient including 5 mM HEPES–KOH pH 7.5 and 0.02% α-DDM. PSI fractions were collected after ultracentrifugation at 134,400g (Beckman Coulter SW 41 Ti rotor, 33,000 rpm) for 14 h (Extended Data Figs. 1b and 5). Before further analysis, sucrose was removed via a PD G25 desalting column followed by concentration with a spin column (regenerated cellulose: 100,000 MWCO).

Biochemical analysis of PSI

For SDS–PAGE (Extended Data Fig. 1c–e), samples were adjusted to 1 µg chlorophyll, supplemented with loading buffer and incubated at 65 °C for 15 min. Proteins were separated by 13% (w/v) SDS–PAGE[51]. Gels were stained with Coomassie Brilliant Blue R-250 or blotted onto nitrocellulose membranes (Amersham) and detected by specific primary antibodies: PsaF (ref. [52]), Pc (Agrisera), PsaA, Lhca5 (ref. [4]), Lhca2 (ref. [4]), Lhca9 (ref. [4]), Lhca3 (ref. [53]), PsaD (ref. [54]), PsaG (ref. [55]), Lhcb/a (ref. [56]), PsbA (Agrisera) and LhcSR3 (ref. [77]). The antibody against PsaA was raised using the peptides STPEREAKKVKIAVDR and VKIAVDRNPVETS-FEK and was obtained from Eurogentec. All primary antibodies were used at a 1:1,000 dilution, except for anti-Lhcb/a (1:2,500) and anti-PsbA (1:10,000). Secondary antibodies for ECL detection were used at a 1:10,000 dilution (goat anti-rabbit IgG (H + L)-HRP conjugate, Bio-Rad).

For quantitative analysis by mass spectrometry, removal of sucrose and protein digestion was carried out following the filter-aided sample preparation, using 2 µg of sequencing grade trypsin (Promega) per fraction[57]. Iodoacetamide and dithiothreitol used in the original protocol were replaced by chloroacetamide and tris(2-carboxyethyl) phosphine, respectively. After overnight digestion at 37 °C, samples
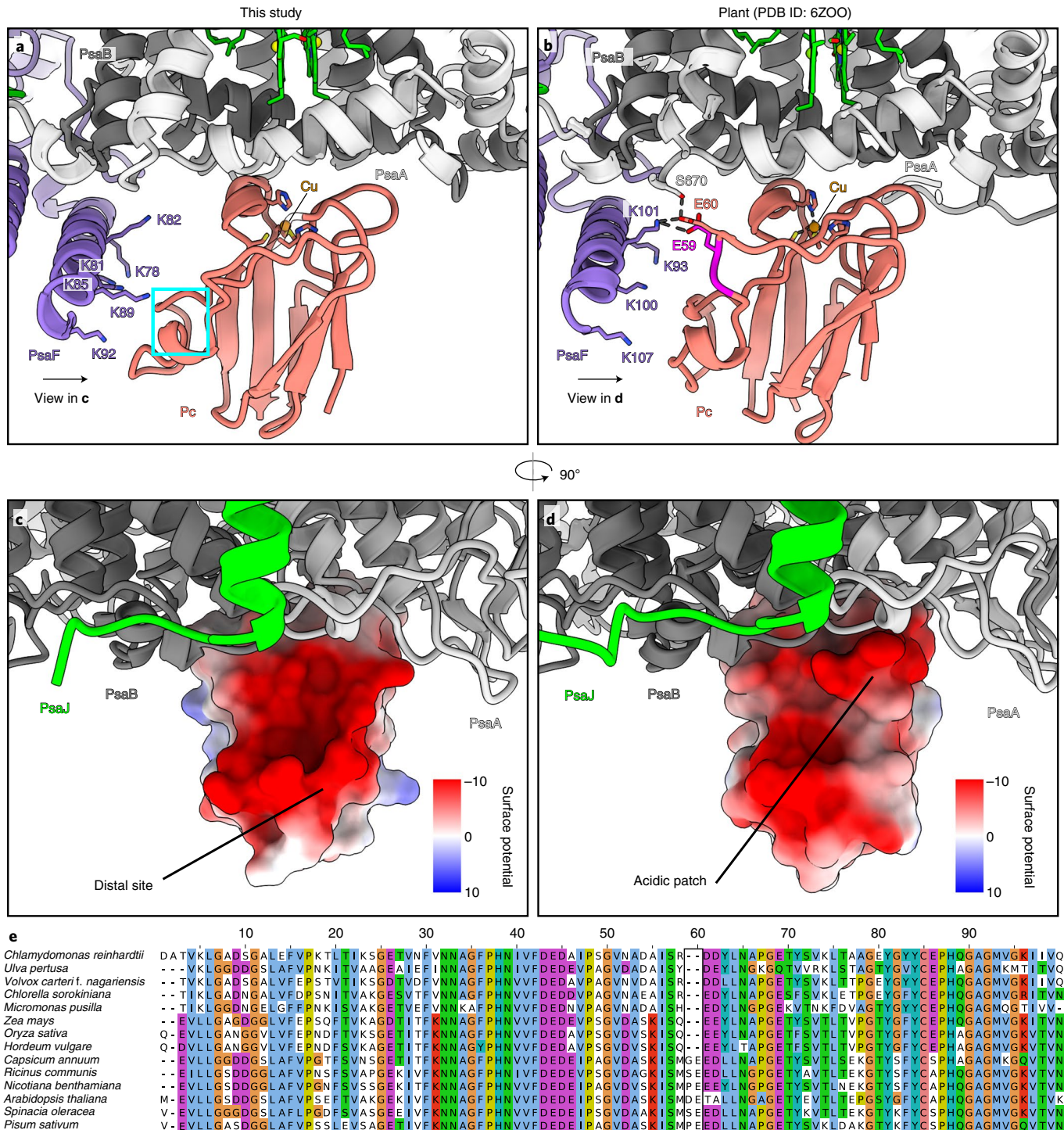
Fig. 5 | Pc binding site. a, Pc binding in *C. reinhardtii* (current work). The positively charged residues of PsaF stabilize the interactions. The corresponding Pc region that deviates from the plant counterpart is cyan. **b**, Pc binding in plants[29]. The two inserted residues are magenta. **c**, Ninety-degree rotated view with Pc surface shown with Coulomb potential from the interface. **d**, The same view for a plant counterpart (PDB ID: 6ZOO), illustrating that the acidic patch is shifted. **e**, Multiple sequence alignment of Pc from different species of the green lineage (algae and plants) showing that the inserted residues 58 and 59 occur in a subset of plants and do not represent a general case.

were acidified by adding trifluoroacetic acid to a final volume of 0.1%. Five percent of the peptide solution was desalted using self-made Sta-geTips according to established protocols[58]. Desalted peptides were dried by vacuum centrifugation and stored at −20 °C until further use. The liquid chromatography (LC)–tandem mass spectrometry (MS/MS) system consisted of an Ultimate 3000 RSLC nanoLC System (Thermo Fisher Scientific) coupled via a Nanospray Flex ion source

(Thermo Fisher Scientific) to a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific). Samples were reconstituted in 5 µl of 2% (v/v) acetonitrile/0.05% (v/v) trifluoroacetic acid in ultrapure water (eluent A1), loaded on a trap column (C18 PepMap 100, 300 µM × 5 mm, 5 µm particle size, 100 Å pore size; Thermo Fisher Scientific) and desalted for 3 min at a flow rate of 15 µl min⁻¹ using eluent A1. Subsequently, the trap column was switched in-line with an Acclaim

PepMap100 reversed phase column (75 μm × 50 cm, 2 μm particle sizes, 100 Å pore size; Thermo Fisher Scientific) for peptide separation. The mobile phases were composed of 0.1% (v/v) formic acid in ultrapure water (eluent A2) and 80% (v/v) acetonitrile/0.08% (v/v) formic acid in ultrapure water (B). Flow rate was 250 nl min⁻¹. The following gradient was applied: 5–35% B over 105 min, 35–99% B over 5 min and 99% B for 20 min. Mass spectrometry full scans (scan range m/z 350–1,400, resolution 70,000 at m/z 200, automatic gain control) target value $3 × 10^6$, maximum injection time 50 ms) were acquired in data-dependent mode, dynamically selecting the 12 most abundant precursor ions for fragmentation by higher-energy C-trap dissociation (27% normalized collision energy, resolution 17,500 at m/z 200, precursor isolation window 1.5 m/z). Dynamic exclusion was set to 'auto' (chromatographic peak width 15 s). AGC target value and intensity threshold for MS/MS were $5 × 10^4$ and $1 × 10^4$, respectively, at 80 ms maximum ion fill time. Singly charged ions, ions with charge state 5 or above and ions with unassigned charge states were rejected. Internal lock mass calibration was enabled on m/z 445.12003. LC–MS/MS data were processed in MaxQuant 1.6.14 for protein identification and label-free quantification[59]. Default settings were used, except for calculation of intensity-based absolute quantitation (iBAQ) values and 'second peptide search', which were enabled and disabled, respectively. iBAQ values were normalized to the PSI core subunit PsaB, and values below 21 were excluded as they represent already low intensity values, which might not be reliable. Spectra were searched against a concatenated database containing protein sequences based on the *Chlamydomonas* v5.6 gene models (Joint Genome Institute, www.phytozome.org), as well as sequences of chloroplast- and mitochondrial-encoded proteins (GenBank BK000554.2 and NC_001638.1). Carbamidomethylation of cysteines was set as a fixed modification. Oxidation of methionine and acetylation of protein N-termini were considered as variable modifications. A false discovery rate of 1% was applied to peptide and protein identifications. Label-free quantitation data were imported into Perseus (version 1.6.15.0)[60] for log₂ transformation, and contaminants, proteins identified only by site and reverse hits, were removed.

Room-temperature absorption spectra (300–750 nm) of PSI monomer and dimer fractions (Extended Data Fig. 1g) were measured with an ultraviolet–visible spectrophotometer (V-650, Jasco) at 10 μg chlorophyll ml⁻¹ and normalized to the red region. Fluorescence emission spectra (650–780 nm) of PSI fractions (Extended Data Fig. 1h) were recorded at 77 K at 1 μg chlorophyll ml⁻¹ with a spectrofluorometer (P-6500, Jasco) upon excitation at 436 nm. Spectra were normalized to the maximum of the emission peaks and smoothed according to Savitzky–Golay[77] using the Jasco spectra analysis program (Spectra Manager II). For 2D-PAGE, thylakoids (0.8 mg chlorophyll ml⁻¹) isolated from control and anoxic 137c wild-type cells were solubilized with 0.9% β-DDM for 20 min. Two-dimensional PAGE[61] and silver staining[52] were performed as described. Excised spots from silver-stained blue native PAGE gels were destained by incubation with 15 mM potassium hexacyanoferrate (III)/50 mM sodium thiosulfate for 8 min and then submitted to tryptic in-gel digestion[62]. No reduction and alkylation of cysteines was performed. LC–MS/MS was implemented[54], where an Ultimate 3000 nano-LC system was coupled via a nanospray source to an LCQ Deca XP Plus mass spectrometer (Thermo Finnigan).

LC–MS/MS data were processed with Proteome Discoverer (Thermo Fisher Scientific, version 2.4). Raw files were searched using the SequestHT and MS Amanda algorithms against a concatenated database containing sequences of nuclear- (Chlamydomonas v5.6 gene models, www.phytozome.org), chloroplast- (GenBank BK000554.2) and mitochondrial-encoded (NC_001638.1) proteins. Search settings were: precursor and fragment mass tolerances: 250 ppm and 0.25 Da, respectively; minimum peptide length: 6; maximum of missed cleavages: 2; variable modifications: oxidation of methionine, N-acetylation of protein N-termini. Identifications were filtered to achieve a peptide and protein level false discovery rate of 0.01.

Cryo-EM data collection and processing

Three microlitres of purified PSI complex at 1 mg chlorophyll ml⁻¹ was applied to glow-discharged (GloQube Quorum, 40 s, 20 mA) holey carbon grids (Quantifoil 300 Au R1.2/R1.3, Electron Microscopy Sciences) and vitrified using a Vitrobot MKIV (2 s blotting time, 4 °C, 100% humidity). The data collection was carried out using a 300KV Titan Krios G2 Microscope (Thermo Fisher Scientific) equipped with a Gatan Bioquantum energy filter and a K3 Summit direct electron detector (Ametek). Movies were recorded using counting mode at a magnification of ×105,000 corresponding to a calibrated pixel size of 0.84 Å. The dose rate was 15.27 e pixel⁻¹ s⁻¹ and the exposure time was 3 s divided into 45 frames, leading to a total dose of 45.8 e Å⁻². Software E Pluribus Unum (EPU) was used to collect 17,439 movies with a defocus range from −0.7 μm to −2.5 μm. Data statistics are shown in Extended Data Table 1.

Extended Data Fig. 2 shows the processing scheme applied. The pre-processing steps were performed using cryoSPARC 3.1.0 (ref. [63]). Movie stacks were motion corrected and dose weighted using MotionCor2 (ref. [64]). Contrast transfer function (CTF) of the motion-corrected micrographs was estimated using CTFFIND4 (ref. [65]). Blob picker and then template picker were used to pick 440,494 particles, and 2D classification in cryoSPARC was performed. Dimeric particles were separated from monomeric by inspection of the 2D-class averages, and for each subpopulation an *ab initio* model was generated using cryoSPARC applying C2 and C1 symmetry, respectively. For each model homogeneous refinement was performed, leading to a nominal resolution of 3.7 Å for the dimer and 3.0 Å for the monomer. Particles (dimer, 69,144; monomer, 123,746) were converted into a Star file format[66] and imported into RELION 3.1.beta[67,68]. Particles were re-extracted (un-binned) and processed in RELION using a box size of 700 pixel and 500 pixel for the dimer and monomer, respectively. Three-dimensional refinement followed by 3D classification was performed imposing C2 symmetry for the dimer and C1 for the monomer. A subset of high-quality particles was selected for the dimer and monomer and subjected to 3D refinement, which resulted in a resolution of 3.3 Å for the dimer and 2.9 Å for the monomer. CTF refinement[69,70] followed by 3D refinement and Bayesian polishing followed by another round of CTF refinement was performed for the dimer as well as for the monomer. A final 3D refinement resulted in an overall resolution of 2.97 Å for the dimer and 2.31 Å for the monomer. The resolution of the dimer could be further improved to 2.74 Å by using signal subtraction of one monomer followed by symmetry expansion and 3D refinement applying C1 symmetry.

To increase the number of particles for classification on the Pc region, dataset 2 was collected from the same dimer band, but with a pixel size of 0.51 Å. The dataset was processed with cryoSPARC 3.1.0 (ref. [63]). After template picking, 864,399 particles were extracted. With a small subset of the extracted particles an *ab initio* reconstruction was generated followed by heterogeneous refinement using five classes, one of which contained *ab initio* reconstruction as reference. The class containing the PSI monomer was then subjected to homogeneous refinement in cryoSPARC 3.1.0, resulting in a reconstruction at 3.88 Å resolution. The particles were then exported to RELION[71], and 3D classification was performed. The class that contained 88,219 good particles was used for further refinement, which improved the overall resolution to 3.5 Å. Applying CTF refinement and Bayesian polishing resulted in further improvement, and the final nominal resolution is 2.68 Å. The data were then merged with the monomer (dataset 1), and signal subtraction followed by focused classification using a mask around the Pc region was performed. A class with 66,080 particles showed the best density for Pc, which was used for model building. The workflow is further illustrated in Extended Data Fig. 2a.

For analysis of the motion between the two monomers of the dimer, we performed a multi-body refinement in RELION[71] followed by a principal component analysis using the program relion_flex_analyse.

Two bodies were chosen, one for each monomer, resulting in 12 eigenvectors describing the motion. Ten maps for each of the three eigenvectors that describe about 78% of the motion in the data were printed out, and the maps with the extreme positions (maps 1 and 10) were used to fit the models that are shown in Extended Data Fig. 6. A Python script was used to estimate the distances between the chlorophylls at the dimer interface for each particle in the data and to plot the results as histograms as depicted in Extended Data Fig. 7e,f.

Model building and refinement

Initially, the available model of the PSI structure (Protein Data Bank (PDB) ID: 6JO5) of *C. reinhardtii* was rigid body fitted into the 2.74 Å map of the symmetry expanded dimer using Chimera v 1.14 (ref. [72]). Model building and real-space refinement was then carried out using Coot v9.1.4 (refs. [44,45]) to complete one monomer. Two copies of the completed monomeric model were then rigid body fitted into the C2 generated 2.97 Å dimer map using Chimera. The model of the monomer was then fitted separately in the highest-resolution 2.3 Å map of the monomer. All protein residues as well as pigments were fitted using Coot[44,45] with locally optimized map weights. The *cis–trans* isomerism of each pigment was judged on the basis of density and modelled accordingly. Newly identified chlorophylls and carotenoids were modelled, when the experimental evidence (density map) supported and the chemical environment matched the surrounding of the pigments. For carotenoids, the density that clearly showed the characteristics of an elongated tetraterpenoid with densities for the four methyl groups, two sticking out on each side of the chain, was identified as a new carotenoid binding site. To further analyse the identity of the corresponding pigment, possible candidates were fitted and compared. A carotenoid that fitted best in terms of density and chemical environment was then selected. In case of luteins, the oxygen of the cyclohexane ring was the main criterion for pigment identity because it is involved in hydrogen bonding. For Chl *b* identification, the densities for the aldehyde group needed to be present as well as the hydrogen bonding occurring with a water molecule that are usually stabilized by other chlorophylls, lipids and protein side chains. Water molecules involved in pigment interactions were placed manually. All other water molecules were picked by Coot[44,45] with the autopicking function followed by manual inspection and correction. All high-resolution features were modelled using Coot[44,45] until the model was completed. For all modelling steps, restraint files for pigments and ligands were used that were generated using the Grade server (http://grade.globalphasing.org). Restraint files were adopted manually if it was required.

For Pc, a model was generated using SWISS model[73]. The model was then rigid body fitted using Chimera. Rotamers were corrected for the residues that were allowed owing to the better local densities. Self-restraints in Coot were activated followed by flexible fitting into the density. All models were refined using Real-Space-Refine from the PHENIX suite[74] using the Grade server restraint files for the ligands and a distance.edit file that was generated by Ready-set in PHENIX. Further, hydrogen atoms were added for refinement to the model using Ready-set. The refinement protocol was optimized using different weight parameters. The refinement statistics are presented in Extended Data Table 1. Multiple rounds of validation and model building were carried out using MolProbity[75] and Coot[44,45]. For further validation, the PDB Validation server was used (https://validate-rcsb-2.wwpdb.org). The structure was analysed using Coot and Chimera. Figures were prepared using ChimeraX[76].

Reporting summary

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

Data availability

The atomic coordinates have been deposited in the RCSB PDB, and EM maps in the Electron Microscopy Data Bank under accession codes: 7ZQD and EMD-14871 (dimer), 7ZQ9 and EMD-14867 (symmetry expanded dimer), 7ZQC and EMD-14870 (monomer), and 7ZQE and EMD-14872 (PSI–Pc). Mass spectrometry datasets: DOI: 10.6019/PXD026990; project accession ID: PXD027067. The following atomic coordinates were used in this study: 1JB0 (PSI trimer from *Synechococcus elongatus*), 6JO5 (PSI from *C. reinhardtii*), 6TCL (PSI tetramer from *Anabaena*), 6ZOO (PSI-Pc from *Pisum sativum*) and 7DZ8 (PSI-LHCII from *C. reinhardtii*). Source data are provided with this paper.

Code availability

All the codes that were used in this study are publicly accessible, and the script for conversion of poses from Euler space into an orientation matrix in Cartesian space is available upon request.

References

1. Qin, X. et al. Structure of a green algal photosystem I in complex with a large number of light-harvesting complex I subunits. *Nat. Plants* **5**, 263–272 (2019).

2. Suga, M. et al. Structure of the green algal photosystem I supercomplex with a decameric light-harvesting complex I. *Nat. Plants* **5**, 626–636 (2019).

3. Su, X. et al. Antenna arrangement and energy transfer pathways of a green algal photosystem-I–LHCI supercomplex. *Nat. Plants* **5**, 273–281 (2019).

4. Ozawa, S.-I. et al. Configuration of ten light-harvesting chlorophyll *a/b* complex I subunits in *Chlamydomonas reinhardtii* photosystem I. *Plant Physiol.* **178**, 583–595 (2018).

5. Huang, Z. et al. Structure of photosystem I-LHCI–LHCII from the green alga *Chlamydomonas reinhardtii* in state 2. *Nat. Commun.* **12**, 1100 (2021).

6. Pan, X. et al. Structural basis of LHCBM5-mediated state transitions in green algae. *Nat. Plants* **7**, 1119–1131 (2021).

7. Chen, M. et al. Distinct structural modulation of photosystem I and lipid environment stabilizes its tetrameric assembly. *Nat. Plants* **6**, 314–320 (2020).

8. Kato, K. et al. Structure of a cyanobacterial photosystem I tetramer revealed by cryo-electron microscopy. *Nat. Commun.* **10**, 4929 (2019).

9. Zheng, L. et al. Structural and functional insights into the tetrameric photosystem I from heterocyst-forming cyanobacteria. *Nat. Plants* **5**, 1087–1097 (2019).

10. Kato, K. et al. Structural basis for the absence of low-energy chlorophylls in a photosystem I trimer from *Gloeobacter violaceus*. *eLife* **11**, e73990 (2022).

11. Jordan, P. et al. Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* **411**, 909–917 (2001).

12. Ben-Shem, A., Frolow, F. & Nelson, N. Crystal structure of plant photosystem I. *Nature* **426**, 630–635 (2003).

13. Amunts, A. & Nelson, N. Plant photosystem I design in the light of evolution. *Structure* **17**, 637–650 (2009).

14. Perez-Boerema, A. et al. Structure of a minimal photosystem I from the green alga *Dunaliella salina*. *Nat. Plants* **6**, 321–327 (2020).

15. Wood, W. H. et al. Dynamic thylakoid stacking regulates the balance between linear and cyclic photosynthetic electron transfer. *Nat. Plants* **4**, 116–127 (2018).

16. Yadav, K. N. S. et al. Supercomplexes of plant photosystem I with cytochrome b6f, light-harvesting complex II and NDH. *Biochim. Biophys. Acta.* **1858**, 12–20 (2017).

17. Caspy, I. et al. Dimeric and high-resolution structures of *Chlamydomonas* photosystem I from a temperature-sensitive photosystem II mutant. *Commun. Biol.* **4**, 1380 (2021).

18. Amunts, A. The revolution evolution. *Nat. Plants* **8**, 14–17 (2021).

19. Mazor, Y., Borovikova, A., Caspy, I. & Nelson, N. Structure of the plant photosystem I supercomplex at 2.6 Å resolution. *Nat. Plants* **3**, 17014 (2017).

20. Wang, J. et al. Structure of plant photosystem I–light harvesting complex I supercomplex at 2.4 Å resolution. *J. Integr. Plant Biol.* **63**, 1367–1381 (2021).

21. Qian, P. et al. Cryo-EM structure of the dimeric *Rhodobacter sphaeroides* RC–LH1 core complex at 2.9 Å: the structural basis for dimerisation. *Biochem. J.* **478**, 3923–3937 (2021).

22. Mühleip, A., McComas, S. E. & Amunts, A. Structure of a mitochondrial ATP synthase with bound native cardiolipin. *eLife* **8**, e51179 (2019).

23. Flygaard, R. K., Mühleip, A., Tobiasson, V. & Amunts, A. Type III ATP synthase is a symmetry-deviated dimer that induces membrane curvature through tetramerization. *Nat. Commun.* **11**, 5342 (2020).

24. Mühleip, A. et al. ATP synthase hexamer assemblies shape cristae of *Toxoplasma* mitochondria. *Nat. Commun.* **12**, 120 (2021).

25. Bennett, C. F. et al. Peroxisomal-derived ether phospholipids link nucleotides to respirasome assembly. *Nat. Chem. Biol.* **17**, 703–710 (2021).

26. Ho, T. T. et al. Photosystem I light-harvesting proteins regulate photosynthetic electron transfer and hydrogen production. *Plant Physiol.* **189**, 329–343 (2022).

27. Amunts, A., Drory, O. & Nelson, N. The structure of a plant photosystem I supercomplex at 3.4 Å resolution. *Nature* **447**, 58–63 (2007).

28. Nellaepalli, S., Ozawa, S.-I., Kuroda, H. & Takahashi, Y. The photosystem I assembly apparatus consisting of YCF3–Y3IP1 and YCF4 modules. *Nat. Commun.* **9**, 2439 (2018).

29. Jensen, P. E., Haldrup, A., Rosgaard, L. & Scheller, H. V. Molecular dissection of photosystem I in higher plants: topology, structure and function. *Physiol. Plant.* **119**, 313–321 (2003).

30. Binte Safie, S. R., Ng, Y. K., Yao, L. & Lee, Y. K. Growth bottlenecks of microalga *Dunaliella tertiolecta* in response to an up-shift in light intensity. *Eur. J. Phycol.* **53**, 509–519 (2018).

31. Mackinder, L. C. M. et al. A spatial interactome reveals the protein organization of the algal CO_2-concentrating mechanism. *Cell* **171**, 133–147.e14 (2017).

32. Amunts, A., Toporik, H., Borovikova, A. & Nelson, N. Structure determination and improved model of plant photosystem I. *J. Biol. Chem.* **285**, 3478–3486 (2010).

33. Crepin, A., Kučerová, Z., Kosta, A., Durand, E. & Caffarri, S. Isolation and characterization of a large photosystem I–light-harvesting complex II supercomplex with an additional lhca1–A4 dimer in *Arabidopsis*. *Plant J.* **102**, 398–409 (2020).

34. Elsemman, I. E. et al. Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies. *Nat. Commun.* **13**, 801 (2022).

35. Shen, L. et al. Architecture of the chloroplast psi–NDH supercomplex in *Hordeum vulgare*. *Nature* **601**, 649–654 (2021).

36. Su, X. et al. Supramolecular assembly of chloroplast NADH dehydrogenase-like complex with photosystem I from *Arabidopsis thaliana*. *Mol. Plant* **15**, 454–467 (2022).

37. Gahura, O. et al. An ancestral interaction module promotes oligomerization in divergent mitochondrial ATP synthases. Preprint at *bioRxiv* https://doi.org/10.1101/2021.10.10.463820 (2022).

38. Matzov, D. et al. The cryo-EM structure of hibernating 100s ribosome dimer from pathogenic S*taphylococcus aureus*. *Nat. Commun.* **8**, 723 (2017).

39. Amunts, A. & Nelson, N. Functional organization of a plant photosystem I: evolution of a highly efficient photochemical machine. *Plant Physiol. Biochem.* **46**, 228–237 (2008).

40. Bergner, S. V. et al. State transition7-dependent phosphorylation is modulated by changing environmental conditions, and its absence triggers remodeling of photosynthetic protein complexes. *Plant Physiol.* **168**, 615–634 (2015).

41. Bienvenut, W. V. et al. Dual lysine and N-terminal acetyltransferases reveal the complexity underpinning protein acetylation. *Mol. Syst. Biol.* **16**, e9464 (2020).

42. Croce, R. & van Amerongen, H. Light harvesting in oxygenic photosynthesis: structural biology meets spectroscopy. *Science* **369**, e9464 (2020).

43. Fredj, A. B., Lakhdar, Z. B. & Ruiz-López, M. F. The structure of chlorophyll a–water complexes: insights from quantum chemistry calculations. *Chem. Commun.* https://doi.org/10.1039/b716800d (2008).

44. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

45. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

46. Caspy, I., Borovikova-Sheinker, A., Klaiman, D., Shkolnisky, Y. & Nelson, N. The structure of a triple complex of plant photosystem I with ferredoxin and plastocyanin. *Nat. Plants* **6**, 1300–1305 (2020).

47. Caspy, I. et al. Structure of plant photosystem I–plastocyanin complex reveals strong hydrophobic interactions. *Biochem. J.* **478**, 2371–2384 (2021).

48. Hippler, M., Drepper, F., Haehnel, W. & Rochaix, J.-D. The N-terminal domain of psaf: precise recognition site for binding and fast electron transfer from cytochrome c 6 and plastocyanin to photosystem I of *Chlamydomonas reinhardtii*. *Proc. Natl Acad. Sci. USA* **95**, 7339–7344 (1998).

49. Li, X. et al. A genome-wide algal mutant library and functional screen identifies genes required for eukaryotic photosynthesis. *Nat. Genet.* **51**, 627–635 (2019).

50. Boynton, J. E. et al. Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. *Science* **240**, 1534–1538 (1988).

51. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).

52. Hippler, M., Drepper, F., Farah, J. & Rochaix, J.-D. Fast electron transfer from cytochrome c6 and plastocyanin to photosystem I of *Chlamydomonas reinhardtii* requires psaf. *Biochemistry* **36**, 6343–6349 (1997).

53. Hippler, M., Klein, J., Fink, A., Allinger, T. & Hoerth, P. Towards functional proteomics of membrane protein complexes: analysis of thylakoid membranes from *Chlamydomonas reinhardtii*. *Plant J.* **28**, 595–606 (2001).

54. Naumann, B., Stauber, E. J., Busch, A., Sommer, F. & Hippler, M. N-terminal processing of LHCA3 is a key step in remodeling of the photosystem I–light-harvesting complex under iron deficiency in *Chlamydomonas reinhardtii*. *J. Biol. Chem.* **280**, 20431–20441 (2005).

55. Ozawa, S.-ichiro, Onishi, T. & Takahashi, Y. Identification and characterization of an assembly intermediate subcomplex of photosystem I in the green alga *Chlamydomonas reinhardtii*. *J. Biol. Chem.* **285**, 20072–20079 (2010).

56. Michel, H., Tellenbach, M. & Boschetti, A. A chlorophyll B-less mutant of *Chlamydomonas reinhardii* lacking in the light-harvesting chlorophyll ab-protein complex but not in its apoproteins. *Biochim. Biophys. Acta* **725**, 417–424 (1983).

57. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).

58. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).

59. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

60. Tyanova, S. et al. The Perseus Computational Platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).

61. Busch, A., Nield, J. & Hippler, M. The composition and structure of photosystem I-associated antenna from *Cyanidioschyzon merolae*. *Plant J.* **62**, 886–897 (2010).

62. Shevchenko, A., Tomas, H., Havli, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–2860 (2006).

63. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. CryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

64. Zheng, S. Q. et al. Motioncor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

65. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

66. Pisoni, E. Code on Zenodo. *Zenodo* https://doi.org/10.5281/zenodo.3576630 (2020).

67. Scheres, S. H. W. Amyloid structure determination in RELION-3.1. *Acta Crystallogr D Struct Biol.* **76**, 94–101 (2020).

68. Zivanov, J., Nakane, T. & Scheres, S. H. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in *RELION*-3.1. *IUCrJ* **7**, 253–267 (2020).

69. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).

70. Zivanov, J., Nakane, T. & Scheres, S. H. A bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).

71. Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. W. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7**, e36861 (2018).

72. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

73. Waterhouse, A. et al. Swiss-model: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).

74. Afonine, P. V. et al. Real-space refinement in phenix for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol.* **74**, 531–544 (2018).

75. Williams, C. J. et al. Molprobity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2017).

76. Goddard, T. D. et al. UCSF Chimerax: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2017).

77. Bianca, N. et al. Comparative quantitative proteomics to investigate the remodeling of bioenergetic pathways under iron deficiency in *Chlamydomonas reinhardtii*. *PROTEOMICS* **7**, 3964–3979 (2007).

78. Abraham, S. & Golay, M. J. E. Smoothing and Differentiation of Data by simplified least squares procedures. *Analytical Chemistry* **36**, 1627–1639 (1964).

Acknowledgements

Author contributions

M.H. and A.A. designed the project. L.M., A.P.-B. and A.N. prepared the sample for cryo-EM and performed initial screening. A.N. collected and processed the cryo-EM data and built the model. L.M., S.K., M.S. and A.V.-M. performed biochemical analysis. V.T. performed computational analysis. Y.T. provided antibodies. T.T.H.H. produced mutant strains. A.N. and A.A. analysed the structure and wrote the manuscript with contributions from L.M. and M.H. All authors contributed to the analysis and the final version of the manuscript.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41477-022-01253-4.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41477-022-01253-4.

Correspondence and requests for materials should be addressed to Michael Hippler or Alexey Amunts.

Peer review information *Nature Plants* thanks Matthew Johnson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

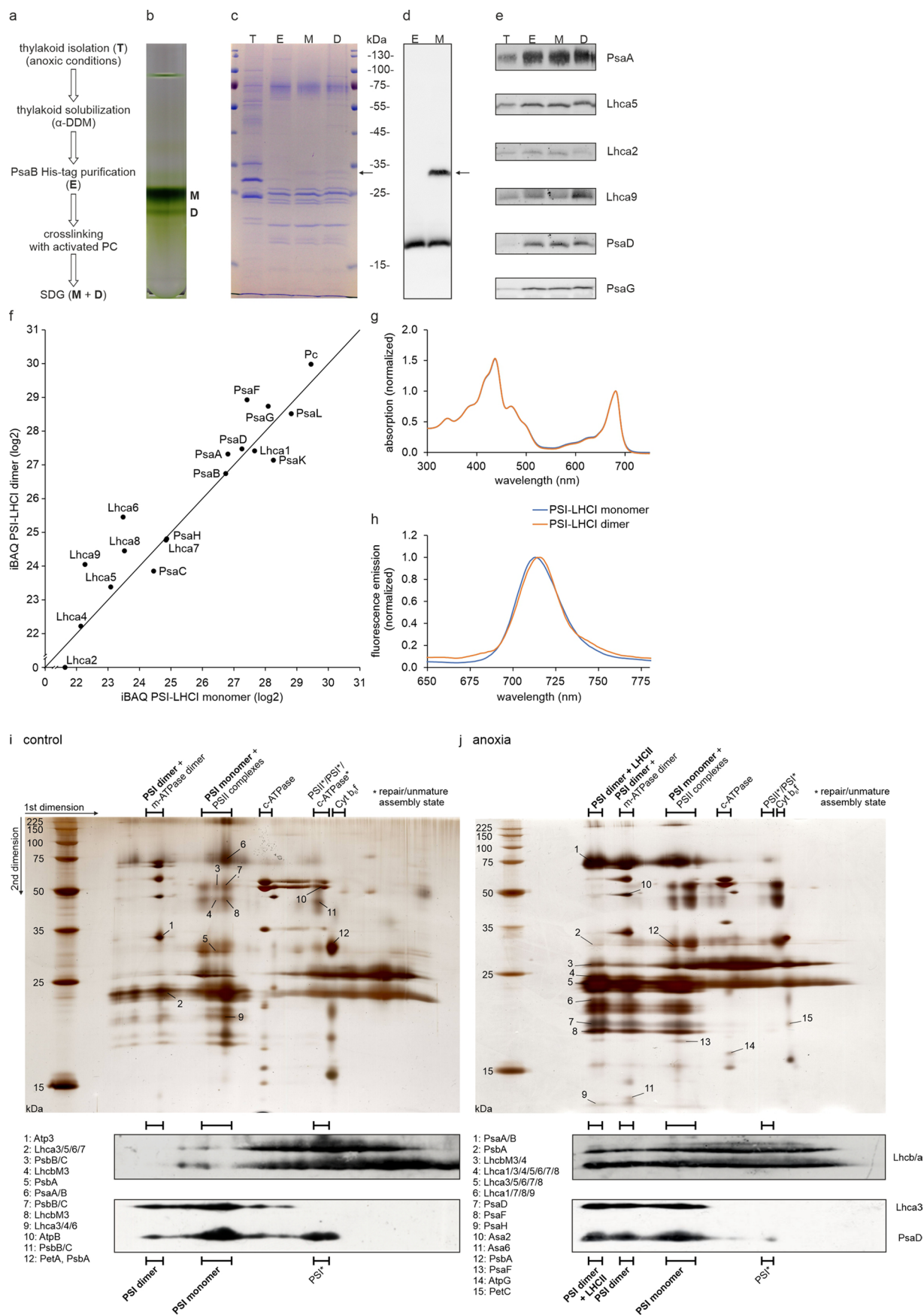**Reprints and permissions information** is available at www.nature.com/reprints.

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

Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Sample preparation and characterization. a, The experimental workflow. **b**, Sucrose density gradient affinity purified PSI (~ 60 μg chl). Monomer (M) and dimer (D) fractions are indicated. **c**, SDS-PAGE analysis (1 μg chl). T: thylakoids; E: PsaB-His elution, M: PSI monomer, D: PSI dimer. The arrow indicates a putative PsaF-Pc crosslinked product. **d**, SDS-PAGE and Western Blot against PsaF confirming the cross-linking with Pc. **e,** SDS-PAGE and Western Blot against PsaA, Lhca5, Lhca2, Lhca9, PsaD and PsaG confirming the presence of small PSI subunits as well as LHCI subunits from both sides of the PSI core. **f**, Quantitative mass spectrometry analysis of the PSI fractions from (b). iBAQ values are normalized to the PSI core subunit PsaB and values below 21 are excluded as they represent already low intensity values, which might not be reliable. PSI-LHCI subunits PsaE and Lhca3 were not detected at quantifiable intensities in this experiment. Detection of these small PSI and Lhca subunits via mass spectrometry depends on the presence of only a few proteotypic tryptic peptides, which could be missed, as measurements were done in a data dependent fashion (12 MS^2 per MS^1). **g**, Room temperature absorption spectra of PSI monomer and dimer fractions. The spectra are normalized to the red region. **h**, Low temperature (77 K) fluorescence emission spectra of PSI monomer and

dimer fractions upon excitation of chlorophyll a at 436 nm. The spectra are normalized to the maximum of the emission peaks and smoothed according to Savitzky-Golay[78] using the Jasco spectra analysis program. Data in panel b to h is representative of n = 3 independently performed experiments with similar results. Notably, the presence of PsaH in the dimer sample, but its absence in the cryo-EM structure, indicates that PsaH is only loosely attached to the PSI dimer and readily lost during the cryo-EM analyses. **i**, 2D-PAGE of β-DDM solubilized *C. reinhardtii* wild type thylakoids, control conditions. **j,** low light and anoxic conditions. In the first dimension, multi-protein complexes were separated in their native form by blue native PAGE. In the second dimension, subunits of these protein complexes were separated by SDS-PAGE. "PSII complexes" refers to PSII complexes of different molecular weight due to varying extents of LHCII association. Labelled silver-stained spots were subjected to LC-MS/MS analysis. For convenience, the figure includes a condensed list of representative subunits. For a list of proteins identified in the respective spots, refer to Supplementary Information Dataset 1. Western Blots of Lhcb/a, Lhca3 and PsaD indicate the distribution of LHCII, LHCI and the PSI core. Data in panels i and j is representative of n = 3 independently performed experiments with similar results.
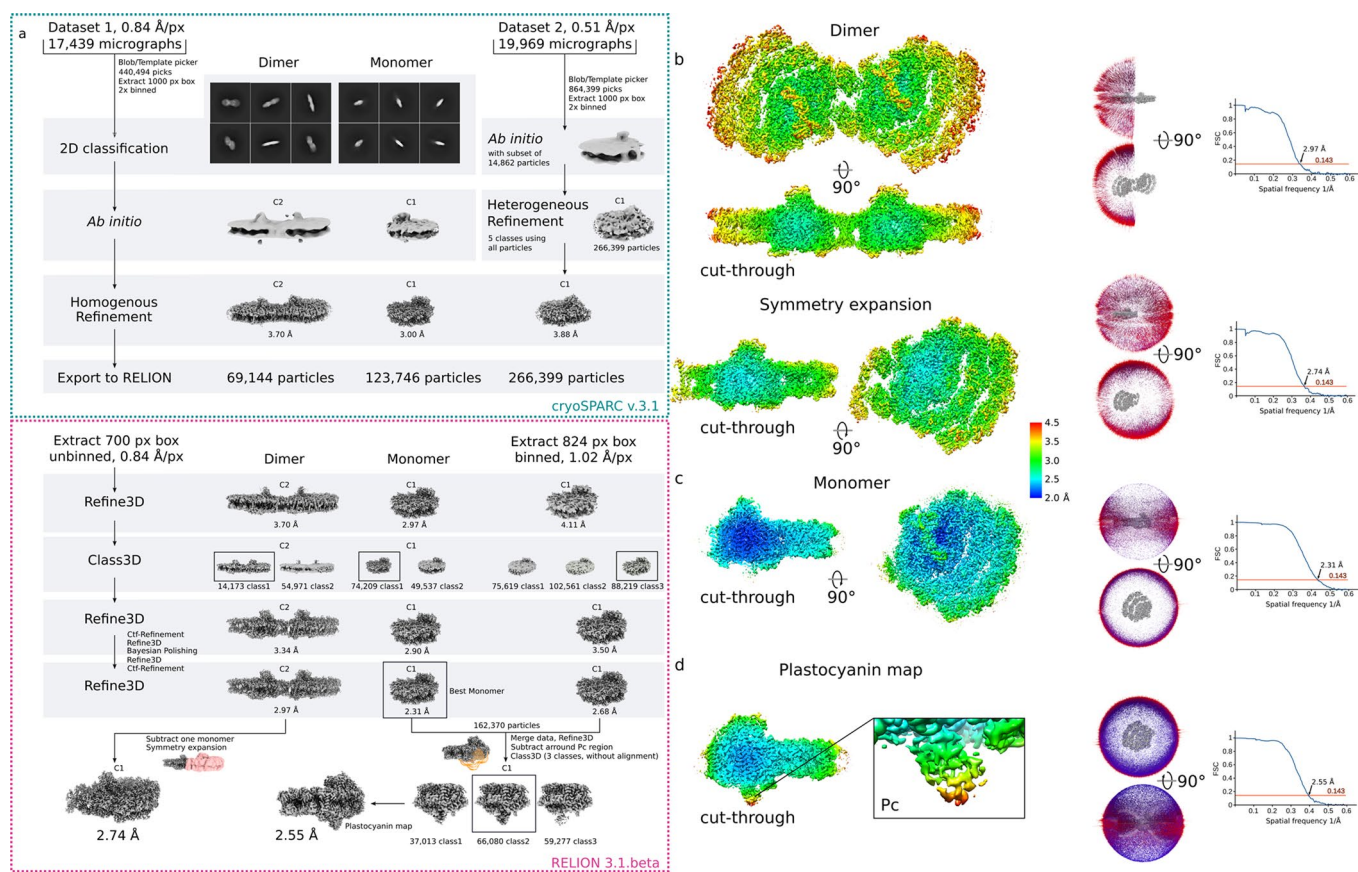
Extended Data Fig. 2 | Cryo-EM processing workflow and local resolution. a, Flowchart of the data processing for the PSI dimer and monomer. The processing steps performed in cryoSPARC and RELION are indicated. To increase the numbers of Pc bound particles Dataset 2 was added for focused classification with signal subtraction. **b**, The map of the dimer and symmetry expanded dimer colored by local resolution in two orientations (left). Angular distribution of the dimer and symmetry expanded dimer and the FSC-curves are shown. **c**, The map of the monomer, angular distribution and FSC-curve. **d**, The map with the best density for plastocyanin, angular distribution and FSC-curve.
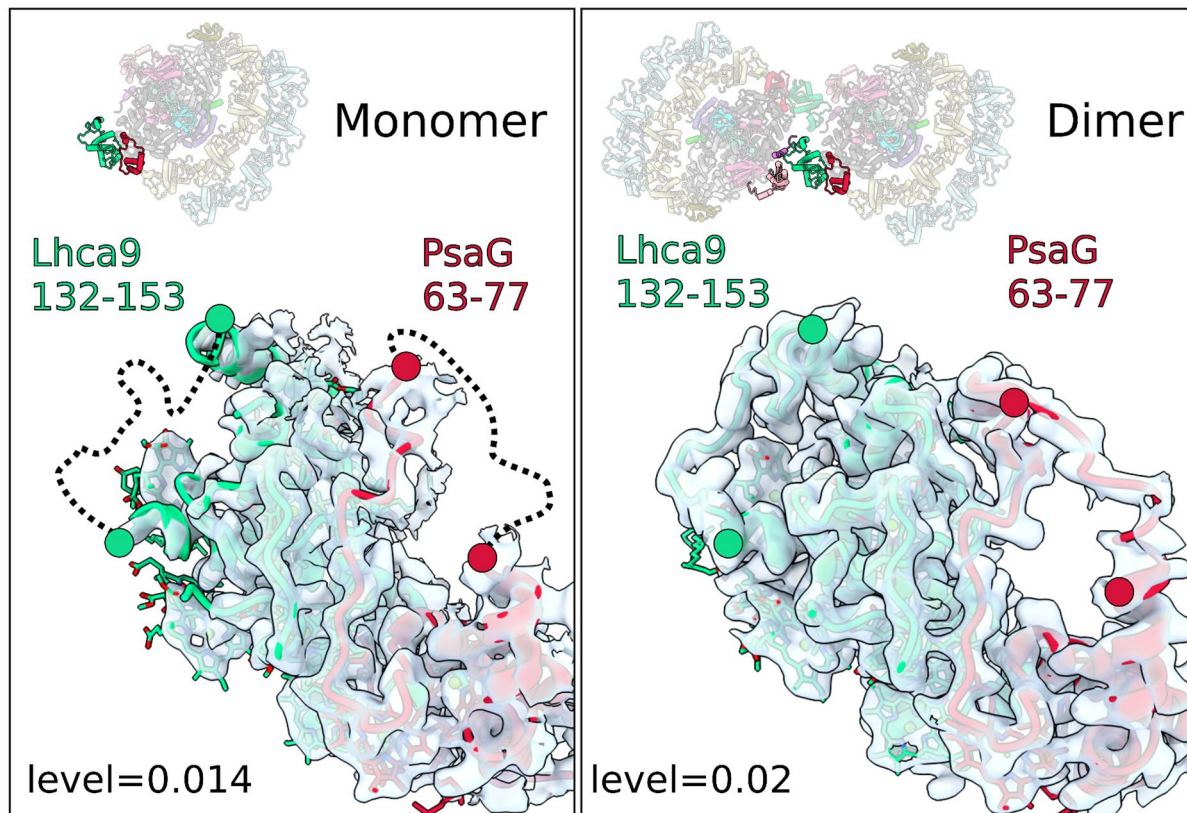
Extended Data Fig. 3 | Stabilization of loops in Lhca9 and PsaG in PSI dimer. Left, in the monomer, the Lhca9 loop 132–153 and PsaG loop 63–77 are disordered (dashed lines). Right, in the dimer, both loop regions are ordered due to a stabilization by the neighbouring monomer.
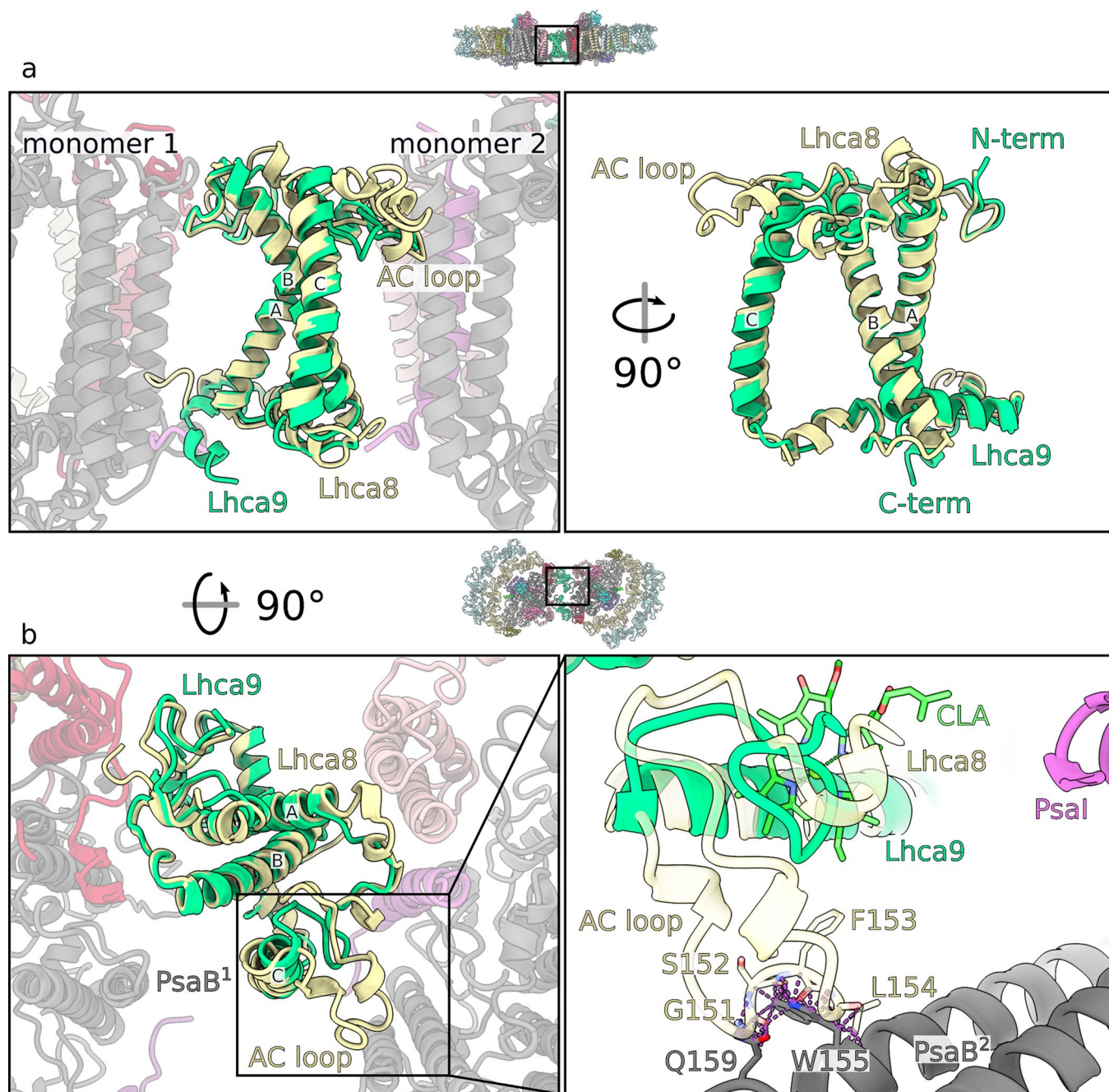
Extended Data Fig. 4 | Lhca9 short AC-loop contributes to the dimerization. a, Superposition of Lhca8 and Lhca9, highlighting the difference in the AC-loop. **b,** The 90°-rotated view illustrates a clash between the longer AC-loop version and PsaB of monomer 2. Clashes between atoms with a VdW overlap > 0.6 Å are highlighted by the dashed lines.
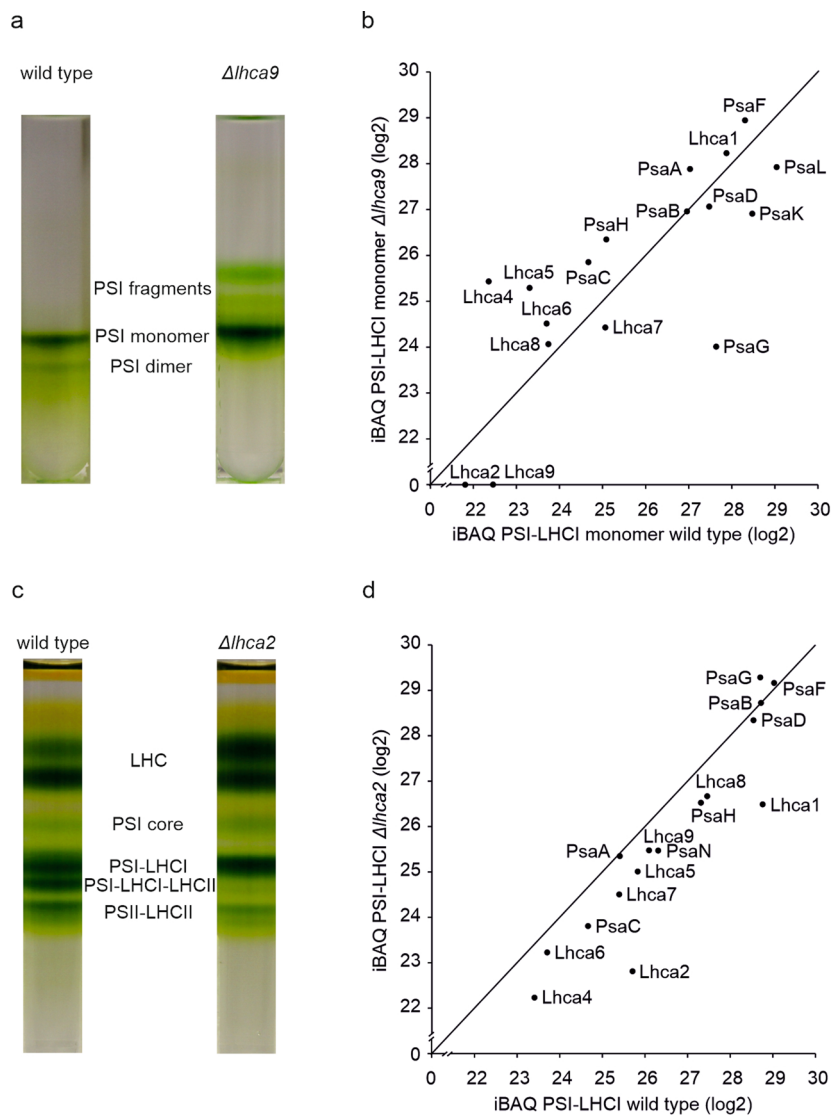
Extended Data Fig. 5 | Analysis of *Δlhca9* and *Δlhca2*. a, Sucrose density gradient of affinity purified PSI from *lhca9* insertional mutant (~60 μg chl) compared to wild type. **b**, Quantitative mass spectrometry analysis of the PSI-LHC monomer fraction. iBAQ values are normalized to the PSI core subunit PsaB and values below 21 are excluded as they represent already low intensity values, which might not be reliable. PSI-LHCI subunits PsaE and Lhca3 were not detected at quantifiable intensities in this experiment. Detection of these small PSI and Lhca subunits via mass spectrometry depends on the presence of only a few proteotypic tryptic peptides, which could be missed, as measurements were done in a data dependent fashion (12 MS[2] per MS[1]). **c**, Sucrose density gradients of α-DDM solubilized thylakoids (~250 μg chl) isolated from wild type and *lhca2* insertional mutant. **d**, Quantitative mass spectrometry analysis of the PSI-LHCI fraction in *lhca2* insertional mutant. iBAQ values are normalized to the PSI core subunit PsaB and values below 21 are excluded as they represent already low intensity values, which might not be reliable. PSI-LHCI subunits PsaE, PsaK and Lhca3 were not detected at quantifiable intensities in this experiment. Detection of these small PSI and Lhca subunits via mass spectrometry depends on the presence of only a few proteotypic tryptic peptides, which could be missed, as measurements were done in a data dependent fashion (12 MS[2] per MS[1]). Gradients in panel a and c are representative of n = 3 independently performed respective experiments with similar results.
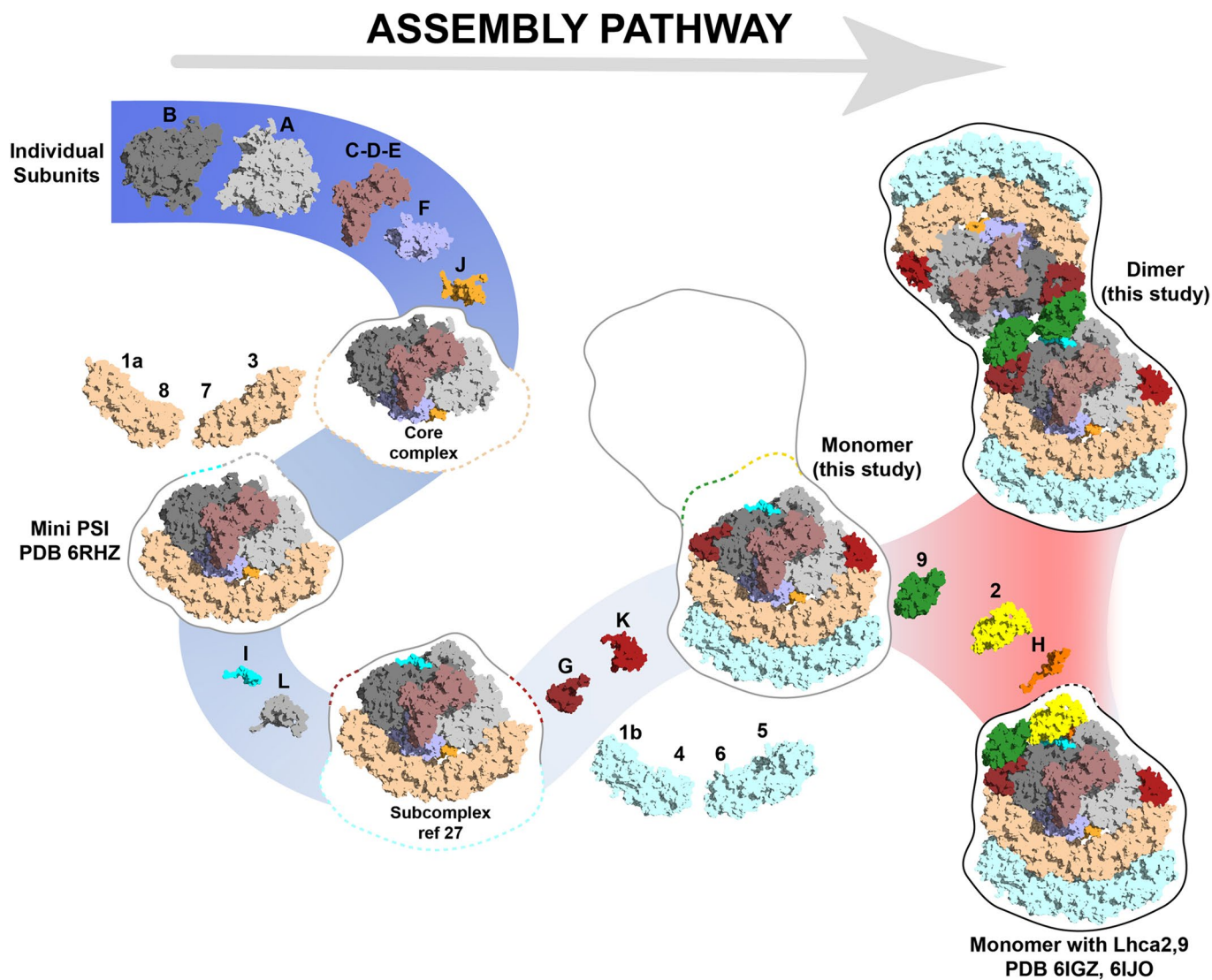
Extended Data Fig. 6 | Assembly pathway of PSI towards dimer. a, Schematic view of known biochemically and structurally defined PSI states and assembling protein subunits. Subunit composition for each state is outlined with the shape of a following state and corresponding color. The divaricating of the assembly pathway takes place at the last stage (red background). The path towards PSI monomer or dimer is dependent on presence/absence of PsaH and Lhca2.
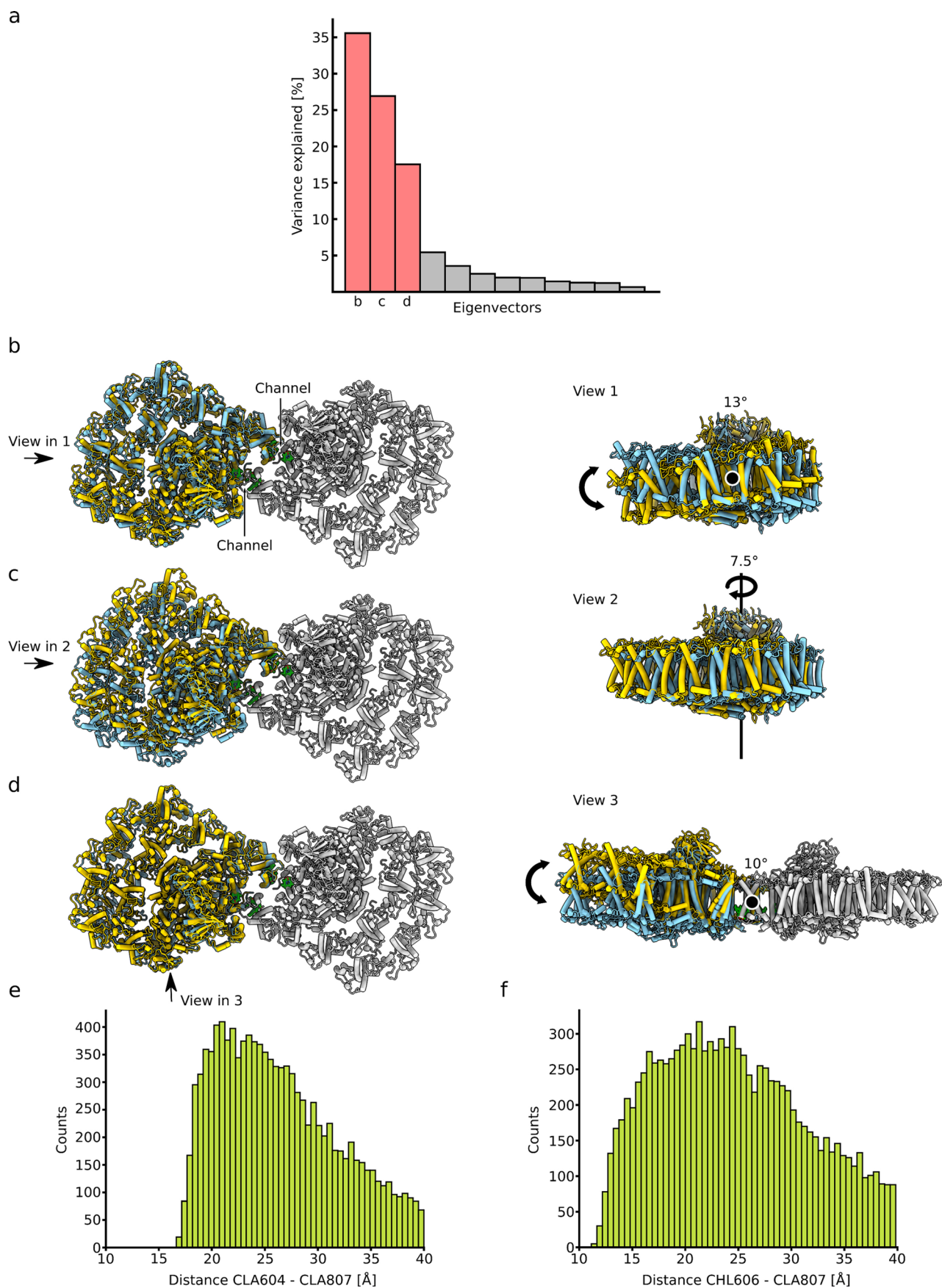
Extended Data Fig. 7 | Multi-body refinement analysis. a, Eigenvectors that explain the variability of the data. The three major eigenvectors account for ~78% of the motion in the PSI dimer. **b-d**, Stromal and side view of the model, showing the maximal motion along the three vectors. **e, f**, The distance between the chlorophylls is plotted based on the relative motion from the multi-body analysis. The y-axis shows the counts of particles that exhibit a certain distance.
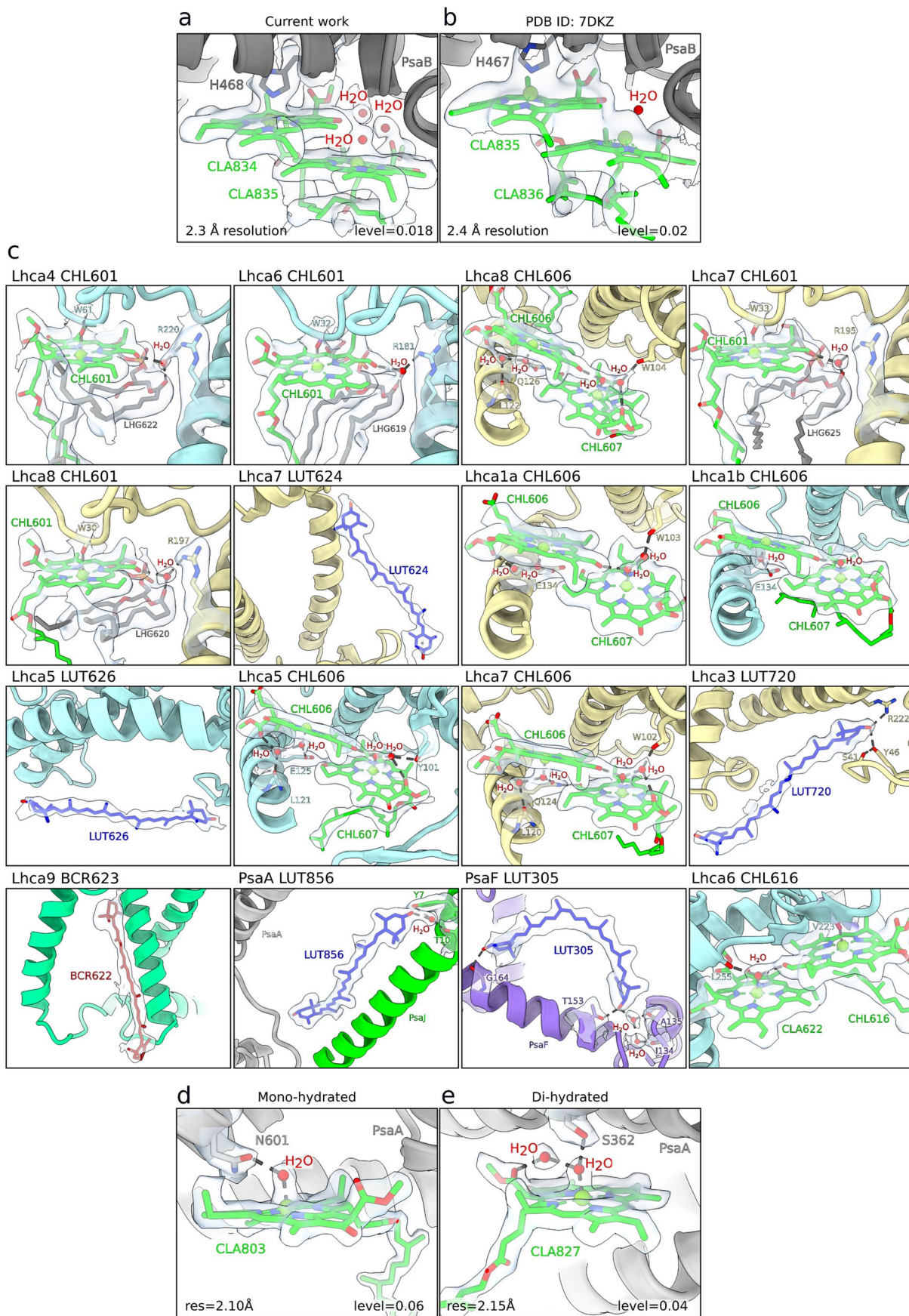
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Identification of new elements and map quality. a, The current density around CLA834-835 is shown in iso-surface representation at a local resolution of 2.3 Å. Three water molecules involved in chlorophyll coordination are identified. Updated libraries are used for correct chlorophyll modeling. **b**, The same region from X-ray study at 2.4 Å resolution[20] lacks the density for waters, whereas the modeled water molecule has no density, and Mg atoms are not in the correct plane. **c**, Close up views of newly identified elements, coordination and chemical environment with the density map. **d**, An example of a mono-hydrated chlorophyll with the corresponding density. **e**, An example of a di-hydrated chlorophyll with the corresponding density. Local resolution and map levels are indicated.
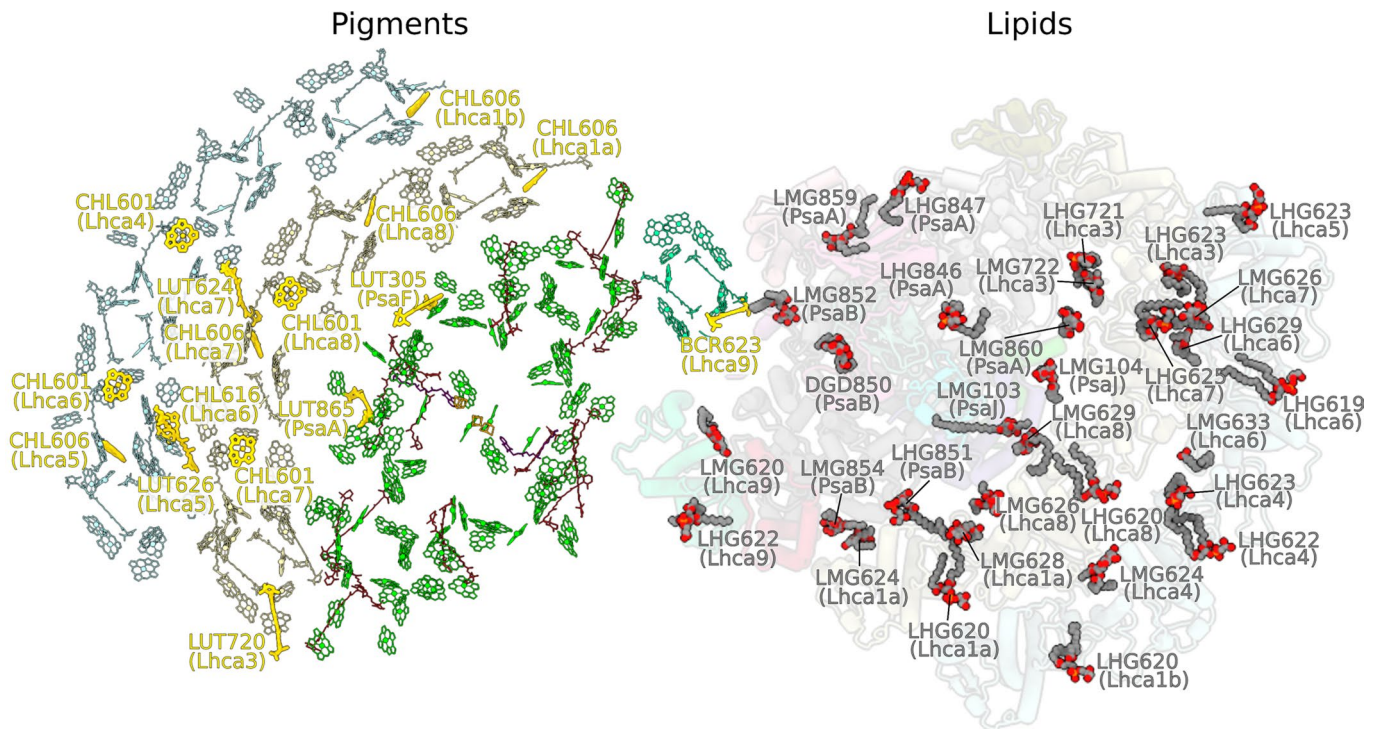
Extended Data Fig. 9 | Overview of the pigments and lipids. Pigments (left) and lipids (right) are shown from stroma. The newly identified pigments are highlighted in bold gold.

Extended Data Table 1 | Cryo-EM statistics. Data collection, processing, model refinement and validation statistics

	Dimer (EMDB-14871) (PDB 7ZQD)	Symmetry expanded dimer (EMDB-14867) (PDB 7ZQ9)	Monomer (EMDB-14870) (PDB 7ZQC)	Plastocyanin (EMDB-14872) (PDB 7ZQE)
Data collection and processing				
Magnification	105,000	105,000	105,000	165,000 / 105,00
Voltage (kV)	300	300	300	300
Electron exposure (e–/Å^2)	45.8	45.8	45.8	48 / 45.8
Defocus range (μm)	-0.7-2.5	-0.7-2.5	-0.7-2.5	-0.3, -1.1 / -0.7-2.5
Pixel size (Å)	0.84	0.84	0.84	0.51 / 0.84
Initial images (no.)	17,439	17,439	17,439	19,969 / 17,439
Symmetry imposed	C2	C1	C1	C1
Final particle images (no.)	14,173	28,346	74,209	66,080
Map resolution (Å) FSC threshold 0.143	2.97	2.74	2.31	2.55
Refinement				
Initial model used (PDB code)	6JO5	6JO5	6JO5	SWISS model
Map sharpening B factor (Å^2)	-28.80	-21.00	-23.05	-21.05
Model composition				
Non-hydrogen atoms	100,830	56,460	50,499	729
Protein residues	8,418	4,712	4,164	98
Ligands	708	398	352	0
Waters	160	88	621	0
B factors (Å^2) (min/max/mean)				
Protein	22.5/125.1/65.5	48.7/154.0/80.3	13.9/117.2/40.0	66.7/83.8/73.5
Ligand	25.5/118.9/66.9	51.0/148.5/79.45	15.9/109.5/39.45	
Waters	25.4/115.3/64.0	51.1/95.5/72.0	13.4/57.5/31.72	
R.m.s. deviations				
Bond lengths (Å)	0.009	0.009	0.009	0.002
Bond angles (°)	1.065	1.069	1.085	0.494
Validation				
MolProbity score	1.16	1.16	1.12	1.20
Clashscore	3.70	3.69	3.29	4.21
Poor rotamers (%)	0.39	0.35	0.24	0.00
Ramachandran plot				
Favored (%)	98.40	98.34	97.98	98.96
Allowed (%)	1.57	1.61	1.99	1.04
Disallowed (%)	0.02	0.04	0.02	0.00

nature portfolio

Corresponding author(s):   ALEXEY AMUNTS

Last updated by author(s):   Aug 26, 2022

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	The datasets were collected EPU 1.9 software on FEI Titan Krios (FEI/Thermofischer) transmission electron microscope operated at 300 keV with a slit width of 20 eV on a GIF quantum energy filter (Gatan). A K2 Summit detector (Gatan) was used at a pixel size of 0.84 Å (magnification of 105,000x) with an exposure rate of 15.27 electrons/pixel/second fractionated over 45 frames. A defocus range of -0.7 to -2.5 μm was used.
Data analysis	Movie frames were aligned and averaged by global and local motion corrections by MotionCor2. Contrast transfer function (CTF) parameters were estimated by CTFFIND4. Particles were picked and initially 2D classified by RELION 3.1. 2D classification and 3D heterogeneous refinement steps were performed in cryoSPARC v.3.1. The models were manually built with Coot 0.914 and stereochemical refinement was performed using phenix.real_space_refine in the PHENIX 1.17.1 suite. The final figures were generated using UCSF Chimera 1.14, UCSF Chimera X 0.91, Pymol 2.0. Mass-spectrometry data was processed in MaxQuant 1.6.14, Perseus 1.6.15.0, Proteome Discoverer 2.4. Spectra were normalized to the maximum of the emission peaks using the Jasco spectra analysis program Spectra Manager II.

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

Data

Policy information about availability of data

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

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

The atomic coordinates were deposited in the RCSB Protein Data Bank (PDB). The local resolution filtered maps, half maps, masks and FSC-curves have been deposited in the Electron Microscopy Data Bank (EMDB).
Mass spectrometry datasets: Project Name: Intensity-based absolute quantification (iBAQ) of components of photosystem I monomers and dimers. Project accession: PXD026990. Project DOI: 10.6019/PXD026990. Project Name: Identification of photosystem I components from Chlamydomonas reinhardtii grown under oxic and anoxic conditions. Project accession: PXD027067. Project name "Label-free quantification of photosystem I monomers and dimers from Chlamydomonas reinhardtii." Project accession: PXD028024. Project DOI: 10.6019/PXD028024.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data where this information has been collected, and consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	A total of 17,439 movies were recorded and analyzed. No statistical analyses has been performed. The number of cryo-EM particles in the single dataset collected was the number of particles available. No predetermined sample size was used for other experiments.
Data exclusions	Particles that were not photosystem I were excluded in the analysis, since they cannot contribute to reconstruction.
Replication	Similar cryo-EM structures were successfully obtained from preliminary datasets.
Randomization	Particle images were randomly assigned into half-sets to obtain gold-standard resolution estimates as described in the text.
Blinding	N/A to cryo-EM study; raw micrographs or particle images are not categorical data. Particles are randomly assigned into half-sets for image processing; hence no blinding is applicable.

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 and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Antibodies

Antibodies used	primary rabbit anti-PsaF, -Pc, -PsaA, -Lhca5, -Lhca2, -Lhca9, -Lhca3, -PsaD, -PsaG, -Lhcb/a, PsbA and LhcSR3 as referenced in the text secondary horseradish peroxidase (HRP) conjugated goat anti-rabbit IgG (H + L) (Bio-Rad)
Validation	validation provided by manufacturer