Chloroplast photosystem I dimer and high resolution model of the complex with plastocyanin

Andreas Naschberger[1,2,†], Laura Mosebach[3,†], Victor Tobiasson[1,2], Sebastian Kuhlgert[3], Martin Scholz[3], Annemarie Perez-Boerema[1], Yuichiro Takahashi[4,5], Michael Hippler[3,6*], Alexey Amunts[1,2,*]

[1]Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, 17165 Solna, Sweden
[2]Department of Medical Biochemistry and Biophysics, Karolinska Institute, 17177 Stockholm, Sweden
[3]Institute of Plant Biology and Biotechnology, University of Münster, 48143 Münster, Germany
[4]Research Institute for Interdisciplinary Science, Okayama University, Okayama 700-8530, Japan
[5]Japan Science and Technology Agency-CREST, 4-1-8 Kawaguchi, Saitama 332-0012, Japan
[6]Institute of Plant Science and Resources, Okayama University, Kurashiki, Japan

[*]Correspondence to: mhippler@uni-muenster.de; amunts@scilifelab.se

[†]These authors contributed equally to this work.

Abstract

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. Our cryo-EM structure identifies the light-harvesting protein Lhca9 as the key element for the dimerization. Furthermore, the absence of Lhca2 and PsaH, gives rise to a head-to-head relative orientation of the PSI-LHCI monomers, in a way that is essentially different from the oligomer formation in cyanobacteria. The interface between the monomers partially overlaps with the surface area that would bind one of the LHCII complexes in state transitions. We also define the most accurate available PSI-LHCI model at 2.3 Å resolution, including a flexibly bound electron donor plastocyanin, and assign correct identities and orientations of all the pigments, as well as 486 water molecules that affect energy transfer pathways.

Main

A chloroplast PSI of green algae consists of the 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 a fluctuating illumination and anoxia, the algal PSI additionally associates with two LHCII trimers [5,6]. Structural studies have shown that the oligomeric state of a chloroplast PSI is a monomer, due to a presence of the subunit PsaH, whereas in cyanobacteria also structures of dimers [7-9] and trimers [10] were reported. Cyanobacterial PSI oligomerizes via direct contacts between subunits PsaI and PsaL, however such an association has been ruled out for a chloroplast PSI due to structural constraints and apparent rigidity [11,12]. Recent structural studies of PSI from a chloroplast of a salt-tolerant alga suggest that its functional core may vary more than previously believed, thus suggesting a potential architectural plasticity in response to environment [13]. 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 [14]. The membrane architecture in dark- and light-adapted membranes contains ordered rows of closely packed PSI dimers, which are more abundant in the dark state [14]. Similarly, closely associated PSI-LHCI complexes were detected in plants by negative stain electron microscopy [15]. 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 chloroplast PSI-LHCI dimers, and information on their structures is lacking. In the absence of any structural data, no evidence is available on composition, elements mediating dimerization, and how the arrangement would differ from the cyanobacterial counterparts.

To study a chloroplast PSI dimer at high resolution, we grew genetically modified *C. reinhardtii* cells containing a His-tag at the N-terminus of PsaB in low light and under anoxic conditions (see Methods). The thylakoid membranes were then solubilised 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. 1). Two PSI fractions were detected on the sucrose gradient. 2D polyacrylamide gel electrophoresis (native/reducing 2D-PAGE) of isolated thylakoids indicated the presence of PSI dimers (Extended Data Fig. 2 and Supplementary Table 1). The heavier green band on the gradient was subjected to single-particle cryo-EM analysis (Supplementary Table 2). 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. 3). Upon symmetry expansion, the resolution was further improved to 2.74 Å (Extended Data Fig. 3). The remaining 74,209 particles containing the monomer were refined to 2.31 Å resolution, representing a considerable improvement on the previously reported 2.9-3.5 Å resolution structures [1-3], as well as X-ray crystal structures of a plant PSI [16,17]. A density corresponding to the bound electron donor Pc was found at the luminal side of both PSI forms, dimer and monomer.

78 To derive a structure of the chloroplast PSI dimer, we first built an accurate model of one
79 monomer using the 2.74 Å resolution map, and then fitted it into the cryo-EM density of the C2
80 refined dimer. Compared to the monomer, all but two core subunits (PsaH, PsaO) and one light-
81 harvesting protein (Lhca2) are found in the dimer (Fig. 1). The structure contains 40 protein
82 subunits, 398 chlorophylls *a*, 60 chlorophylls *b*, 56 beta-carotenes, 54 luteins, 2 violaxanthins, 2
83 neoxanthins, 4 phylloquinones, 6 iron-sulphur clusters, and 32 lipids (Fig. 1). In addition, two
84 unaccounted densities corresponding to two loops in the stromal side of Lhca9 and PsaG could be
85 interpreted in the dimer, due to the stabilization by the adjacent monomer (Extended Data Fig. 4).
86 The first better-defined density is Lhca9 loop region 132-153 which is stabilised due to a direct
87 interaction with PsaL of the second monomer within the dimer. As a result, the area is closely
88 packed with PsaG, and therefore also the PsaG loop region 63-77 is better resolved in the dimer
89 (Extended Data Fig. 4). Finally, cofactors of Lhca9 could be modelled at the interface between
90 the monomers. No density for the His-tag on PsaB could be detected.

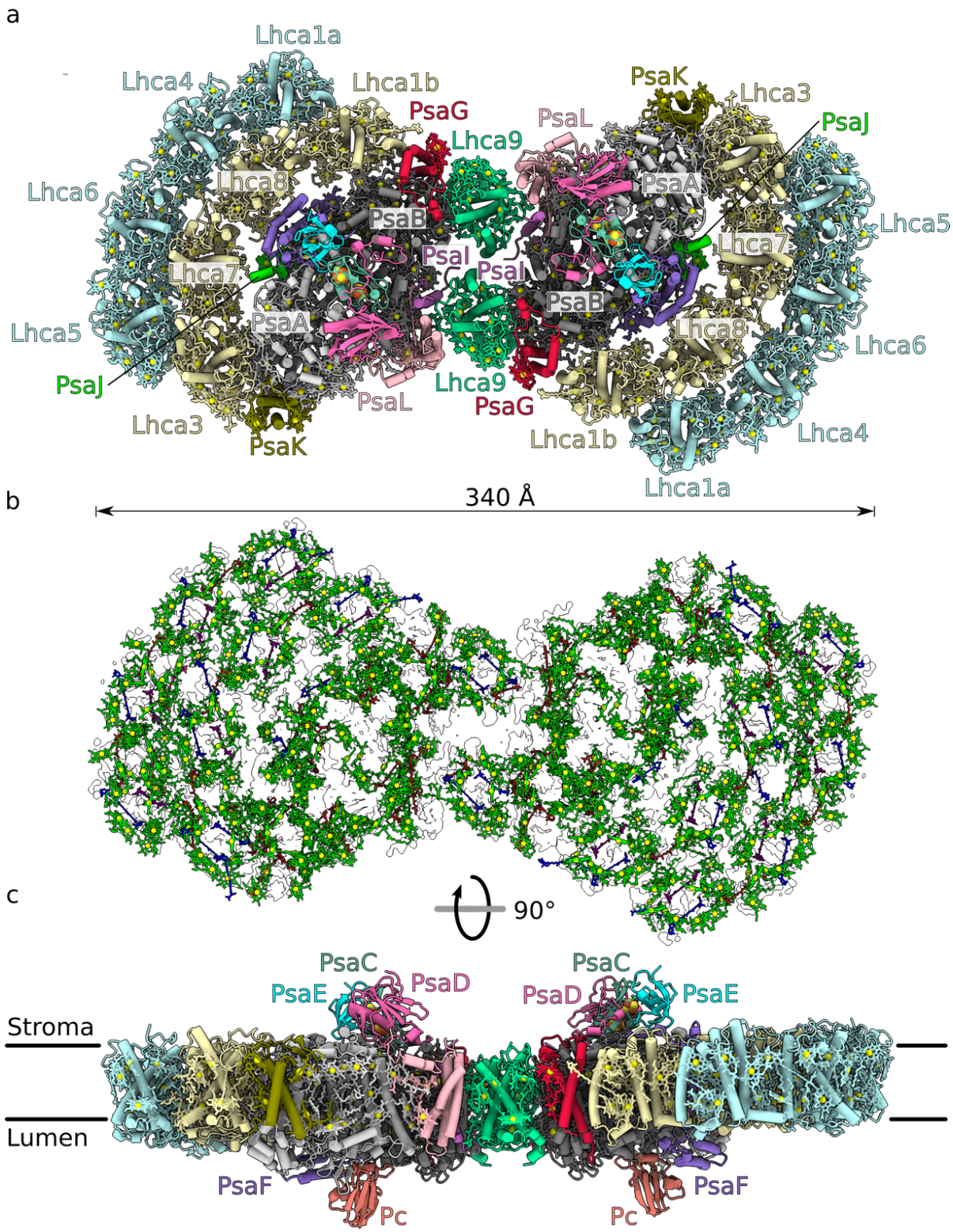


Fig. 1: Overall structure of the PSI dimer. **a**, View of individual proteins from stroma. **b**, Arrangement of the pigments in the outline of the map: chlorophylls green (Mg yellow), luteins blue, beta-carotenes red, violaxanthin purple, and neoxanthin pink **c**, Overall view along the membrane.

The structural basis for the algal chloroplast PSI dimerization is fundamentally different from cyanobacteria (Fig. 2a). In cyanobacteria, PSI dimerises via the stromal region of PsaL [7-9] and trimerises via the lumenal C-terminus of PsaL, assisted by PsaI [10]. In our structure of the chloroplast PSI-LHCI dimer, neither PsaL nor PsaI interact with each other between the neighbouring units. Instead, PsaH that preserved a monomer is removed, and Lhca9 and its associated cofactors act as a symmetrical linker between the monomers, highlighting the importance of the light-harvesting antenna proteins for regulation of the macro-organisation. Lhca9 is distinct among the light-harvesting proteins due to a truncated loop between helices A and C, and lack of the associated chlorophyll [6]. Based on this difference, we rationalised how Lhca9 allows for dimerization, as a longer AC-loop would clash with the neighbouring PsaB (Extended Data Fig. 5), although conformational changes cannot be excluded.

Consequently, the two Lhca9 copies tether the PSI monomers in a head-to-head fashion, resulting in a 340-Å long structure (Fig. 1 and Fig. 2b). They form interactions of four types covering the 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 cofactors in the membrane that include a newly modelled, beta-carotene 623, and five chlorophylls (604, 610, 611, 612, 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 beta-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 only be detected by high-resolution structural studies. The involvement of lipids in PSI oligomerization is consistent with the formation of supercomplexes in other bioenergetic membranes [18-21].

To solidify the structural observations, we engineered a *Δlhca9* 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. 6). Next, we analysed sucrose density gradients of solubilized thylakoids from *Δlhca2* and *Δlhca9* mutants without the His-tag (Extended Data Fig. 7). Compared to the wild type, in *Δlhca2*, the abundance of the heavier band was reduced, and Lhca9 is present, supporting its independent association with PSI. However, in *Δlhca9*, no PSI dimer fraction was observed, and Lhca2 presence is reduced. This data is consistent with the proposed structural role of Lhca9 in the PSI dimer formation.

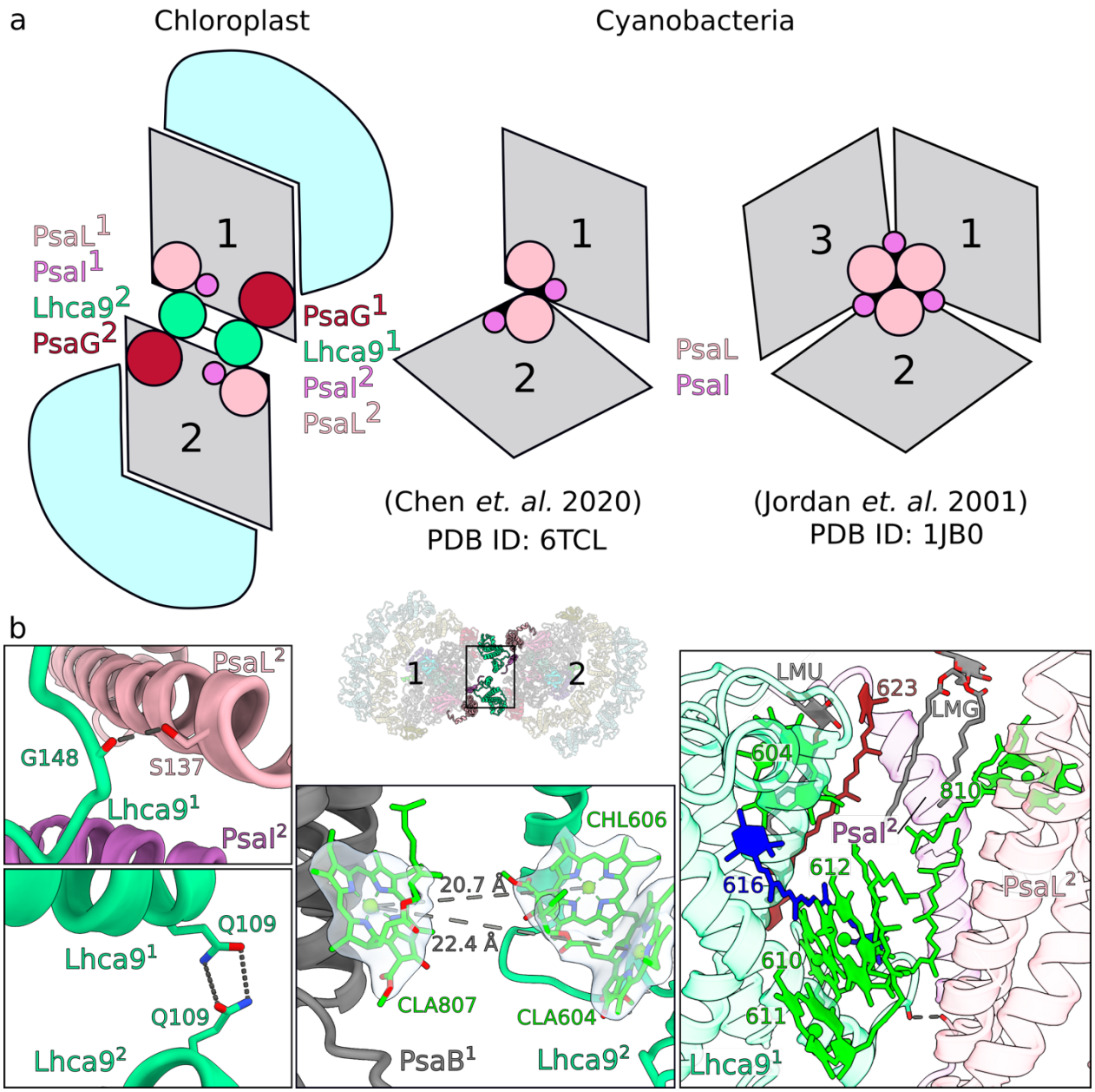


Fig. 2: Dimerization of PSI. a, Schematic representation: Chloroplast PSI-LHCI dimer associated via Lhca9, cyanobacterial PSI dimer (PDB ID: 6TCL) and trimer (PDB ID: 1JB0) associated via PsaI/L. The PSI core is grey, LHCI light-blue. **b**, The dimer interface is formed by: hydrogen bonds between PsaL and Lhca9, and between two Lhca9 copies (left); potential energy transfer paths between the two monomers (center); pigments and lipids (right).

The specific interactions between the monomers are enabled due to unoccupied positions of Lhca2 and PsaH (Fig. 3a). Compared to the state transition complex [5,6], Lhca9 from the neighbouring monomer is positioned in the membrane, where Lhca2 resides in PSI-LHCI-LHCII (Fig. 3b,c), and their three transmembrane helices would overlap with each other (Fig. 3c). The presence of the PsaH transmembrane helix is not compatible with the Lhca9^2-associated cofactors CLA9, LMG852, BCR9 that extend from the neighbouring monomer in the dimer (Fig. 3c). In

addition, the superposition shows that there would be a clash between PsaG and Lhca1 of the inner belt with LHCII-2 (Fig. 3c). Since Lhca2 and PsaH are required for the lateral binding of LHCII to the core in state transitions, the structure of the algal PSI dimer would not facilitate LHCII binding at this position. However, our 2D-PAGE indicated a comigration of LHCII polypeptides with the dimer fraction, and therefore a structural adaptation cannot be excluded. The antagonistic relationship of Lhca9[2] and Lhca2 might reflect a regulation of PSI dimerization that is associated with the membrane shape adaptation in response to environment [14].

Compared to algae, plant PSI binds one LHCII trimer with a stronger association to the core, the docking site of which is further away from the PsaL side [22]. Larger than monomer fractions of a plant PSI have been reported [23], and we do not observe a moiety from the dimer that would affect the contact surface of a plant LHCII. Since 2D projections from negative stain images of *Arabidopsis* PSI suggest the presence of a Lhca hetero-dimer that is analogous to alga [24], a state transition complex can be in principle structurally established with the PSI dimer in plants (Fig. 3d).

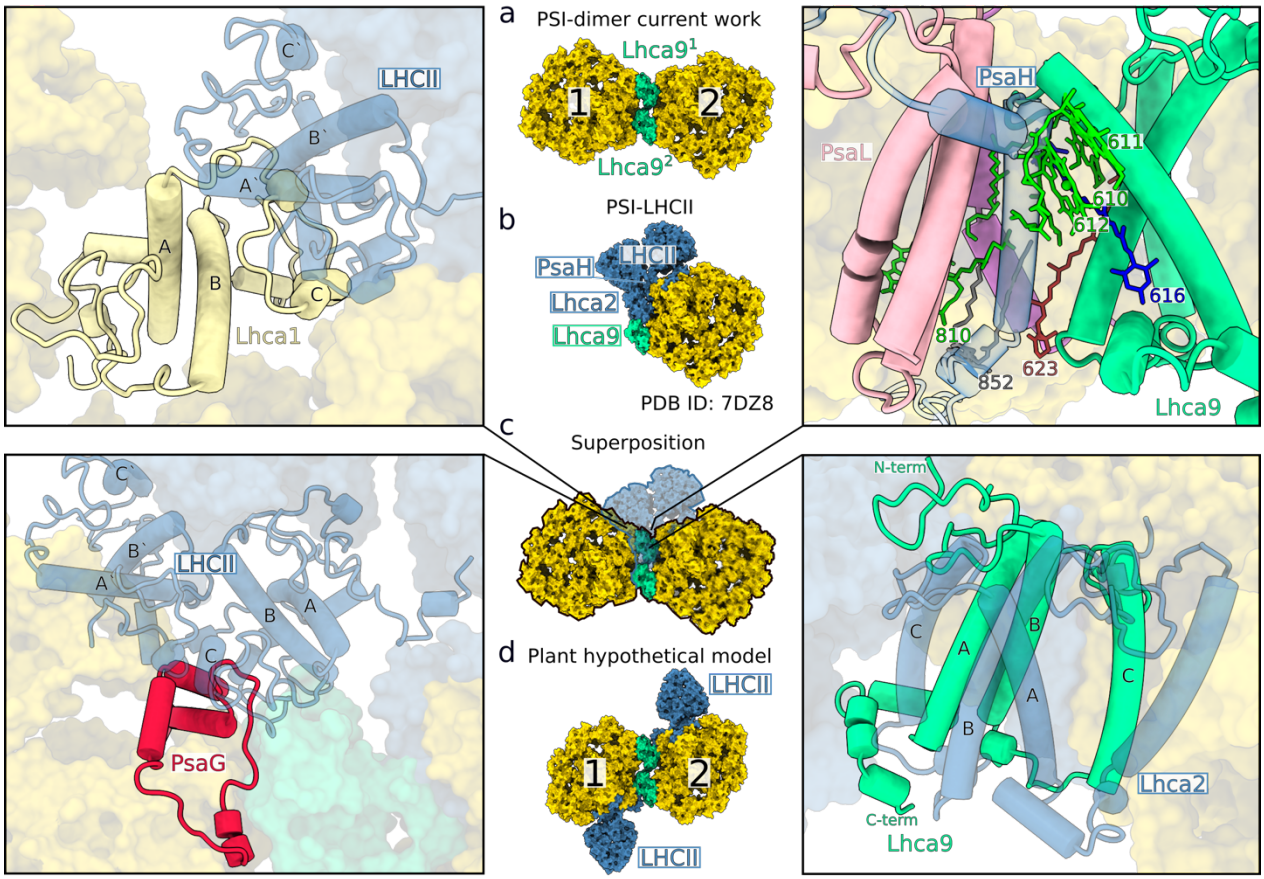


Fig. 3: PSI dimer and LHCII association in *C. reinhardtii*. a, The PSI dimer of the current work. **b,** The PSI with LHCII in state transition (PDB ID: 7DZ8). **c,** Superposition of PSI dimer with PSI-LHCII shows clashes of Lhca2:LHCII with PsaG, PsaL, Lhca1, Lhca9 from the neighbouring PSI monomer. d, A hypothetical model of a plant PSI dimer with a single LHCII copy is presented.

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. 8). 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 8a). 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. 8b-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 (Lhca9[2]), CHL606 (Lhca9[2]), 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 a considerable variability (Extended Data Fig. 8e, 8f). Therefore, similar to the cyanobacterial PSI dimer, an excitation coupling between the two monomers is less favourable *in vitro*, and this is consistent with measurements of 77 K fluorescence spectra that showed only a minor shift between monomer and dimer (Extended Data Fig. 1). 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.

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 unprecedented structural details of chlorophylls, carotenoids, and water molecules (Fig. 4, Extended Data Fig. 4 and 10, Supplementary Table 2). The map can improve the level of detail not only when compared to the previous cryo-EM studies of algal PSI [1-3], but also the plant PSI maps obtained by X-ray crystallography [16,17] (Extended Data Fig. 9). 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 beta-carotene and more accurate estimation of the coordination of 53 chlorophylls (Fig. 4b, Extended Data Fig. 10). Particularly, Chl *b* molecules are identified at positions 601 and 606 in Lhca4, Lhca5, Lhca6, Lhca7 and Lhca8. The two newly modeled 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 (Fig. 4b). The newly modeled beta-carotene 622 is in Lhca9 and could be identified due to structural stabilization of the interface region in the dimer (Fig. 4b). Since chl *b* limits free diffusion of excitation energy [25], 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).

In addition, on the luminal side of PSI, we observed a density corresponding to the associated electron donor Pc. Signal subtraction, followed by focused 3D classification, allowed us to rigid body fit a model for Pc into the density at the local resolution of ~3.5 Å (Extended Data Fig. 3). We then performed flexible fitting using self-similarity-restraints in *Coot* [26,27]. With respect to the PSI-Pc interactions, comparison between our model with a plant counterpart [28,29] revealed two main differences (Extended Data Fig. 11). In *C. reinhardtii*, the negatively charged residues of the Pc acidic patch are shifted by ~5 Å due to the missing residues P58-E59, and therefore, the interaction with K101 of PsaF is weakened (Extended Data Fig. 11). Instead, the binding strength

203 is compensated with the PsaF region K78-K92, which has six lysines (78, 81, 82, 85, 89, 92)
204 increasing a positively charged concentration at a distal site (Extended Data Fig 11c,d), thus
205 supporting additional electrostatic interactions with the acidic residues of Pc. The importance of
206 the distal lysine residues is supported by site-directed mutagenesis of the *C. reinhardtii* N-
207 terminal PsaF domain and functional analyses of electron transfer between Pc and mutant as well
208 as wild type PSI [30]. Thus, algae and plant Pc adapted their slightly different interfaces for optimal
209 interactions with PSI.

210 A striking feature of the high resolution cryo-EM map is resolvable density for multiple newly
211 detected water molecules, which particularly aided in modeling the coordination of chlorophylls
212 (Fig. 4c). Thus, we report the most complete available experimental picture of a chemical
213 environment for chlorophyll binding (Supplementary Table 3). Particularly, it allows
214 distinguishing between mono- and di-hydrated forms (Extended Data Fig. 12), which largely
215 escaped detection by X-ray crystallography (Extended Data Fig. 9). This is mechanistically
216 important, because the di-hydrated derivative is chemically more stable, as illustrated by quantum
217 chemical calculations [31]. We observe that, other than the previously reported CLA824 [16], only
218 two waters can be involved in penta-coordinated Mg for all the chlorophylls. Remarkably, waters
219 play a coordinative role for most of Chl *b*, for which the relative ratio of water coordination is
220 four times higher than for Chl *a* (Supplementary Table 3). The difference between Chl *a* and Chl
221 *b* is a methyl versus a formyl group, thus water serves as a hydrogen bond donor to the latter,
222 while it also interacts with charged/polar protein residues or lipids. Therefore, the immediate
223 surrounding of Chl *b* molecules is more enriched with non-protein material than previously
224 thought, which plays a role in tuning the photophysics and the transport properties of excitation
225 energy in PSI. Together, the presented model allows now for comparison of PSI phylogenetic
226 conservation also on the level of chlorophyll coordination and solvent positioning.

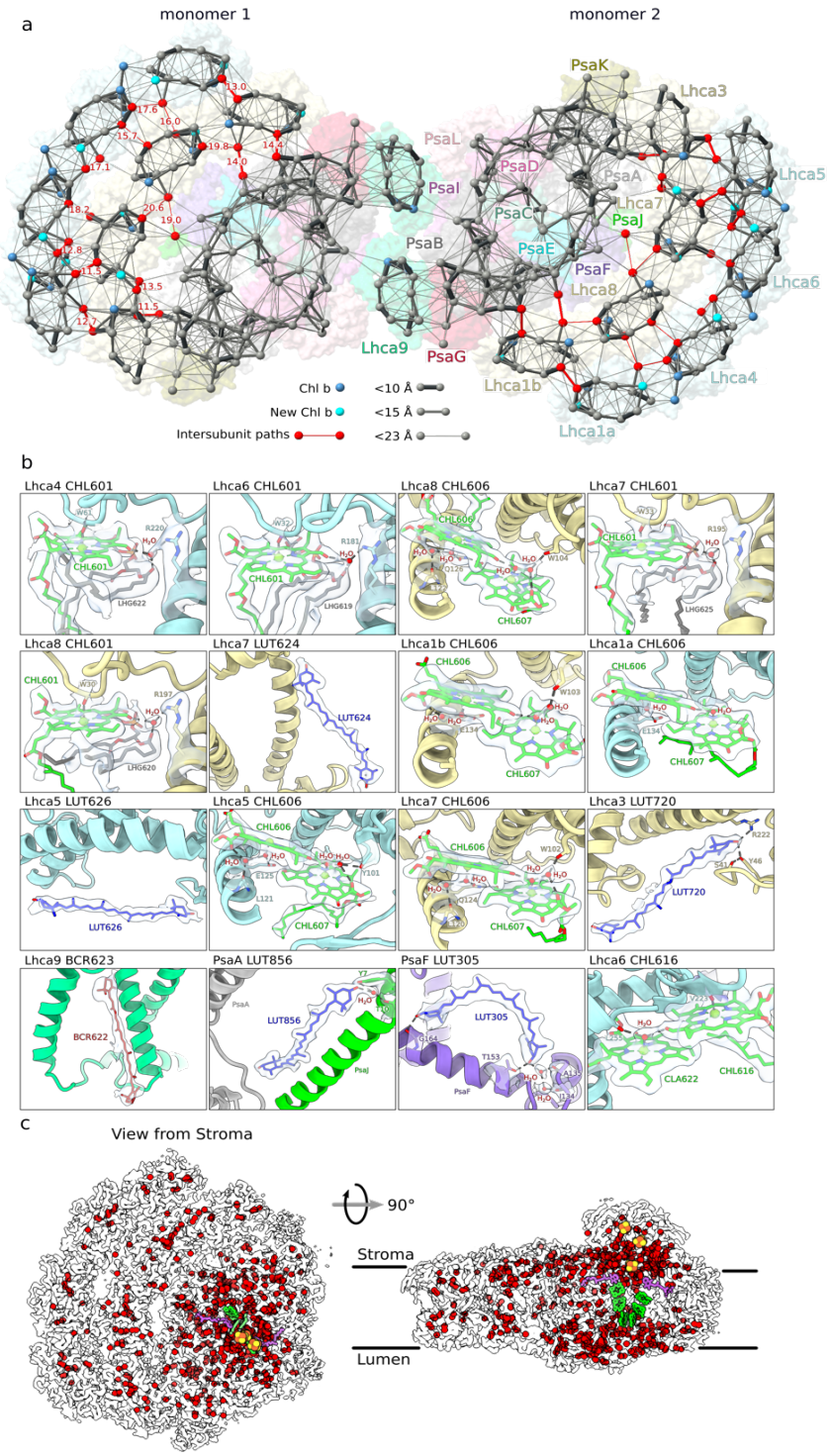


Fig. 4: High resolution features of co-factors and solvation. a, Energy transfer pathways. Distribution of chlorophylls indicated by Mg atoms. Chl *a* gray, Chl *b* blue, newly identified Chl *b* cyan, pathways within 23 Å are connected by lines, and the line width reflects the distance. The most likely intersubunit pathways are red. **b**, Close up views of newly identified elements, coordination and chemical environment with the density map. **c**, Water molecules modeled in PSI are shown as red spheres (oxygen atoms) in the outline of the map, chlorophyll green, phylloquinones purple, iron-sulfur clusters yellow-red.

235 # Methods

236 **Strains and growth conditions**

237 Experiments were performed a previously described strain expressing PsaB fused with His_{20}-tag
238 after the third residue from the N-terminus [2] and an *lhca9* insertional mutant [32] transformed with
239 the corresponding PsaB His tag plasmid. Chloroplast transformation and transformant selection
240 was performed as described previously. Further experiments were performed with the *C.*
241 *reinhardtii* wild type strain 137c as well as *lhca2* and *lhca9* insertional mutant[32] back-crossed
242 into wild type 137c.
243 All strains were maintained on TAP medium [33], solidified with 1.5% w/v agar at 25 °C in the
244 presence of ~20 µmol photons m^{-2} s^{-1} photosynthetically active, continuous illumination. For
245 experiments, strains were cultured in TAP medium on a rotary shaker (120 rpm) at 25 °C in the
246 presence of ~20 µmol photons m^{-2} s^{-1} photosynthetically active, continuous illumination.
247

248 **Purification of PSI**

249 *C. reinhardtii* cells featuring were incubated in anoxic conditions (~10^8 cells mL^{-1} in TAP
250 medium + 10 mM glucose, 40 U mL^{-1} glucose oxidase, 50 U mL^{-1} catalase) and dim light (~20
251 µmol photons m^{-2} s^{-1}) for 60 min. All following steps were performed at 4 °C and dim light. Cells
252 were disrupted in 0.33 M sucrose, 25 mM HEPES-KOH pH = 7.5, 5 mM $MgCl_2$, 1 mM PMSF, 1
253 mM benzamidine, 5 mM aminocaproic acid with a nebulizer (2 bar, two passages). Thylakoid
254 membranes were extracted via sucrose density step gradient ultracentrifugation (0.5 M – 1.3 M –
255 1.8 M sucrose, 5 mM HEPES-KOH pH = 7.5, 10 mM EDTA, 1 mM benzamidine, 5 mM
256 aminocaproic acid).
257 Isolated thylakoids were set to 1 mg chl mL^{-1} in 5 mM HEPES-KOH pH = 7.5 and solubilized by
258 addition of an equal volume of 2 % α-DDM for 10 min. Unsolubilized material was separated by
259 centrifugation. The supernatant was diluted four times to 125 µg chl mL^{-1} and 0.25 % α-DDM.
260 The sample was loaded onto a TALON metal affinity column (1 mL resin mg chl^{-1}) in 5 mM
261 HEPES-KOH pH = 7.5, 100 mM NaCl, 5 mM $MgSO_4$, 10 % glycerol at a flow rate of ~0.5 mL
262 min^{-1}. The column was washed with 10 times the bed volume of 5 mM HEPES-KOH pH = 7.5,
263 100 mM NaCl, 5 mM $MgSO_4$, 10 % glycerol, 0.02 % α-DDM at a flow rate of ~1 mL min^{-1}. A
264 second wash was performed with 10 times the bed volume of 5 mM HEPES-KOH pH = 7.5, 100
265 mM NaCl, 5 mM $MgSO_4$, 10 % glycerol, 0.02 % α-DDM and 5 mM imidazole at a flow rate of
266 ~1 mL min^{-1}. The PSI was eluted with 5 mM HEPES-KOH pH = 7.5, 100 mM NaCl, 5 mM
267 $MgSO_4$, 10 % glycerol, 0.02 % α-DDM and 150 mM imidazole. The PSI was concentrated with a
268 spin column (regenerated cellulose: 100,000 MWCO) to ~ 1.5 mg chl mL^{-1}, diluted five times
269 with 30 mM HEPES-KOH pH = 7.5, 0.02 % α-DDM and reconcentrated twice.
270 Pc crosslinking was performed in 30 mM HEPES-KOH pH = 7.5, 1 mM ascorbate, 0.1 mM
271 DAD, 3 mM $MgCl_2$ with 0.1 mg chl mL^{-1} PSI particles and 20 µM activated Pc for 45 min at
272 room temperature. Pc activation was performed in 10 mM MOPS-KOH pH = 6.5 with 100 µM
273 recombinant PC, 5 mM EDC and 10 mM sulfo-NHS for 20 min at room temperature. The
274 crosslinker was removed and the buffer exchanged to 30 mM HEPES-KOH pH = 7.5 via a PD

275 G25 desalting column followed by ultrafiltration with a centricon (regenerated cellulose: 10,000
276 MWCO).
277 The crosslinked PSI particles (~ 60 µg chl per gradient) or solubilized thylakoids of non-tagged
278 strains were loaded onto a 1.3 M – 0.1 M sucrose density gradient including 5 mM HEPES-KOH
279 pH = 7.5 and 0.02 % α-DDM. PSI fractions were collected after ultracentrifugation at 33,000 rpm
280 (Beckman Coulter SW 41 Ti) for 14 h (Extended Data Fig. 1b and 6a). Prior to further analysis,
281 sucrose was removed via a PD G25 desalting column followed by concentration with a spin
282 column (regenerated cellulose: 100,000 MWCO).
283

284 **Biochemical analysis of PSI**

285 For SDS-PAGE (Extended Data Fig. 1c-e), samples were adjusted to 1 µg chl, supplemented
286 with loading buffer and incubated at 65 °C for 15 min. Proteins were separated by 13% (w/v)
287 SDS-PAGE [34]. Gels were stained with Coomassie Brilliant Blue R-250 or blotted onto
288 nitrocellulose membranes (Amersham) and detected by specific primary antibodies: PsaF [35], PC
289 (Agrisera), PsaA, Lhca5, Lhca2, Lhca9 [4], Lhca3, PsaD [36], PsaG and LhcB/A [37]. Secondary
290 antibodies for ECL detection were anti-rabbit (Invitrogen).
291 For quantitative analysis by mass spectrometry (Extended Data Fig. 1f and 6b), removal of
292 sucrose and protein digestion was carried out following the FASP protocol, using 2 µg of
293 sequencing grade trypsin (Promega) per fraction [38]. Iodoacetamide and dithiothreitol used in the
294 original protocol were replaced by chloroacetamide and tris (2-carboxyethyl) phosphine,
295 respectively. After over-night digestion at 37°C, samples were acidified by adding trifluoroacetic
296 acid (TFA) to a final volume of 0.1%. Five percent of the peptide solution were desalted using
297 self-made Stage tips according to established protocols [39]. Desalted peptides were dried by
298 vacuum centrifugation and stored at -20°C until further use. The LC-MS/MS system consisted of
299 an Ultimate 3000 RSLC nanoLC System (Thermo Fisher Scientific) coupled via a Nanospray
300 Flex ion source (Thermo Fisher Scientific) to a Q Exactive Plus mass spectrometer (Thermo
301 Fisher Scientific). Samples were reconstituted in 5 µl of 2 % (v/v) acetonitrile/0.05% (v/v) TFA
302 in ultrapure water (eluent A1), loaded on a trap column (C18 PepMap 100, 300 µM x 5 mm, 5
303 µm particle size, 100 Å pore size; Thermo Fisher Scientific) and desalted for 3 min at a flow rate
304 of 15 µL min^{-1} using eluent A1. Subsequently, the trap column was switched in-line with an
305 Acclaim PepMap100 reversed phase column (75 µm x 50 cm, 2 µm particle sizes, 100 Å pore
306 size; Thermo Fisher Scientific) for peptide separation. The mobile phases were composed of 0.1
307 % (v/v) formic acid in ultrapure water (eluent A2) and 80 % (v/v) acetonitrile/0.08 % (v/v)
308 formic acid in ultrapure water (B). Flow rate was 250 nL/min. The following gradient was
309 applied: 5-35 % B over 105 min, 35-99 % B over 5 min, 99 % B for 20 min. MS full scans (scan
310 range m/z: 350–1400, resolution 70,000 at m/z 200, AGC target value 3e6, maximum injection
311 time 50 ms) were acquired in data-dependent mode, dynamically selecting the 12 most abundant
312 precursor ions for fragmentation by higher-energy C-trap dissociation (HCD, 27 % normalized
313 collision energy, resolution 17,500 at m/z 200, precursor isolation window 1.5 m/z). Dynamic
314 exclusion was set to 'auto' (chromatographic peak width: 15 s). AGC target value and intensity
315 threshold for MS/MS were 5e4 and 1e4, respectively, at 80 ms maximum ion fill time. Singly

316 charged ions, ions with charge state 5 or above and ions with unassigned charge states were
317 rejected. Internal lock mass calibration was enabled on m/z 445.12003. LC-MS/MS data was
318 processed in MaxQuant 1.6.14 for protein identification and label-free quantification [40]. Default
319 settings were used, except for calculation of iBAQ values and "second peptide search" which
320 were enabled and disabled, respectively. Spectra were searched against a concatenated database
321 containing protein sequences based on the *Chlamydomonas* v5.6 gene models (Joint Genome
322 Institute, www.phytozome.org), as well as sequences of chloroplast- and mitochondrial-encoded
323 proteins (GenBank BK000554.2 and NC_001638.1). Carbamidomethylation of cysteines was set
324 as a fixed modification. Oxidation of methionine and acetylation of protein N-termini were
325 considered as variable modifications. A false discovery rate (FDR) of 1% was applied to peptide
326 and protein identifications. LFQ data was imported into Perseus (version 1.6.15.0) [41] for log2-
327 transformation, and contaminants, proteins only identified by site and reverse hits were removed.
328 iBAQ values were normalized to the PSI core subunit PsaB.
329 Room temperature absorption spectra (300 – 750 nm) of PSI monomer and dimer fractions
330 (Extended Data Fig. 1g) were measured with a UV-/Vis spectrophotometer (V-650, Jasco) at 10
331 µg chl mL^{-1} and normalized to the red region. Fluorescence emission spectra (650 – 780 nm) of
332 PSI fractions (Extended Data Fig. 1h) were recorded at 77 K at 1 µg chl mL^{-1} with a
333 spectrofluorometer (P-6500, Jasco) upon excitation at 436 nm. Spectra were normalized to the
334 maximum of the emission peaks and smoothed according to [42]. For 2D-PAGE (Extended Data
335 Fig. 2), thylakoids (0.8 mg chl mL^{-1}) isolated from control and anoxic 137c wild type cells were
336 solubilized with 0.9 % β-DDM for 20 min. 2D-PAGE [43], silver staining, tryptic in-gel digestion [37]
337 and LC-MS/MS was operated as described [36], where a Ultimate 3000 nano-LC system was
338 coupled via a nanospray source to an LCQ Deca XP mass spectrometer (Thermo Finnigan).
339 LC-MS/MS data were processed with Proteome Discoverer (Thermo Fisher Scientific, version
340 2.4). Raw files were searched using the SequestHT and MS Amanda algorithms against a
341 concatenated database containing sequences of nuclear- (Chlamydomonas v5.6 gene models,
342 www.phytozome.org), chloroplast- (GenBank BK000554.2) and mitochondrial-encoded
343 (NC_001638.1) proteins Search settings were: precursor and fragment mass tolerances 250 ppm
344 and 0.25 Da, respectively; minimum peptide length: 6; maximum of missed cleavages: 2; variable
345 modifications: oxidation of methionine, N-acetylation of protein N-termini. Identifications were
346 filtered to achieve a peptide and protein level false discovery rate of 0.01.
347

348 **Cryo-EM data collection and processing**

349 3 µL of purified PSI complex at 1 mg Chl mL^{-1} were applied to glow-discharged (GloQube
350 Quorum, 40 sec, 20mA) holey carbon grids (Quantifoil 300 Au R1.2/R1.3, Electron Microscopy
351 Sciences) and vitrified using a Vitrobot MKIV (2 sec blotting time, 4°C, 100% humidity). The
352 data collection was carried out using a 300KV Titan Krios G2 Microscope (Thermo Fisher
353 Scientific) equipped with a Gatan Bioquantum energy filter and a K3 Summit direct electron
354 detector (Ametek). Movies were recorded using counting mode at a magnification of ×105,000
355 corresponding to a calibrated pixel size of 0.84 Å. The dose rate was 15.27 e/px/sec and the
356 exposure time was 3 sec divided into 45 frames leading to a total dose of 45.8 e $Å^{-2}$. EPU was

357 used to collect 17,439 movies with a defocus range from -0.7 to -2.5µm. Data statistics is shown
358 in Supplementary Table 2.

359 The Extended Data Fig. 3 shows the processing scheme applied. The pre-processing steps were
360 performed using cryoSPARC 3.1.0 [44]. Movie stacks were motion corrected and dose weighted
361 using MotionCor2 [45]. Contrast Transfer Function (CTF) of the motion corrected micrographs was
362 estimated using CTFFIND4 [46]. Blob picker and then template picker were used to pick 440,494
363 particles and 2D classification in cryoSPARC was performed. Dimeric particles were separated
364 from monomeric by inspection of the 2D-class averages and for each sub-population an *Ab initio*
365 model was generated using cryoSPARC applying C2 and C1 symmetry, respectively. For each
366 model homogenous refinement was performed leading to a nominal resolution of 3.7 Å for the
367 dimer and 3.0 Å for the monomer. Particles (dimer: 69,144 monomer: 123,746) were converted
368 into a Star file format [47] and were imported into RELION 3.1.beta [48,49]. Particles were re-
369 extracted (un-binned) and processed in RELION using a box size of 700 pixel and 500 pixel for
370 the dimer and monomer, respectively. 3D Refinement followed by 3D classification was
371 performed imposing C2 symmetry for the dimer and C1 for the monomer. A subset of high
372 quality particles was selected for the dimer and monomer and subjected to 3D refinement which
373 resulted in a resolution of 3.3 Å for the dimer and 2.9 Å for the monomer. CTF refinement [50,51]
374 followed by 3D refinement and Bayesian polishing followed by another round of CTF refinement
375 was performed for the dimer as well as for the monomer. A final 3D refinement resulted in an
376 overall resolution of 2.97 Å for the dimer and 2.31 Å for the monomer. The resolution of the
377 dimer could be further improved to 2.74 Å by using signal subtraction of one monomer followed
378 by symmetry expansion and 3D refinement applying C1 symmetry.

379 In order to increase the number of particles for classification on the Pc region, Dataset 2 was
380 collected from the same dimer band, but with a pixel size of 0.51 Å. The dataset was processed
381 with cryoSPARC 3.1.0 [44]. After template picking 864,399 particles were extracted. With a small
382 subset of the extracted particles an *Ab initio* reconstruction was generated followed by
383 Heterogeneous Refinement using 5 classes, one of which contained *Ab initio* reconstruction as
384 reference. The class containing the PSI monomer was then subjected to Homogeneous
385 Refinement in cryoSPARC 3.1.0 resulting in a reconstruction at 3.88 Å resolution. The particles
386 were then exported to RELION [47] and 3D classification was performed. The class that contained
387 88,219 good particles was used for further refinement which improved the overall resolution to
388 3.5 Å. Applying CTF-refinement and Bayesian Polishing resulted in further improvement, and
389 the final nominal resolution is 2.68 Å. The data was then merged with the monomer (Dataset 1)
390 and signal subtraction followed by focused classification using a mask around the Pc region was
391 performed. A class with 66,080 particles showed the best density for Pc which was used for
392 model building. The workflow is further illustrated in the Extended Data Fig 3a.

393 For analysis of the motion between the two monomers of the dimer we performed a Multibody
394 Refinement in RELION [52] followed by a principal component analysis using the program
395 relion_flex_analyse. Two bodies were chosen, one for each monomer, resulting in twelve
396 Eigenvectors describing the motion. Ten maps for each of the three Eigenvectors which describe

397 about 78% of the motion in the data were printed out and the maps with the extreme positions
398 (map1 and 10) were used to fit the models that are shown in Extended Data Figure 8. A Python
399 script was used to estimate the distances between the chlorophylls at the dimer interface for each
400 particle in the data and to plot the results as histograms as depicted in Extended Data Figure 8e,f.
401

402 **Model building and refinement**

403 Initially, the available model of the PSI structure (PDB ID: 6JO5) of *C. reinhardtii* was rigid
404 body fitted into the 2.74 Å map of the symmetry expanded dimer using Chimera v 1.14 [53]. Model
405 building and real-space refinement was then carried out using Coot v9.1.4 [26,27] to complete one
406 monomer. Two copies of the completed monomeric model were then rigid body fitted into the C2
407 generated 2.97 Å dimer map using Chimera. The model of the monomer was then fitted
408 separately in the highest resolution 2.3 Å map of the monomer. All protein residues as well as
409 pigments were fitted using *Coot* with locally optimized map weights. The cis-trans isomerism of
410 each pigment was judged based on density and modeled accordingly. Newly identified
411 chlorophylls and carotenoids were modeled, when the experimental evidence (density map)
412 supported and the chemical environment matched the surrounding of the pigments. For
413 carotenoids, the density that clearly showed the characteristics of an elongated tetraterpenoid with
414 densities for the four methyl groups, two sticking out on each side of the chain, was identified as
415 a new carotenoid binding site. To further analyze the identity of the corresponding pigment
416 possible candidates were fitted and compared. A carotenoid that fitted best in terms of density
417 and chemical environment was then selected. In case of luteins, particularly the oxygen of the
418 cyclohexane-ring was the main criterion for pigment identity, because it is involved in hydrogen
419 bonding. For the Chl *b* identification the densities for the aldehyde group needed to be present as
420 well as the hydrogen bonding occurring with a water molecule that are usually stabilized by other
421 chlorophylls, lipids, and protein side chains. Water molecules involved in pigment interactions
422 were placed manually. All other water molecules were picked by *Coot* with the auto picking
423 function followed by manual inspection and correction. All high resolution features were
424 modeled using *Coot* until the model was completed. For all modeling steps restraint files for
425 pigments and ligands were used which were generated using the Grade server
426 (http://grade.globalphasing.org). Restraint files were adopted manually if it was required.

427 For plastocyanin, a model was generated using SWISS model [54]. The model was then rigid body
428 fitted using Chimera. Rotamers were corrected for the residues that were allowed due to the
429 better local densities. Self-restraints in Coot were activated followed by flexible fitting into the
430 density. All models were refined using Real-Space-Refine from the PHENIX suite [55] using the
431 Grade server restraint files for the ligands and a distance .edit file which was generated by Ready-
432 set in PHENIX. Further, hydrogen atoms were added for refinement to the model using Ready-
433 set. The refinement protocol was optimized using different weight parameters. The refinement
434 statistics are shown in Supplementary Table 2. Multiple rounds of validation and model building
435 were carried out using MolProbity [56] and *Coot*. For further validation the PDB Validation server
436 was used (https://validate-rcsb-2.wwpdb.org).

The structure was analyzed using *Coot* and Chimera. Figures were prepared using ChimeraX [57].

Data availability

The atomic coordinates have been deposited in the Protein Data Bank (PDB) and are available under the accession codes: PDB ID: XXX for the dimer, PDB ID: XXXX for the symmetry expanded dimer, PDB ID: XXXX for the high resolution monomer, and PDB ID: XXXX for the PSI-Pc model. The local resolution filtered maps, half maps, masks and FSC-curves have been deposited in the Electron Microscopy Data Bank with the accession codes EMD-XXXXX for the dimer, EMD-XXXXX for the symmetry expanded dimer, EMD-XXXXX for the high resolution monomer, and EMD-XXXXX for the Plastocyanin map. 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.

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, doi:10.1038/s41477-019-0379-y (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, doi:10.1038/s41477-019-0438-4 (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, doi:10.1038/s41477-019-0380-5 (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, doi:10.1104/pp.18.00749 (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, doi:10.1038/s41467-021-21362-6 (2021).

6 Pan, X. *et al.* Structural basis of LhcbM5-mediated state transitions in green algae. *Nat Plants*, doi:10.1038/s41477-021-00960-8 (2021).

7 Chen, M. *et al.* Distinct structural modulation of photosystem I and lipid environment stabilizes its tetrameric assembly. *Nat Plants* **6**, 314-320, doi:10.1038/s41477-020-0610-x (2020).

8 Kato, K. *et al.* Structure of a cyanobacterial photosystem I tetramer revealed by cryo-electron microscopy. *Nat Commun* **10**, 4929, doi:10.1038/s41467-019-12942-8 (2019).

9 Zheng, L. *et al.* Structural and functional insights into the tetrameric photosystem I from heterocyst-forming cyanobacteria. *Nat Plants* **5**, 1087-1097, doi:10.1038/s41477-019-0525-6 (2019).

10 Jordan, P. *et al.* Three-dimensional structure of cyanobacterial photosystem I at 2.5 A resolution. *Nature* **411**, 909-917, doi:10.1038/35082000 (2001).
11 Ben-Shem, A., Frolow, F. & Nelson, N. Crystal structure of plant photosystem I. *Nature* **426**, 630-635, doi:10.1038/nature02200 (2003).
12 Amunts, A. & Nelson, N. Plant photosystem I design in the light of evolution. *Structure* **17**, 637-650, doi:10.1016/j.str.2009.03.006 (2009).
13 Perez-Boerema, A. *et al.* Structure of a minimal photosystem I from the green alga Dunaliella salina. *Nat Plants* **6**, 321-327, doi:10.1038/s41477-020-0611-9 (2020).
14 Wood, W. H. J. *et al.* Dynamic thylakoid stacking regulates the balance between linear and cyclic photosynthetic electron transfer. *Nat Plants* **4**, 116-127, doi:10.1038/s41477-017-0092-7 (2018).
15 Yadav, K. N. *et al.* Supercomplexes of plant photosystem I with cytochrome b6f, light-harvesting complex II and NDH. *Biochim Biophys Acta Bioenerg* **1858**, 12-20, doi:10.1016/j.bbabio.2016.10.006 (2017).
16 Mazor, Y., Borovikova, A., Caspy, I. & Nelson, N. Structure of the plant photosystem I supercomplex at 2.6 A resolution. *Nat Plants* **3**, 17014, doi:10.1038/nplants.2017.14 (2017).
17 Wang, J. *et al.* Structure of plant photosystem I-light harvesting complex I supercomplex at 2.4 A resolution. *J Integr Plant Biol*, doi:10.1111/jipb.13095 (2021).
18 Muhleip, A., McComas, S. E. & Amunts, A. Structure of a mitochondrial ATP synthase with bound native cardiolipin. *Elife* **8**, doi:10.7554/eLife.51179 (2019).
19 Flygaard, R. K., Muhleip, A., Tobiasson, V. & Amunts, A. Type III ATP synthase is a symmetry-deviated dimer that induces membrane curvature through tetramerization. *Nat Commun* **11**, 5342, doi:10.1038/s41467-020-18993-6 (2020).
20 Muhleip, A. *et al.* ATP synthase hexamer assemblies shape cristae of Toxoplasma mitochondria. *Nat Commun* **12**, 120, doi:10.1038/s41467-020-20381-z (2021).
21 Bennett, C. F. *et al.* Peroxisomal-derived ether phospholipids link nucleotides to respirasome assembly. *Nat Chem Biol*, doi:10.1038/s41589-021-00772-z (2021).
22 Pan, X. *et al.* Structure of the maize photosystem I supercomplex with light-harvesting complexes I and II. *Science* **360**, 1109-1113, doi:10.1126/science.aat1156 (2018).
23 Amunts, A., Toporik, H., Borovikova, A. & Nelson, N. Structure determination and improved model of plant photosystem I. *J Biol Chem* **285**, 3478-3486, doi:10.1074/jbc.M109.072645 (2010).
24 Crepin, A., Kucerova, 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, doi:10.1111/tpj.14634 (2020).
25 Croce, R. & van Amerongen, H. Light harvesting in oxygenic photosynthesis: Structural biology meets spectroscopy. *Science* **369**, doi:10.1126/science.aay2058 (2020).
26 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).
27 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).
28 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, doi:10.1038/s41477-020-00779-9 (2020).

29 Caspy, I. *et al.* Structure of plant Photosystem I-plastocyanin complex reveals strong hydrophobic interactions. *Biochem J*, doi:10.1042/BCJ20210267 (2021).

30 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 c6 and plastocyanin to photosystem I of Chlamydomonas reinhardtii. *Proc Natl Acad Sci U S A* **95**, 7339-7344, doi:10.1073/pnas.95.13.7339 (1998).

31 Ben Fredj, A., Ben Lakhdar, Z. & Ruiz-Lopez, M. F. The structure of chlorophyll a-water complexes: insights from quantum chemistry calculations. *Chem Commun (Camb)*, 718-720, doi:10.1039/b716800d (2008).

32 Li, X. *et al.* A genome-wide algal mutant library and functional screen identifies genes required for eukaryotic photosynthesis. *Nat Genet* **51**, 627-635, doi:10.1038/s41588-019-0370-6 (2019).

33 Hoober, J. K. The Chlamydomonas Sourcebook. A Comprehensive Guide to Biology and Laboratory Use. Elizabeth H. Harris. Academic Press, San Diego, CA, 1989. xiv, 780 pp., illus. $145. *Science* **246**, 1503-1504, doi:10.1126/science.246.4936.1503-a (1989).

34 Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685, doi:10.1038/227680a0 (1970).

35 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, doi:10.1021/bi970082c (1997).

36 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, doi:10.1074/jbc.M414486200 (2005).

37 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, doi:10.1046/j.1365-313x.2001.01175.x (2001).

38 Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362, doi:10.1038/nmeth.1322 (2009).

39 Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols* **2**, 1896-1906, doi:10.1038/nprot.2007.261 (2007).

40 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, doi:10.1038/nbt.1511 (2008).

41 Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**, 731-740, doi:10.1038/nmeth.3901 (2016).

42 Savitzky, A. & Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* **36**, 1627-1639, doi:10.1021/ac60214a047 (1964).

43 Busch, A., Nield, J. & Hippler, M. The composition and structure of photosystem I-associated antenna from Cyanidioschyzon merolae. *Plant J* **62**, 886-897, doi:10.1111/j.1365-313X.2010.04202.x (2010).

44 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, doi:10.1038/nmeth.4169 (2017).

45 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332, doi:10.1038/nmeth.4193 (2017).
46 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221, doi:10.1016/j.jsb.2015.08.008 (2015).
47 Asarnow, D., Palovcak, E. & Cheng, Y. UCSF pyem v0.5. Zenodo https://doi.org/10.5281/zenodo.3576630. (2019).
48 Scheres, S. H. W. Amyloid structure determination in RELION-3.1. *bioRxiv* (2019).
49 Zivanov, J., Nakane, T. & Scheres, S. H. W. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ* **7**, 253-267, doi:10.1107/S2052252520000081 (2020).
50 Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, doi:10.7554/eLife.42166 (2018).
51 Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5-17, doi:10.1107/S205225251801463X (2019).
52 Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* **7**, doi:10.7554/eLife.36861 (2018).
53 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).
54 Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296-W303, doi:10.1093/nar/gky427 (2018).
55 Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544, doi:10.1107/S2059798318006551 (2018).
56 Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293-315, doi:10.1002/pro.3330 (2018).
57 Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25, doi:10.1002/pro.3235 (2018).

Acknowledgements

The cryo-EM data were collected at the SciLifeLab facility (funded by the KAW, EPS, and Kempe foundations) and the Karolinska Institutet 3D-EM facility. We thank D. Kimanius for providing a script for conversion of poses from Euler space into an orientation matrix in Cartesian space, S. Hawat for mass spectrometric analyses, V. Adlfar for isolation of Pc. Funding: Swedish Foundation for Strategic Research (ARC19-0051), Knut and Alice Wallenberg Foundation (2018.0080), Deutsche Forschungsgemeinschaft (739/13-2), Federal states (NRW 313-WO44A), German-Israeli Foundation for Scientific Research and Development (G-1483-207/2018). AA is supported by the EMBO Young Investigator Programme, MH is supported by the RECTOR programme (University of Okayama, Japan).

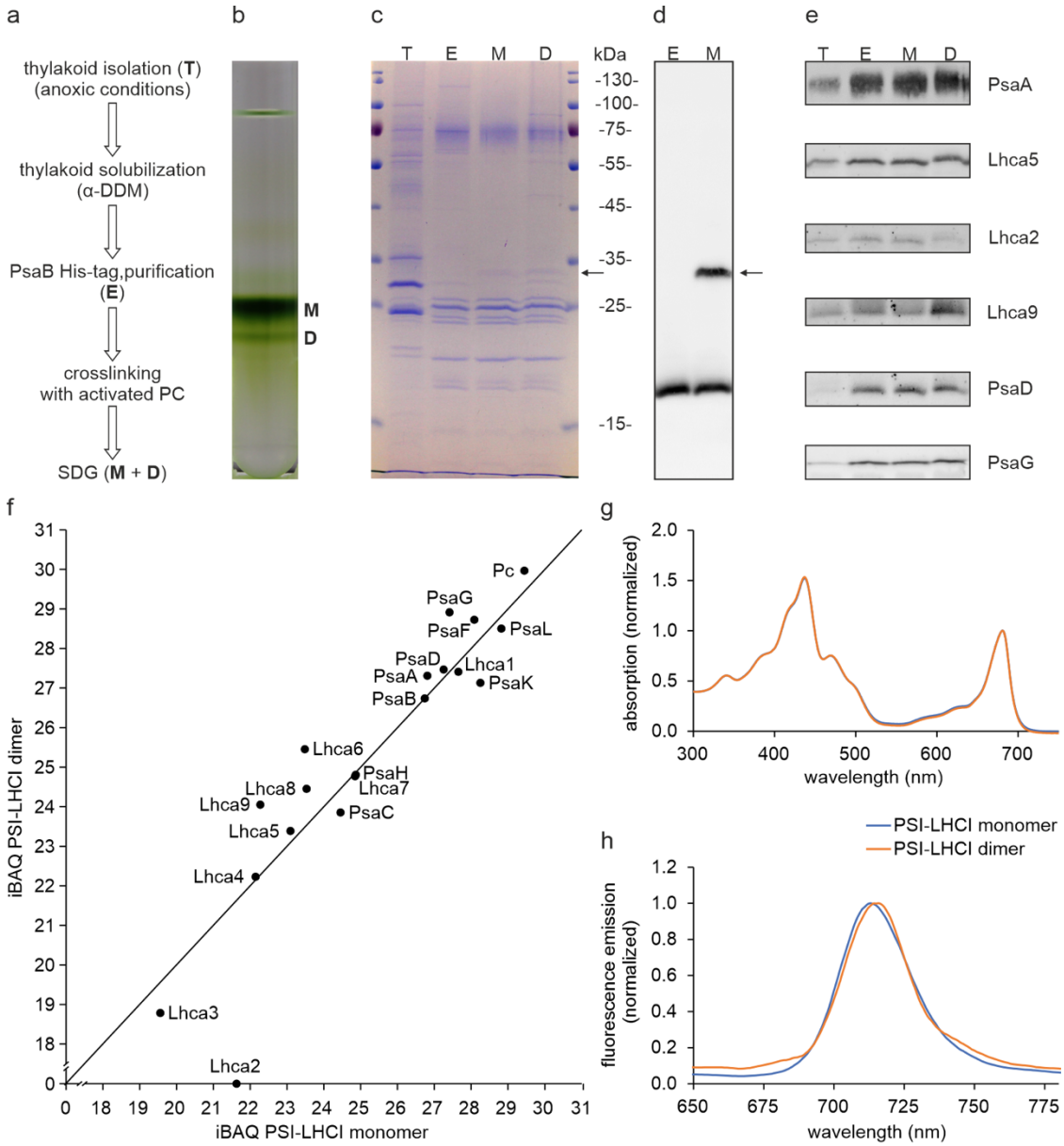
Author contributions

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.

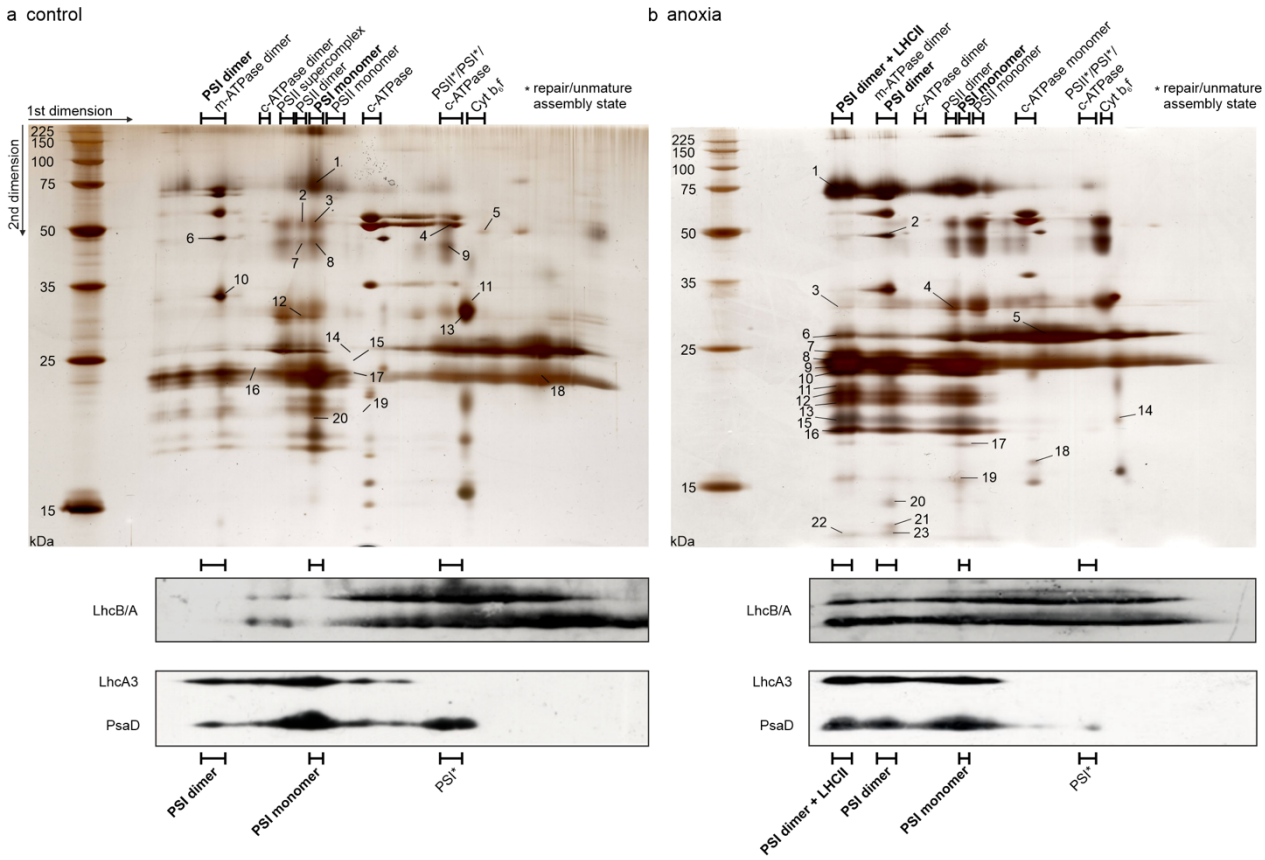
L.M. and S.K. performed biochemical analysis. V.T. performed computational analysis. Y.T. provided antibodies. A.N. and A.A. analysed the structure and wrote the manuscript with contributions from L.M. and M.H. All authors contributed to the analysis and the final version of the manuscript.

Competing interests

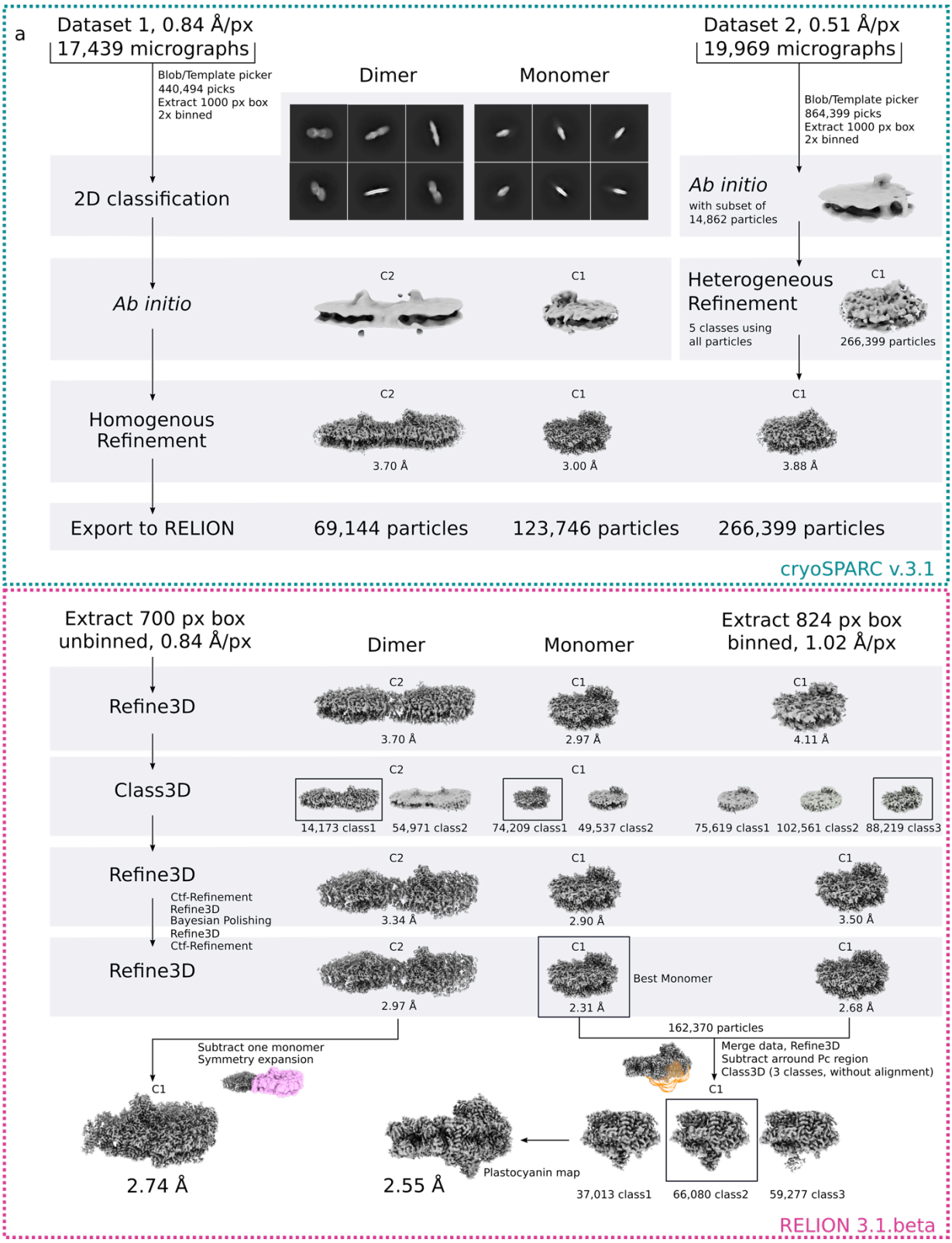
The authors declare no competing interests.



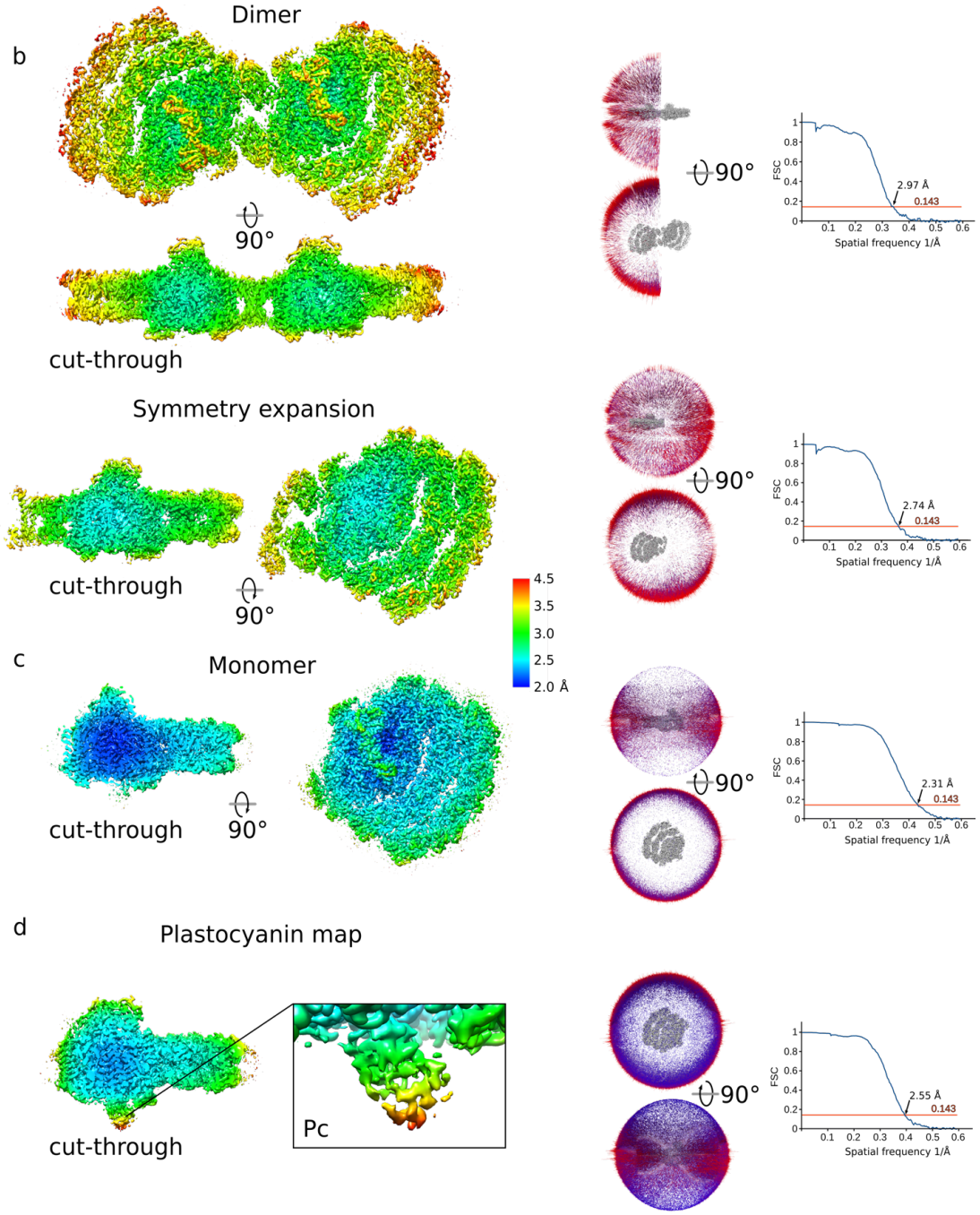
Extended Data Fig. 1: Sample preparation and characterisation. a, The experimental workflow. **b**, Sucrose density gradient affinity purified PSI (~ 60 µg chl). Monomer (M) and dimer (D) fractions are indicated. **c**, SDS-PAGE analysis (1 µg chl). T: thylakoids; E: PSAB-His elution, M: PSI monomer, 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. **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 [42].



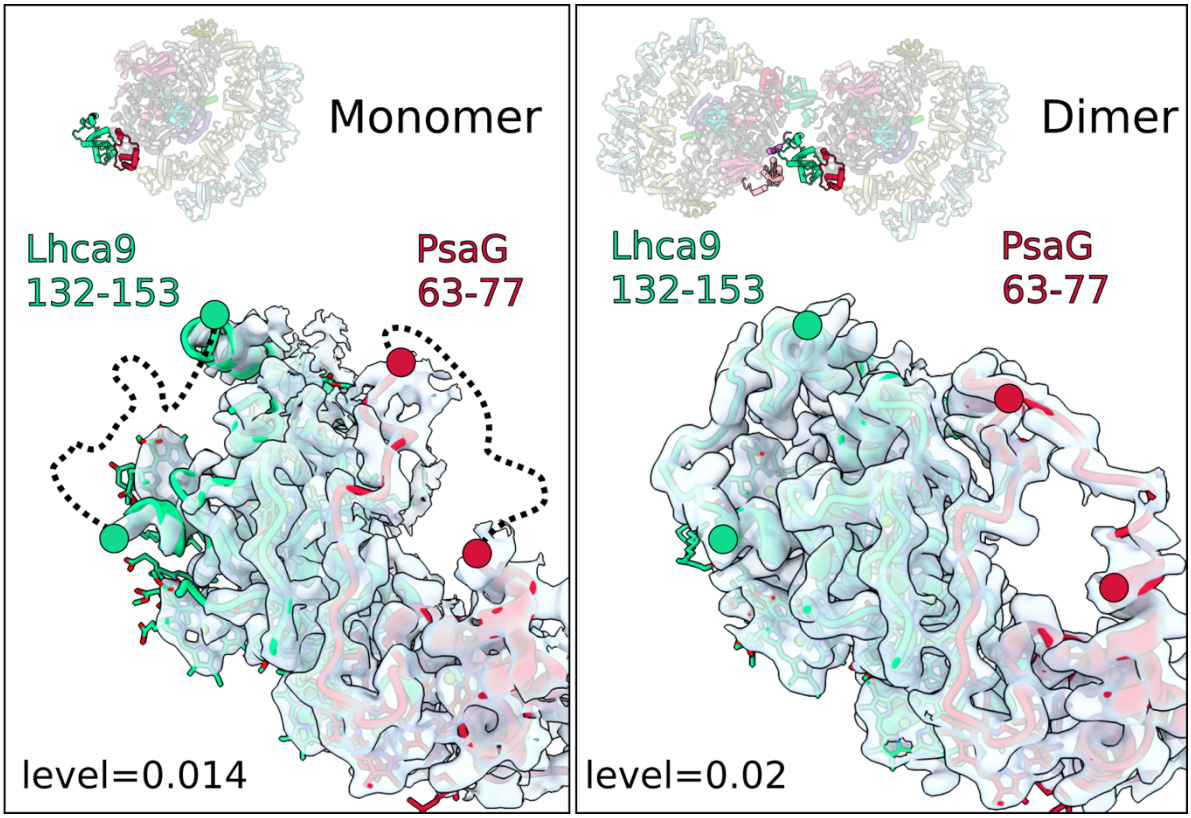
Extended Data Fig. 2: 2D-PAGE of solubilized *C. reinhardtii* wild type thylakoids. a, control conditions. **b**, low light and anoxic conditions. In the first dimension, multi-protein complexes were separated in their native form by blue native PAGE. In the second dimension, subunits of these protein complexes were separated by SDS-PAGE. Labelled silver-stained spots were subjected to LC-MS/MS analysis. Western Blots of LhcB/A, Lhca3 and PsaD indicate the distribution of LHCII, LHCI and the PSI core.


a
Dataset 1, 0.84 Å/px
17,439 micrographs
Blob/Template picker
440,494 picks
Extract 1000 px box
2x binned
Dimer
Monomer
Dataset 2, 0.51 Å/px
19,969 micrographs
Blob/Template picker
864,399 picks
Extract 1000 px box
2x binned
2D classification
Ab initio
with subset of
14,862 particles
Ab initio
C2
C1
Heterogeneous
Refinement
5 classes using
all particles
C1
266,399 particles
Homogenous
Refinement
C2
3.70 Å
C1
3.00 Å
C1
3.88 Å
Export to RELION
69,144 particles
123,746 particles
266,399 particles
cryoSPARC v.3.1
Extract 700 px box
unbinned, 0.84 Å/px
Dimer
Monomer
Extract 824 px box
binned, 1.02 Å/px
Refine3D
C2
3.70 Å
C1
2.97 Å
C1
4.11 Å
Class3D
C2
14,173 class1
54,971 class2
C1
74,209 class1
49,537 class2
75,619 class1
102,561 class2
88,219 class3
Refine3D
Ctf-Refinement
Refine3D
Bayesian Polishing
Refine3D
Ctf-Refinement
C2
3.34 Å
C1
2.90 Å
C1
3.50 Å
Refine3D
C2
2.97 Å
C1
2.31 Å
Best Monomer
C1
2.68 Å
162,370 particles
Subtract one monomer
Symmetry expansion
Merge data, Refine3D
Subtract arround Pc region
Class3D (3 classes, without alignment)
C1
C1
2.74 Å
2.55 Å
Plastocyanin map
37,013 class1
66,080 class2
59,277 class3
RELION 3.1.beta

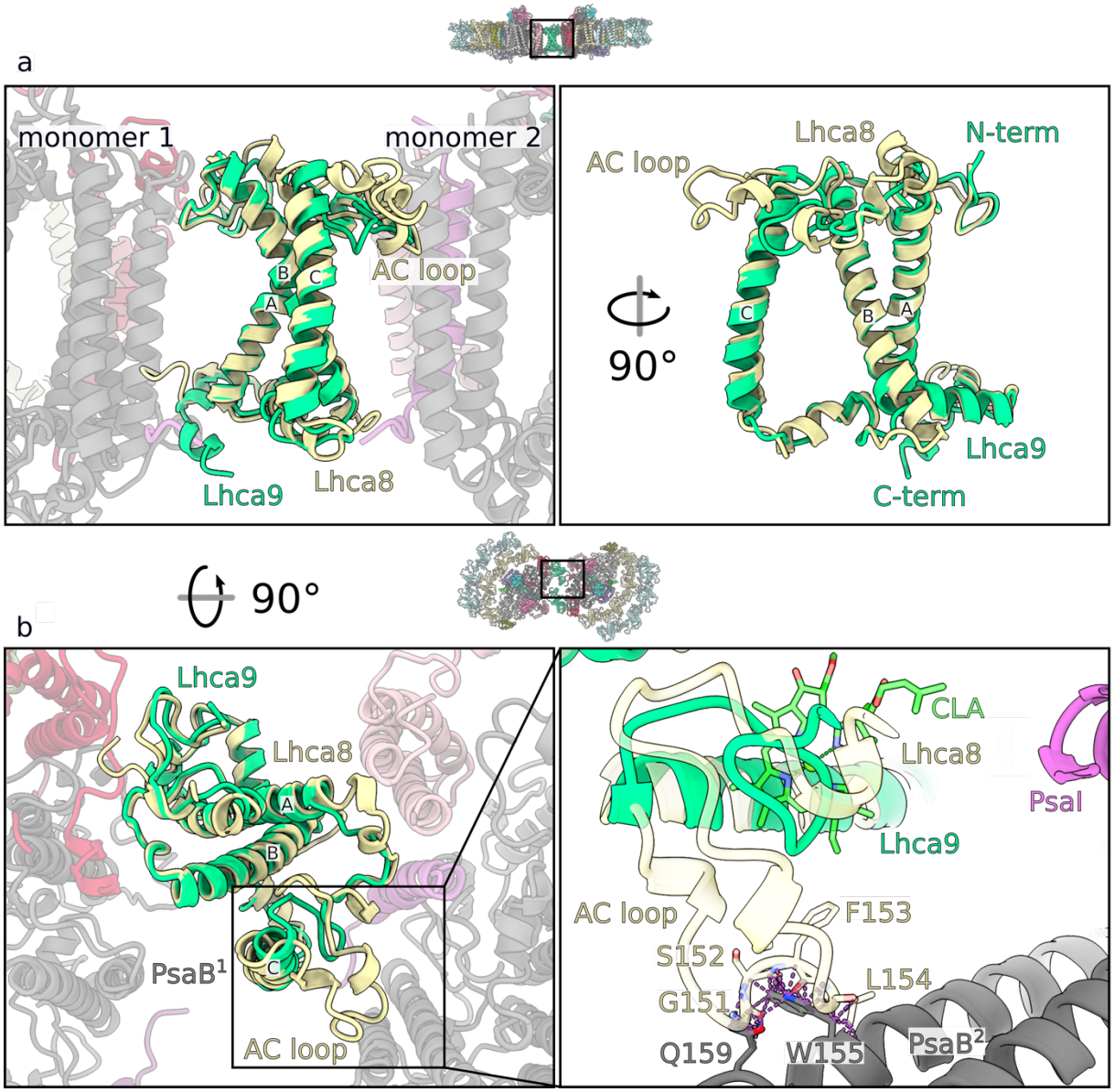
645



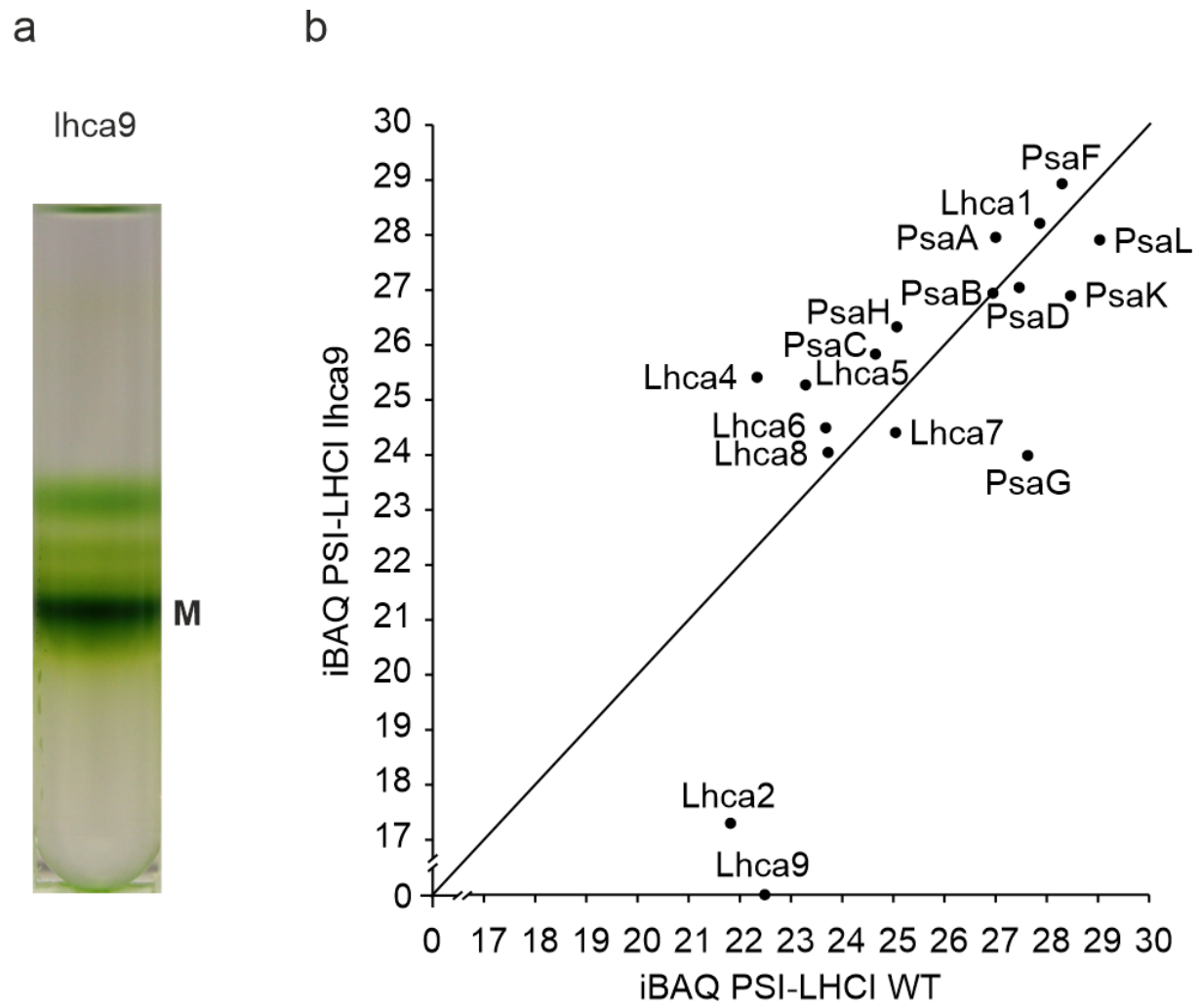
Extended Data Fig. 3: Cryo-EM processing workflow and local resolution. a, Flowchart of the data processing for the PSI dimer and monomer. The processing steps performed in cryoSPARC and RELION are indicated. To increase the numbers of Pc bound particles Dataset 2 was added for focused classification 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.



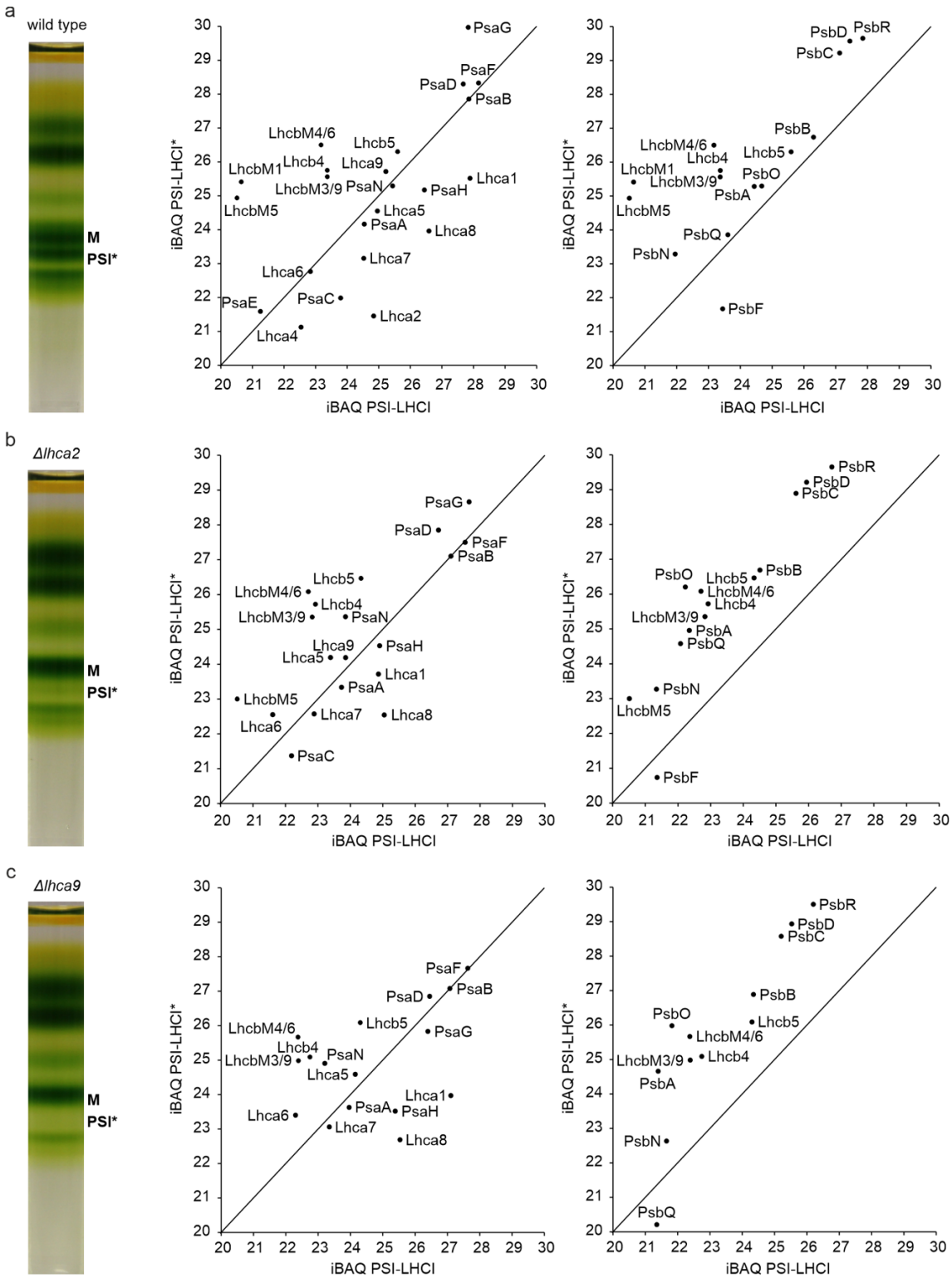
Extended Data Fig. 4: Stabilization of loops in Lhca9 and PsaG in PSI dimer. Left, in the monomer, the Lhca9 loop 132-153 and PsaG loop 63-77 are disordered (dashed lines). Right, in the dimer, both loop regions are ordered due to a stabilization by the neighboring monomer.



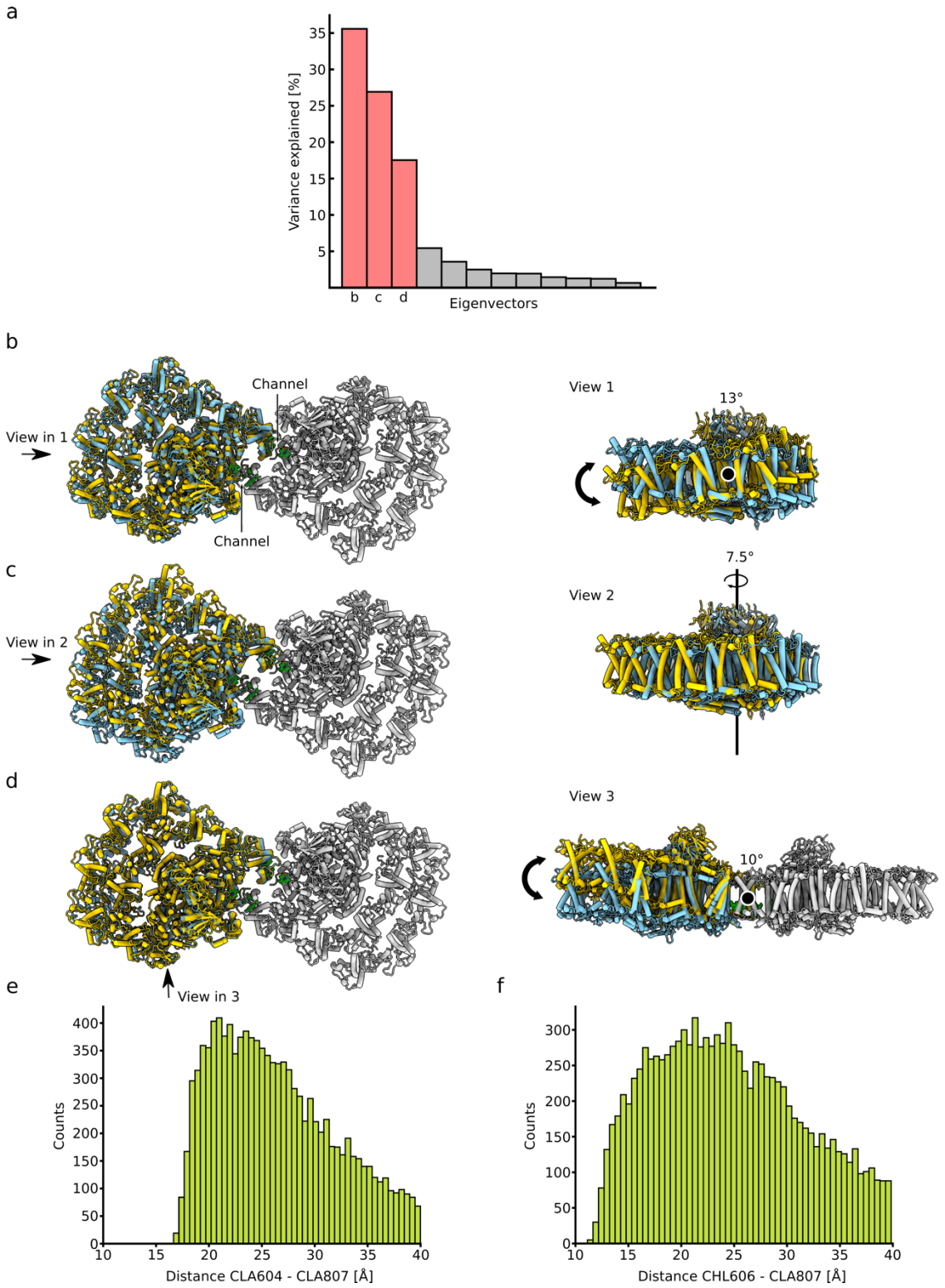
Extended Data Fig. 5: Lhca9 short AC-loop contributes to the dimerization. a, Superposition of Lhca8 and Lhca9, highlighting the difference in the AC-loop. **b**, The 90°-rotated view illustrates a clash between the longer AC-loop version and PsaB of monomer 2. Clashes between atoms with a VdW overlap >0.6 Å are highlighted by the dashed lines.



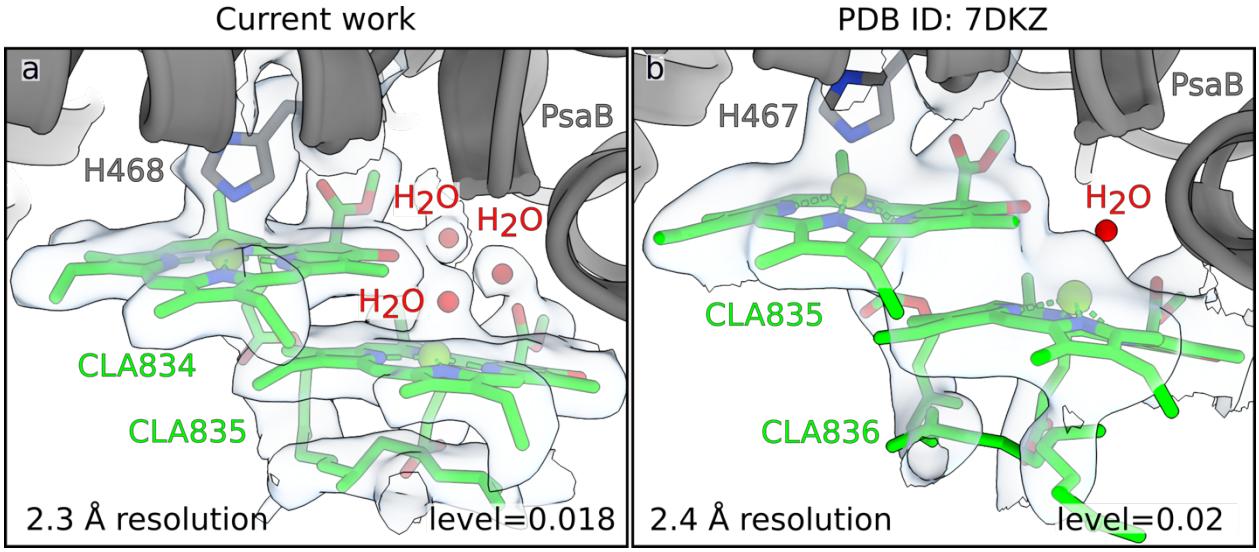
Extended Data Fig. 6: PSI dimer fraction is absent in *Δlhca9*. a, Sucrose density gradient of affinity purified PSI from *lhca9* insertional mutant (~ 60 µg chl). The monomer (M) fraction is indicated. **b,** Quantitative mass spectrometry analysis of the PSI monomer fraction from (**a**) compared to wild type (Extended Data Fig. 1). iBAQ values are normalized to the PSI core subunit PsaB.



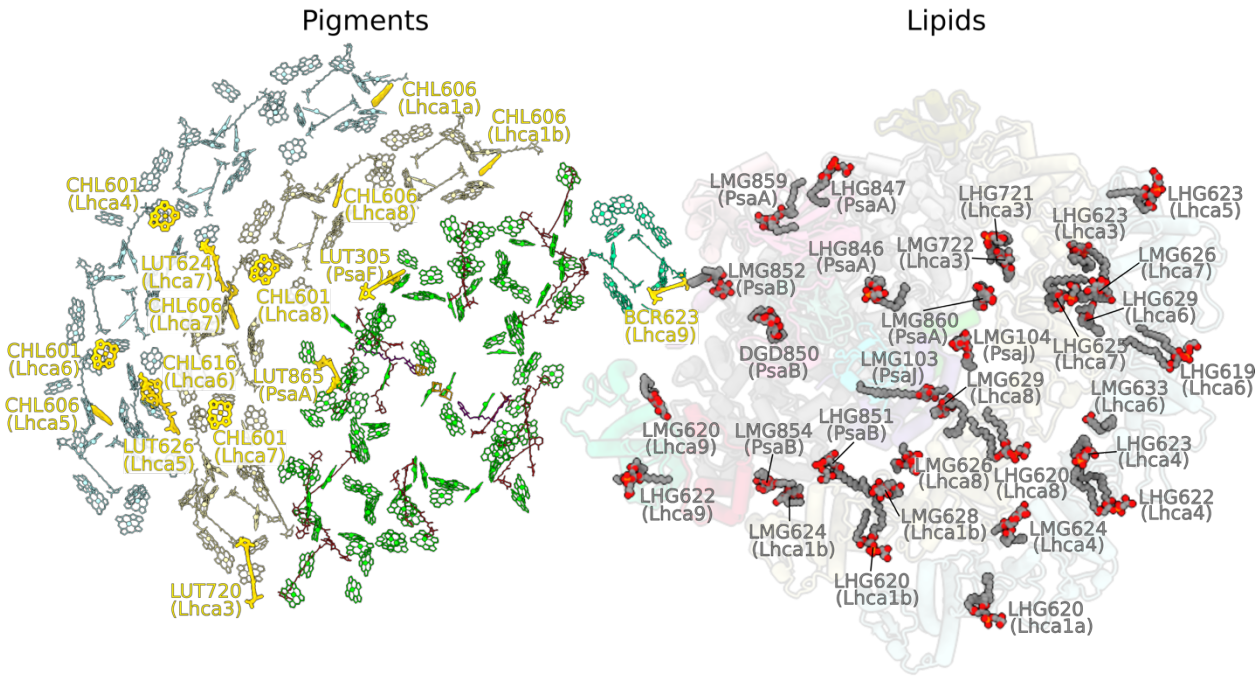
Extended Data Fig. 7: Sucrose density gradient analysis of non-tagged wild type, *Δlhca2* and *Δlhca9*. Sucrose density gradients of solubilized thylakoids isolated from wild type (**a**), *Δlhca2* (**b**) and *Δlhca9* (**c**) and quantitative mass spectrometry analysis of the labelled PSI fractions. iBAQ values are normalized to the PSI core subunit PsaB. 'M' is PSI monomer, 'PSI*' is a lower PSI band.



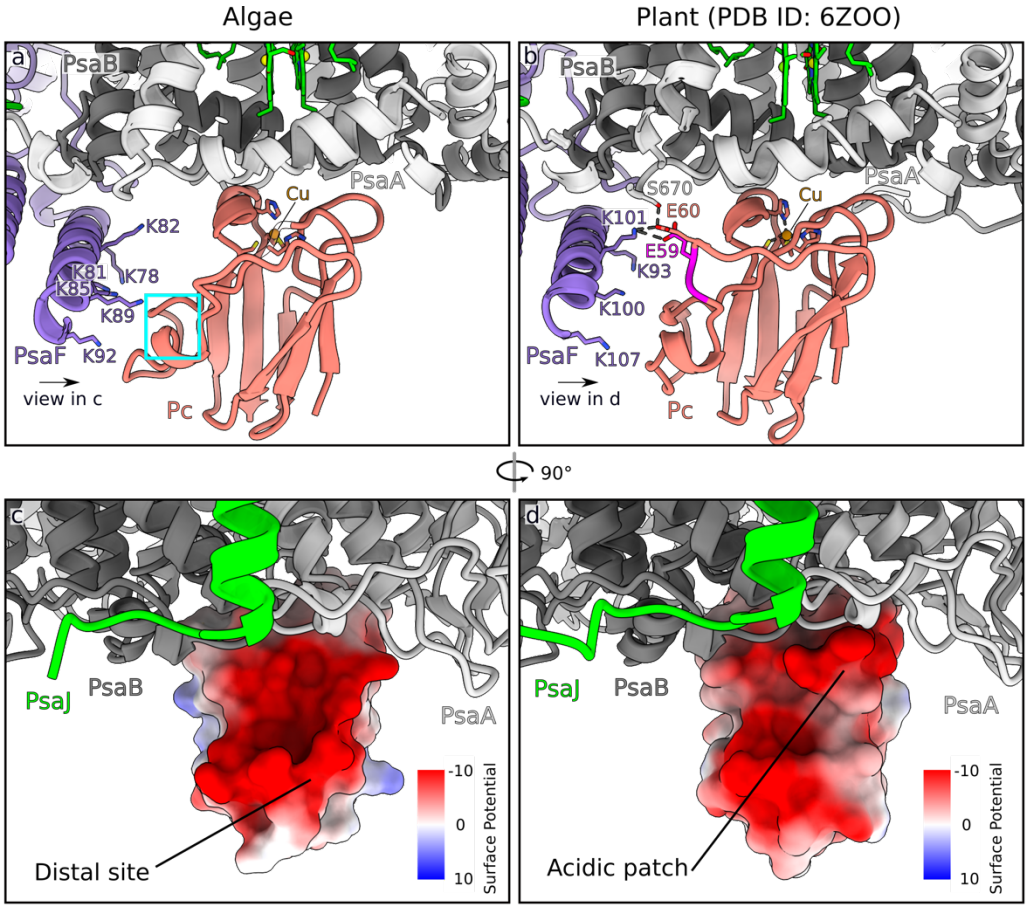
Extended Data Fig. 8: Multi-body refinement analysis. a, Eigenvectors that explain the variability of the data. The three major eigenvectors account for ~78% of the motion in the PSI dimer. **b-d**, Stromal and side view of the model, showing the maximal motion along the three vectors. **e, f**, The 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.



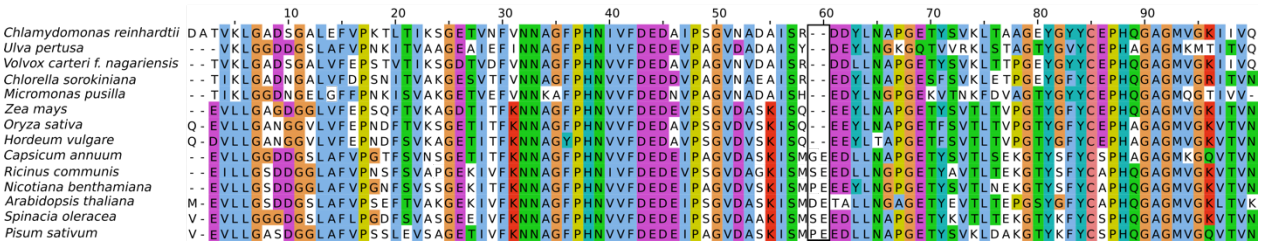
Extended Data Fig. 9: Comparison of the map/model quality with X-ray study. a, The current density around CLA834-835 is shown in iso-surface representation at a local resolution of 2.3 Å. Three water molecules involved in chlorophyll coordination are identified. Updated libraries are used for correct chlorophyll modeling. **b**, The same region from X-ray study at 2.4 Å resolution [17] lacks the density for waters, whereas the modeled water molecule has no density, and Mg atoms are not in the correct plane.



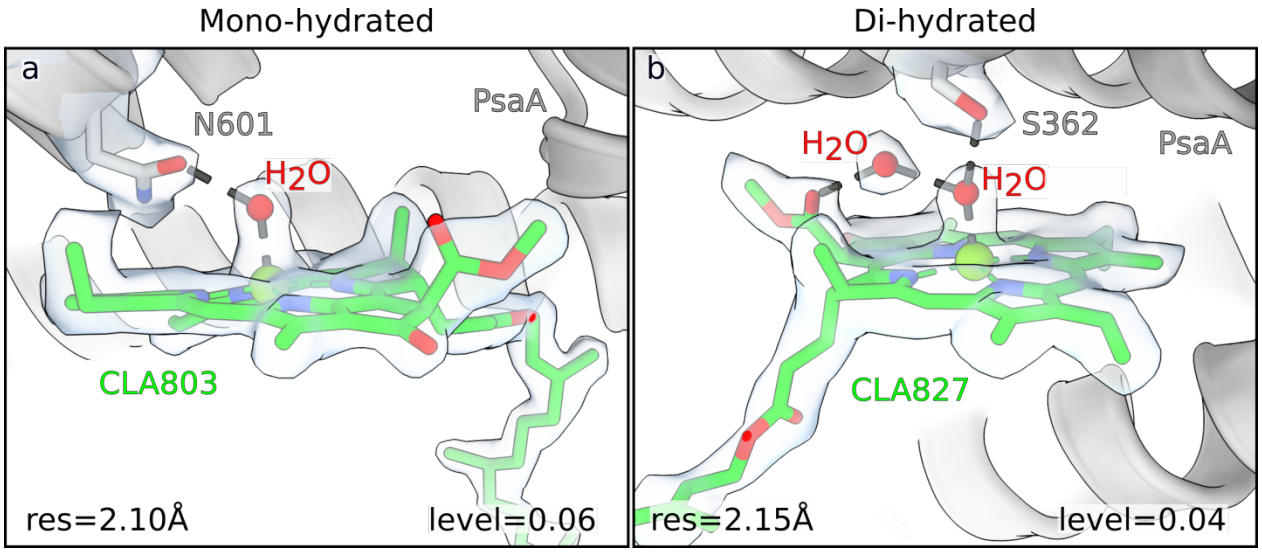
Extended Data Fig. 10: Overview of the pigments and lipids. Pigments (left) and lipids (right) are shown from stroma. The newly identified pigments are highlighted in bold gold.



e



Extended Data Fig. 11: Plastocyanin binding site. a, Pc binding in *C. reinhartdii* (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**, 90° 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 different species of the green lineage (algae and plants) showing that the inserted residues 58, 59 occur in a subset of plants and do not represent a general case.



Extended Data Fig. 12: Mono- and di-hydrated chlorophylls. a, An example of a mono-hydrated chlorophyll with the corresponding density. **b,** An example of a di-hydrated chlorophyll with the corresponding density. Local resolution and map levels are indicated.

715 **Supplementary Table 1: Protein composition of silver-stained spots from a 2D-PAGE of**
716 **solubilized *C. reinhardtii* wild type thylakoids isolated from control and anoxic conditions.**

Spot	Identified proteins
control	
1	PsaA, PsaB
2	β mannosidase, PsbB, PsbC
3	Asa4, AtpB, PsbA, PsbB, PsbC
4	Aaa1, AtpB, LhcbM3, PsbA, PsbB, PsbC
5	β mannosidase, Aaa1, ChlP, PsbB
6	PsaH
7	Lhcb5, LhcbM3, Nad5, PsbB, PsbC
8	LhcbM3, PsbA, PsbB, PsbC
9	LhcbM3, PsbB, PsbC
10	Asa3, Asa4, Atp3
11	Lhcb4, Lhcb5, PetA, PsbA, PsbB, PsbC, PsbD
12	Lhcb4, Lhcb5, LhcbM4, PsbA, PsbB, PsbD
13	Lhcb4, Lhcb5, PetA, PsbA, PsbB, PsbD
14	Lhcb4, Lhcb5, LhcbM3, LhcbM4, PsbA, PsbB, PsbD
15	Lhcb4, Lhcb5, LhcbM3, PsbA, PsbB, PsbD,
16	β mannosidase, Lhca3, Lhca5, Lhca6, Lhca7, LhcbM1
17	Lhcb4, Lhcb5, LhcbM3, LhcbM4, PsbA, PsbD
18	LhcbM1
19	LhcbM3, LhcbM4, LhcbM5, PsbA, PsbO
20	Asa2, Lhca3, Lhca4, Lhca6, Lhcb5, LhcbM2, PsbB, PsbC
anoxia	
1	Atp2, PsaA, PsaB
2	Asa2, PsbC
3	Lhcb5, LhcbM3, PsbA, PsbD
4	Lhcb4, Lhcb5, PsbA, PsbD
5	LhcbM3, LhcbM4
6	LhcbM3, LhcbM4, PsaB, PsbO
7	Lhca1, Lhca3, Lhca4, Lhca5, Lhca6, Lhca7, Lhca8
8	Lhca3, Lhca5, Lhca6, Lhca7, Lhca8, Lhcb5, LhcbM1
9	Lhca3, Lhca5, Lhca6, Lhca7, Lhca8
10	Lhca3, Lhca7, Lhca8
11	Lhca1, Lhca7, Lhca8, Lhca9,
12	Lhca1, Lhca7, Lhca8, Lhca9, PsaD
13	Asa7, Lhca1, Lhca9, Lhcb4, PsaD, PsbO
14	PetC
15	Asa7, Lhca1, PsaB, PsaD
16	PsaF
17	Nuo13, PsaF
18	AtpG

19	Nuo3, NuoP5
20	PsaH
21	Asa6, PsaH
22	PsaH
23	PsaH

719 **Supplementary Table 2: Cryo-EM data collection, refinement and validation statistics of**
720 **PSI monomer and dimer of *C. reinhartdii***

	Dimer (EMDB-xxxx) (PDB xxxx)	Symmetry expanded dimer (EMDB-xxxx) (PDB xxxx)	Monomer (EMDB-xxxx) (PDB xxxx)	Plastocyanin (EMDB-xxxx) (PDB xxxx)
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 particle 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)	2.97	2.74	2.31	2.55
FSC threshold 0.143				
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	50,370	729
Protein residues	8,418	4,712	4,164	98
Ligands	708	398	352	0
Waters	160	88	515	0
B factors (Å^2) (min/max/mean)				
Protein	22.5/125.1/65.5	48.7/154.0/80.3	13.9/117.2/39.9	66.7/83.8/73.5
Ligand	25.5/118.9/66.9	51.0/148.5/79.45	15.9/109.5/39.58	
Waters	25.4/115.3/64.0	1.1/95.5/72.0	13.4/57.5/32.1	
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	1.15	1.20
Clashscore	3.70	3.69	3.36	4.21
Poor rotamers (%)	0.39	0.35	0.24	0.00
Ramachandran plot				
Favored (%)	98.40	98.34	97.89	98.96
Allowed (%)	1.57	1.61	2.11	1.04
Disallowed (%)	0.02	0.04	0.00	0.00

721

722 **Supplementary Table 3: Chemical moieties involved in coordination of chlorophylls (per**
723 **monomer).**

Chlorophyll *a*										**Total: 199**
	His	Met	1x H_2O	2x H_2O	Gln	Back-bone	Lipid	Asp	Glu	Asn
All subunits	84	2	13	22	13	11	10	5	28	11
PsaA	32	1	3	4	3	1	1			
PsaB	31	1	5		1		1	1		
PsaF				2				1		
PsaG	1							1		
PsaJ									1	
PsaK	1			1		1		1		
PsaL	1		1							
Lhca1_b	2			3		1	1		3	1
Lhca1_a	2		1	2		1	1		3	1
Lhca3	2		1	2	2	2			3	1
Lhca4	1			1	1		1	1	3	2
Lhca5	3			2	1	2	1		3	1
Lhca6	3		1	1	1		1		3	1
Lhca7	2		1	2	1	1	1		3	1
Lhca8	2			2	1	1	1		3	1
Lhca9	1				2	1	1		3	2

724

Chlorophyll *b*										**Total: 30**
	His	Met	1x H_2O	2x H2O	Gln	Back-bone	Lipid	Asp	Glu	Asn
All subunits			1	19		7		3		
Lhca1_b				2		1				
Lhca1_a			1	1		1				
Lhca3				1						
Lhca4				3		1		1		
Lhca5				3				1		
Lhca6				3		2		1		
Lhca7				2		1				
Lhca8				2		1				
Lhca9				2						

725