

# SCIENTIFIC REPORTS



OPEN

## Pea PSII-LHCII supercomplexes form pairs by making connections across the stromal gap

Pascal Albanese<sup>1,2</sup>, Roberto Melero<sup>3</sup>, Benjamin D Engel<sup>4</sup>, Alessandro Grinzato<sup>5</sup>, Paola Berto<sup>5</sup>, Marcello Manfredi<sup>6,7</sup>, Angelica Chiodoni<sup>8</sup>, Javier Vargas<sup>3</sup>, Carlos Óscar Sánchez Sorzano<sup>3</sup>, Emilio Marengo<sup>7</sup>, Guido Saracco<sup>8</sup>, Giuseppe Zanotti<sup>5</sup>, Jose-Maria Carazo<sup>3</sup> & Cristina Pagliano<sup>1</sup>

In higher plant thylakoids, the heterogeneous distribution of photosynthetic protein complexes is a determinant for the formation of grana, stacks of membrane discs that are densely populated with Photosystem II (PSII) and its light harvesting complex (LHCII). PSII associates with LHCII to form the PSII-LHCII supercomplex, a crucial component for solar energy conversion. Here, we report a biochemical, structural and functional characterization of pairs of PSII-LHCII supercomplexes, which were isolated under physiologically-relevant cation concentrations. Using single-particle cryo-electron microscopy, we determined the three-dimensional structure of paired C<sub>2</sub>S<sub>2</sub>M PSII-LHCII supercomplexes at 14 Å resolution. The two supercomplexes interact on their stromal sides through a specific overlap between apposing LHCII trimers and via physical connections that span the stromal gap, one of which is likely formed by interactions between the N-terminal loops of two Lhcb4 monomeric LHCII subunits. Fast chlorophyll fluorescence induction analysis showed that paired PSII-LHCII supercomplexes are energetically coupled. Molecular dynamics simulations revealed that additional flexible physical connections may form between the apposing LHCII trimers of paired PSII-LHCII supercomplexes in appressed thylakoid membranes. Our findings provide new insights into how interactions between pairs of PSII-LHCII supercomplexes can link adjacent thylakoids to mediate the stacking of grana membranes.

Photosystem II (PSII) is a multisubunit pigment-protein complex that is embedded in the thylakoid membranes of oxygenic photosynthetic organisms. It uses solar energy to catalyse the splitting of water into dioxygen, protons and electrons, thus providing the molecular oxygen and chemical energy that sustain most of the life on Earth<sup>1</sup>.

PSII forms supercomplexes that are composed of two moieties: a central PSII core containing the catalytic reaction center and a peripheral antenna system responsible for harvesting light and transferring energy to the reaction center. The PSII core, occurring mainly in dimeric form<sup>2–4</sup>, is highly conserved among all oxygenic photosynthetic organisms, with the exception of the extrinsic proteins of the oxygen evolving complex (PsbO, PsbV and PsbU in cyanobacteria and diatoms; PsbO, PsbP and PsbQ in plants and green algae)<sup>5</sup>. Conversely, the peripheral antenna system is evolutionarily divergent; it is composed of extrinsic phycobilisomes in cyanobacteria, while it consists of intrinsic LHCII light harvesting complexes in green algae and plants<sup>5,6</sup>. The evolution of different antenna systems was accompanied by the differentiation of thylakoid membranes into functionally

<sup>1</sup>Applied Science and Technology Department–BioSolar Lab, Politecnico di Torino, Viale T. Michel 5, 15121, Alessandria, Italy. <sup>2</sup>Department of Biology, University of Padova, Via Ugo Bassi 58 B, 35121, Padova, Italy.

<sup>3</sup>Biocomputing Unit, Centro Nacional de Biotecnología–CSIC, Darwin 3, Cantoblanco, 28049, Madrid, Spain.

<sup>4</sup>Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, 82152, Martinsried, Germany.

<sup>5</sup>Department of Biomedical Sciences, University of Padova, Via Ugo Bassi 58 B, 35121, Padova, Italy. <sup>6</sup>ISALIT–

Department of Science and Technological Innovation, University of Eastern Piedmont, Viale T. Michel 11, 15121, Alessandria, Italy. <sup>7</sup>Department of Science and Technological Innovation, University of Eastern Piedmont, Viale T. Michel 11, 15121, Alessandria, Italy. <sup>8</sup>Center for Sustainable Future Technologies – CSFT@POLITO, Istituto Italiano

di Tecnologia, Corso Trento 21, 10129, Torino, Italy. Correspondence and requests for materials should be addressed to C.P. (email: [cristina.pagliano@polito.it](mailto:cristina.pagliano@polito.it))

distinct appressed (stacked) and non-appressed (unstacked) regions, which in higher plants are called grana and stroma lamellae, respectively<sup>7</sup>.

Two related processes are believed to govern grana stacking: (i) interactions between adjacent membrane surfaces<sup>8</sup> and (ii) lateral segregation of photosystems within the thylakoid membrane plane<sup>9,10</sup>. At neutral pH, thylakoid membranes carry a net negative charge, and the maintenance of the grana stacks requires shielding cations (e.g.,  $Mg^{2+}$ ,  $K^+$  and  $Na^+$ )<sup>11,12</sup>. Indeed, it has long been known that exposure of thylakoid membranes to low-ionic strength buffers induces grana unstacking, and the readdition of low concentrations of divalent cations ( $\geq 5$  mM  $Mg^{2+}$ ) or higher concentrations of monovalent cations ( $\geq 150$  mM  $K^+$  or  $Na^+$ ) results in spontaneous restacking<sup>11</sup>. Photosynthetic membrane proteins are segregated into different thylakoid domains; grana regions contain mainly PSII and LHCII, complexes with flat stromal surfaces that do not project into the narrow stromal gap between thylakoids. Conversely, stroma lamellae and grana end membranes accommodate Photosystem I (PSI), with its light harvesting antenna complex (LHCI), and ATP synthase (ATP-ase), which have bulky stromal protrusions. Unlike the photosystems, cytochrome  $b_6/f$  (Cyt  $b_6/f$ ) is evenly distributed between stacked and unstacked thylakoid regions<sup>13,14</sup>. This lateral heterogeneity between PSI and PSII is thought to be intimately linked to membrane appression, allowing higher plants to optimize photosynthesis in ever-changing light conditions<sup>15</sup>.

Grana formation seems to be driven by interactions between the stromal domains of LHCII that face each other from adjacent stacked membranes<sup>16,17</sup>. However, grana with apparently normal stacked architecture form in *Arabidopsis* mutants that are virtually devoid of LHCII<sup>18</sup>, suggesting that other mechanisms are involved in grana formation<sup>19</sup>. In addition, it has been proposed that the ordered assembly of semi-crystalline 2D arrays of PSII-LHCII supercomplexes in grana thylakoids might promote grana stacking by mediating specific contacts between adjacent membrane discs<sup>20</sup>. While the lateral heterogeneity of thylakoid membrane proteins appears to drive grana staking, there seems to be a reciprocal relationship where thylakoid architecture enforces this lateral heterogeneity; the unstacking of grana with low-ionic strength buffer causes the random redistribution of photosystems within the lateral thylakoid membrane plane as well as the detachment of some LHCII from supercomplexes<sup>21,22</sup>.

In plant grana thylakoids, the PSII core associates with a variable number of LHCII antennas to form different types of PSII-LHCII supercomplexes<sup>14,23,24</sup>. The  $C_2S_2$  supercomplex consists of a dimeric PSII core ( $C_2$ ), which strongly binds two LHCII trimers (S-trimers) via two copies of the monomeric Lhcb4 and Lhcb5 subunits. Larger  $C_2S_2M_{1-2}$  supercomplexes contain two copies of monomeric Lhcb6, with one or two additional LHCII trimers (M-trimers) bound with moderate strength to the dimeric PSII core via Lhcb4 and Lhcb6<sup>14</sup>.

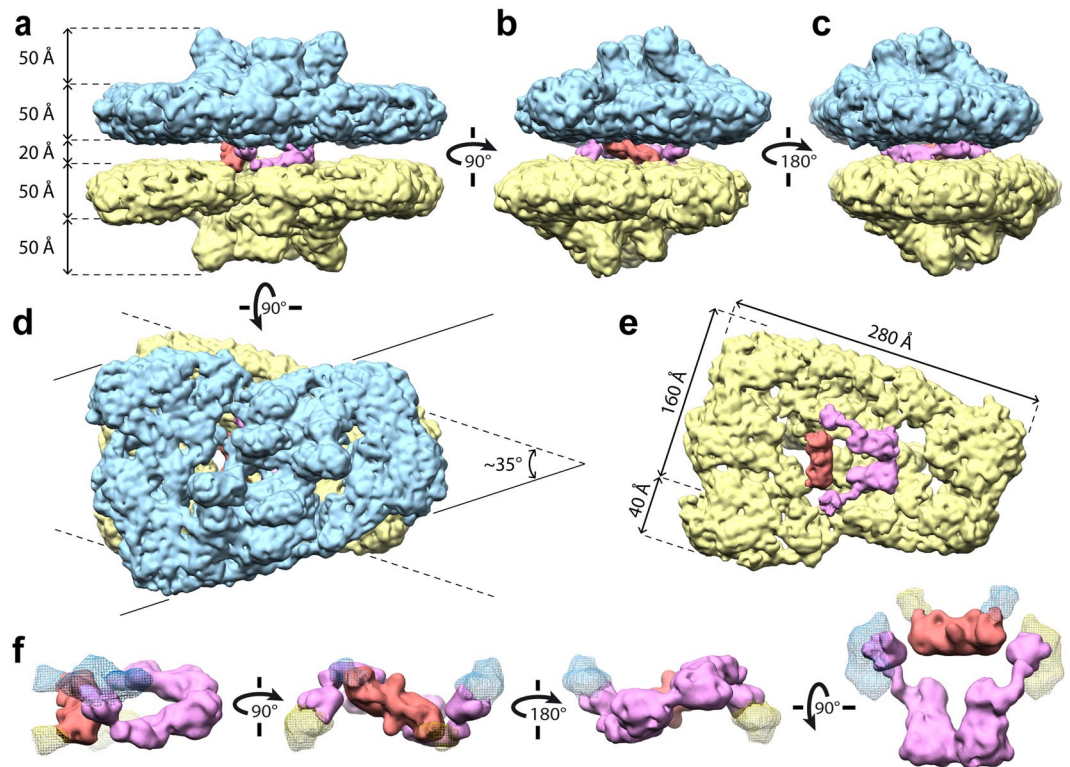
Two 3D maps of the  $C_2S_2$  PSII-LHCII supercomplex isolated from spinach have been obtained by single-particle cryo-electron microscopy (cryo-EM); the first structure was generated 17 years ago at 17 Å resolution<sup>25</sup>, while the second structure was recently solved at 3.2 Å resolution<sup>26</sup>. The new structure's significant improvement in resolution allowed the precise positioning of the pigments bound to the PSII-LHCII supercomplex. Thus, there is now a clear structural description for the energy transfer from the peripheral LHCII antennas to the PSII reaction center. The next step for the mechanistic understanding of plant photosynthesis is to expand this structural description to the organization and coordination of multiple PSII-LHCII supercomplexes within the grana stack. Here, we pursue this goal with a structural and functional description of paired  $C_2S_2M$  PSII-LHCII supercomplexes that can link adjacent grana membranes.

## Results and Discussion

**Isolation of paired PSII-LHCII supercomplexes.** The differentiation of thylakoid membranes into grana and stroma lamellae is a ubiquitous feature of higher plant chloroplasts<sup>7</sup>. Even though the number and diameter of discs within grana stacks may vary depending on fluctuating environmental conditions<sup>27-29</sup>, different plant species grown under the same illumination regime can differ intrinsically in their grana organization<sup>30</sup>. We used pea plants as starting material for PSII-LHCII supercomplex isolation because their chloroplasts contain large, highly stacked grana<sup>30</sup>.

In order to investigate how PSII-LHCII supercomplexes interact with each other within grana stacks, thylakoids were isolated from pea leaves in the presence of divalent cation concentrations (i.e., 5 mM  $Mg^{2+}$ ) that resemble the native chloroplast ionic environment<sup>31</sup>, preserving the stacked morphology of the grana membranes. PSII-LHCII supercomplexes were subsequently isolated by a quick (1 min) direct solubilization of stacked thylakoids with the mild detergent n-dodecyl- $\alpha$ -D-maltoside ( $\alpha$ -DDM) followed by sucrose density gradient ultracentrifugation in the presence of the same concentration of divalent cations, according to an optimized protocol previously described<sup>32</sup>. Physiological cation concentrations were maintained throughout the entire purification, and either  $Ca^{2+}$  or  $Mg^{2+}$  cations in the sucrose gradient buffer produced the same isolation profile (data not shown). It is important to point out that our isolation procedure greatly differs from that used by Wei *et al.*<sup>26</sup> to produce the recent cryo-EM structure of the unpaired  $C_2S_2$  plant supercomplex. In their study, grana membranes were first isolated by solubilizing thylakoids with Triton X-100 and then unstacked by washing with a metal chelating agent (i.e., EDTA), followed by further solubilization with  $\alpha$ -DDM and ultracentrifugation. The buffers were devoid of salt throughout the purification, and thus the resulting structure lacks the native interactions between supercomplexes found within stacked grana.

**Cryo-EM structure of paired  $C_2S_2M$  supercomplexes at 14 Å resolution.** To investigate the structure of the isolated PSII-LHCII supercomplexes, the detergent-solubilized PSII-LHCII particles were frozen in vitreous ice and then imaged by transmission electron microscopy using a direct electron detector. From 6,834 micrographs collected (Supplementary Fig. 1a), 33,729 particles were manually picked for further data processing. After several image sorting steps and iterative rounds of 2D classification (Supplementary Fig. 1b), 6,776 particles corresponding to contaminants and blurred or broken complexes were discarded. The initial model (Supplementary Fig. 1c), built from 44 representative 2D classes, was used for the 3D classification of 26,953 particles, resulting in

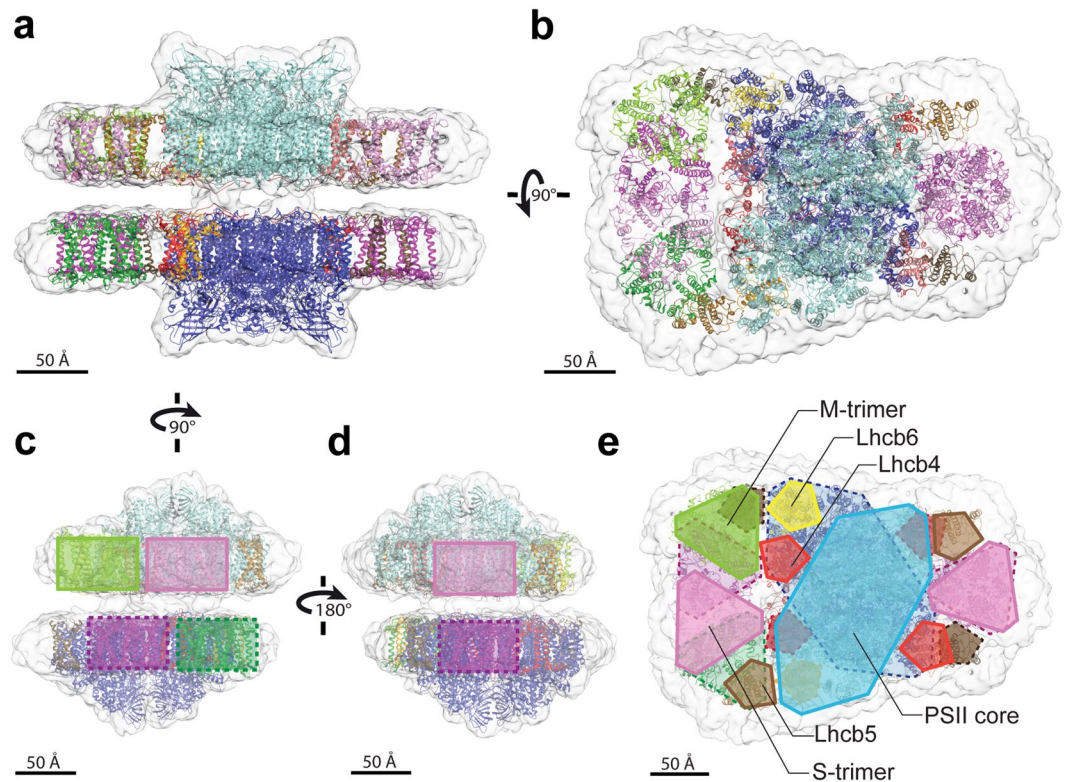


**Figure 1.** Three-dimensional cryo-EM density map of paired  $C_2S_2M$  PSII-LHCII supercomplexes. (a–c) Side view (a) and end views (b,c) within the membrane plane. (d) Top view towards the luminal surface of the paired supercomplexes, rotated  $90^\circ$  from panel (a). Solid and dashed lines extending from the top (cyan) and bottom (yellow) supercomplexes, respectively, illustrate the  $\sim 35^\circ$  rotational offset between the supercomplexes. (e) The same view as panel (d) with the top supercomplex removed, revealing the top view of the lower supercomplex's stromal surface from the perspective of the stromal gap. The stromal connections are also displayed. (f) Structural features of the “hinge” (magenta) and “knot” (red) stromal connections shown in different orientations starting on the left with the same side view as panel (a). Contact sites to the upper and lower supercomplexes are indicated with cyan and yellow mesh densities, respectively.

three classes: paired  $C_2S_2M$  supercomplexes ( $\sim 53\%$ ), unpaired  $C_2S_2M$  supercomplexes ( $\sim 21\%$ ) and paired  $C_2S_2$  supercomplexes ( $\sim 26\%$ ) (Supplementary Fig. 1d). From this classification, the majority of the isolated particles were found to be in a paired conformation ( $\sim 79\%$ ), with the  $C_2S_2M$  form ( $\sim 72\%$ ) more abundant than the  $C_2S_2$  ( $\sim 28\%$ ). Most particles contained LHCII M-trimers ( $\sim 74\%$ ), irrespective of their paired or unpaired behaviour. This finding agrees with previous results showing that  $C_2S_2M$  is the predominant supercomplex in the thylakoid membranes of pea plants grown in moderate light intensity<sup>33</sup>, the same light condition used in this study.

Paired  $C_2S_2M$  particles, representing the most abundant 3D class and the largest isolated supercomplexes, were further analysed (Supplementary Fig. 1e,f). Several refinement cycles led to a final 3D cryo-EM map at an overall resolution of  $14 \text{ \AA}$  (Fig. 1a–d), as judged by the “gold standard” Fourier shell correlation (FSC = 0.143) criterion<sup>34</sup> (Supplementary Fig. 2). The 3D map depicts two  $C_2S_2M$  supercomplexes with their stromal surfaces facing each other. The supercomplexes do not completely overlap, but rather have a rotational offset of  $\sim 35^\circ$  around the membrane plane's normal vector (Fig. 1d). The 3D map of each  $C_2S_2M$  supercomplex, including the detergent shell, is  $280 \text{ \AA}$  along its long-axis in the membrane plane,  $200 \text{ \AA}$  along the M-trimer side and  $160 \text{ \AA}$  along the S-trimer side (Fig. 1e). The thickness of the supercomplex is  $\sim 50 \text{ \AA}$  at its periphery, while the central protrusions extend for another  $\sim 50 \text{ \AA}$  on the luminal side (Fig. 1a). The central regions of the paired supercomplexes are connected on their stromal sides by two distinct densities that we have termed the “hinge” and “knot” as a description of their shapes (Fig. 1e,f). The  $\sim 20 \text{ \AA}$  stromal distance between the paired supercomplexes is consistent with reported stromal gap values for grana thylakoids, which range from  $2 \text{ nm}$ <sup>14</sup> to  $3.2\text{--}3.6 \text{ nm}$ <sup>20,35</sup>. In total, the supercomplex pair has a thickness of  $220 \text{ \AA}$ , which is compatible with the *in situ* stacking periodicity measured for intact pea chloroplasts<sup>20</sup>.

**Model fitting reveals different degrees of overlap between facing LHCII S- and M-trimers.** To interpret our 3D cryo-EM map, we performed rigid-body fitting using the atomic coordinates from two structures: the  $3.2 \text{ \AA}$  cryo-EM structure of the unpaired  $C_2S_2$  PSII-LHCII supercomplex from spinach<sup>26</sup> (PDB ID: 3JCU) for modeling the central dimeric PSII core together with the monomeric Lhcb4 and Lhcb5 subunits, and the  $2.5 \text{ \AA}$  crystal structure of the LHCII trimer from pea<sup>36</sup> (PDB ID: 2BHW) for modeling the LHCII S- and M-trimers. The atomic coordinates of the PSII dimeric core, the LHCII trimers, and the monomeric Lhcb4 and Lhcb5 fit well into the cryo-EM density map (Fig. 2). While modeling the PSII dimeric core, we did not fit the



**Figure 2.** Fitting the cryo-EM density map of paired  $C_2S_2M$  supercomplexes with high-resolution structures. Side view within the membrane plane (a) and top view towards the luminal surface (b) of the paired  $C_2S_2M$  supercomplexes. The following structures were placed into the cryo-EM map by rigid-body fitting: the PSII dimeric core and monomeric Lhcb4 and Lhcb5 from spinach<sup>26</sup> (PDB: 3JCU depleted of subunits PsbP, PsbQ, PsbTn; upper PSII dimer in cyan, lower PSII dimer in blue, upper Lhcb4 in pale red, lower Lhcb4 in dark red, upper Lhcb5 in pale brown, lower Lhcb5 in dark brown), the LHCII trimer from pea<sup>36</sup> (PDB: 2BHW; upper S-trimers in pink, lower S-trimers in violet, upper M-trimer in pale green, lower M-trimer in dark green), the predicted structure of monomeric Lhcb6 from pea generated by the PHYRE2 algorithm<sup>39</sup> (upper Lhcb6 in yellow, lower Lhcb6 in orange). (c–e) Schematic representations showing the positions of the LHCII trimers (c,d) and of all fitted supercomplex components (e), superimposed on the cryo-EM density map, showing end views within the membrane plane (c,d) and a top view towards the luminal surface (e). Colors match the structures in panels (a,b); solid lines for upper supercomplex components, dashed lines for lower supercomplex components.

extrinsic subunits PsbP, PsbQ and PsbTn, since previous biochemical and proteomic characterization showed that PsbQ and PsbTn are absent from the supercomplexes and PsbP is present in sub-stoichiometric amounts relative to the core subunits<sup>32,37</sup>. Although a higher-resolution crystal structure of spinach Lhcb4 (2.8 Å) is available<sup>38</sup>, it lacks 73 amino acid residues at the N-terminus, which were nearly all detected in our sample by mass spectrometry (Supplementary Table 1). Thus, we decided to fit our cryo-EM map with the more complete structure of Lhcb4 obtained by Wei *et al.*<sup>26</sup>. An atomic structure is not yet available for the monomeric Lhcb6 antenna, which links the LHCII M-trimer to the CP47 subunit of the PSII core. To model Lhcb6 within our cryo-EM map, we used the 3D structure predicted by the PHYRE2 program<sup>39</sup> for the *P. sativum* Lhcb6 protein sequence, which showed 50% coverage in the mass spectrometry analysis of our sample (see Supplementary Table 1).

Despite the limited resolution of our 3D map, fitting high-resolution components of the PSII-LHCII supercomplex clearly showed that LHCII trimers face each other across the stromal gap with various degrees of overlap. On the side of the paired supercomplexes containing only S-trimers, the apposing LHCII trimers completely overlapped and were in close proximity to the core complex, with the same position observed in the structure by Wei *et al.*<sup>26</sup> (Fig. 2b,d). Conversely, on the opposite side of the paired supercomplexes containing additional M-trimers, the S-trimers were more distantly positioned with respect to the core complex if compared to the structure by Wei *et al.*<sup>26</sup>, and the apposing S- and M-trimers only partially overlapped (Fig. 2b,c). The overlaps observed between pairs of facing LHCII trimers, including S- with S- and S- with M- trimers, were specific and closely resembled two arrangements observed in the crystal structure of multilayer packed LHCII from spinach reported by Wan *et al.*<sup>40</sup>. The crystal packing of LHCII in this study, obtained with a physiologically-relevant cation concentration and in the presence of native membrane lipids, may replicate the stacking of LHCII in native grana membranes. In these crystals, cations are sandwiched between the neighbouring LHCII layers to screen the surface charges and mediate interactions by forming strong salt bridges between highly conserved negatively

charged residues on facing LHCII trimers. Although we did not resolve physical contacts between apposing LHCII trimers in our cryo-EM map (Fig. 1a), the conservation of these negatively charged residues in the Lhcb1 and Lhcb2 subunits was confirmed in our preparation (Supplementary Fig. 3a, Supplementary Table 1). This evidence provides a plausible explanation for the complete overlap observed between two S-trimers (Fig. 2b,d), which are composed of Lhcb1 and Lhcb2<sup>3</sup>, and thus have affinity for each other across the stromal gap due to salt bridge formation. It also helps explain the partial overlap between S- and M-trimers on the other side of the paired supercomplexes (Fig. 2b,c); the M-trimer contains Lhcb3<sup>41</sup>, which may lack affinity for other LHCII subunits, thus reducing trimer overlap on that side.

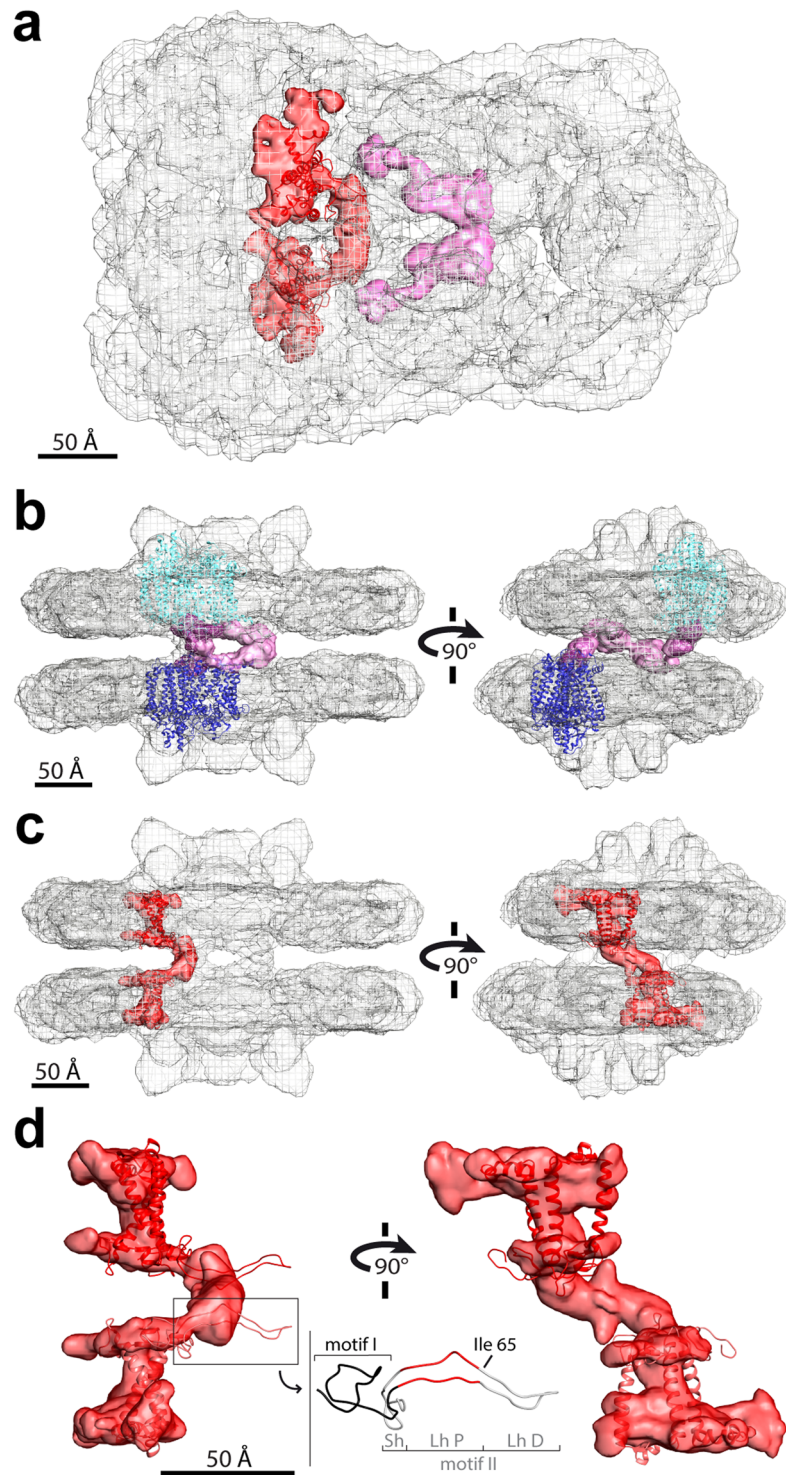
Our cryo-EM map provides clear evidence at intermediate resolution for interactions between apposing LHCII trimers within paired PSII-LHCII supercomplexes. In so doing, it bridges the resolution gap between previous reports of LHCII interactions from X-ray crystallography of isolated LHCII trimers<sup>40</sup> and cryo-electron tomography of C<sub>2</sub>S<sub>2</sub> supercomplex arrays embedded within intact grana membranes<sup>20</sup>.

**“Hinge” and “knot” densities make physical connections across the stromal gap.** Well-defined electron densities were visible in the stromal gap between the paired supercomplexes, strongly indicating the presence of two different centrally positioned physical connections with distinct shapes and specific connections to the two C<sub>2</sub>S<sub>2</sub>M supercomplexes.

Towards the side of the paired supercomplexes containing only S-trimers, we observed a dimer of bulky densities with a “hinge” shape, which made defined contacts to PSII core subunits in each of the facing supercomplexes (Figs 1f and 3a,b). Comparison to the Wei *et al.*<sup>26</sup> structure showed that this “hinge” density connects one PSII monomer in the upper supercomplex to one PSII monomer in the lower supercomplex, binding areas containing loops of the D1, CP43 and D2 proteins (Fig. 3b). Extensive biochemical and proteomic characterization of the sample used in this cryo-EM study<sup>32,37</sup> revealed the presence of an additional subunit of the PSII core that was not found in the Wei *et al.*<sup>26</sup> unpaired C<sub>2</sub>S<sub>2</sub> structure, a 10 kDa protein named PsbR. To date, the structure of PsbR is unknown and its positioning is still controversial. Although an extrinsic luminal location was hypothesized for PsbR based on its capacity to crosslink with the luminal PsbP subunit, PsbR’s hydrophobic C-terminus was suggested to be responsible for anchoring PsbR to the thylakoid membrane, leading to the protein’s insolubility<sup>42</sup>. Studies with *Arabidopsis psbR* knockout mutants revealed a reduction of the amount of PsbP bound to the PSII core and indicated a role for PsbR in stabilizing the supramolecular organization of PSII-LHCII supercomplexes<sup>43</sup>. However, it is not yet clear whether PsbR provides a direct docking site for PsbP or rather if it indirectly enhances the binding of PsbP on the luminal side by inducing a structural rearrangement in the PSII core. Taking all of this evidence into account, a possible interpretation of the “hinge” density in our cryo-EM map is that it is an extrinsic protein, potentially corresponding to PsbR. This protein, interacting with the stromal loops of the PSII core proteins, may induce a structural reorganization of PSII on the luminal side that favors the binding of PsbP, which stabilizes the association of LHCII with PSII within the membrane plane. In addition, by making connections that span the stromal gap between facing supercomplexes, the “hinge” can facilitate the stacking of adjacent grana thylakoids.

Towards the side of the paired supercomplexes containing the M-trimers, we observed a defined stromal density with a “knot” shape, forming an arm that extends diagonally across the stromal gap and contacts the supercomplexes on the opposite sides from the “hinge” binding sites (Figs 1f and 3a,c,d). This density is characterized by a symmetric and well-structured appearance, even when no D2 point group symmetry is imposed during image processing (data not shown). From its position in our fitted cryo-EM map, it is evident that the “knot” connection arises from the stromal surfaces of the Lhcb4 subunits, and thus can be assigned with high confidence to the N-terminal loops of two Lhcb4 subunits that face each other from two apposing supercomplexes (Fig. 3c,d). Lhcb4 contains a flexible ~90 amino acid N-terminal loop, easily lost during the crystallization process<sup>38</sup>, with dimensions that support its possible involvement in forming the “knot”. According to the structure by Wei *et al.*<sup>26</sup>, the N-terminal loop of Lhcb4 forms two motifs with irregular coil structures: motif I (Pro12–Lys41) and motif II (Pro42–Phe87) (Fig. 3d). Motif I forms a short loop near the stromal surface. Motif II contains a short hairpin loop (Ala73–Phe87) and a second hairpin loop that is approximately 40 Å long (Pro42–Ser72), extending into the stromal gap and then running nearly parallel to the stromal surface. In our structure, the connecting “knot” density that we attributed to the N-terminal loop of Lhcb4 confidently accommodates the proximal part of the motif II long hairpin formed approximately by Pro42–Gln47 and Ile66–Ser72. However, the distal part of the hairpin, corresponding approximately to Thr48–Ile65, protrudes from our 3D map (Fig. 3c,d). Spectroscopic analyses<sup>44</sup> revealed that Ile65 is highly mobile, likely serving as a point of flexibility in the loop, while the Thr48–Ile65 region contains a helical stretch and a random coil. It is therefore plausible that the distal part of the long hairpin is responsible for “tying the knot” by homodimerizing to form the structured area at the center of the connecting density between the two Lhcb4 subunits.

The stromal gap-spanning connection between apposing Lhcb4 subunits may be secured by cation-mediated ionic interactions through the formation of salt bridges. Although the N-terminus of Lhcb4 contains some positively-charged amino acids, it is characterized by an overall net negative charge (Supplementary Fig. 3b). In particular, the distal part of the motif II long hairpin (Thr48–Ile65) is enriched in negatively-charged amino acids (Supplementary Fig. 3b) that are highly conserved in higher plants (Supplementary Fig. 4). Cations may screen these surface charges, forming strong salt bridges between negatively charged residues that mediate the interaction between the ends of two facing hairpin loops. Conversely, the proximal Ile66–Ser72 portion of the long hairpin as well as the short hairpin (Ala73–Phe87) each contain a Thr residue that becomes phosphorylated under high-light conditions<sup>45</sup>, indicating that these regions may regulate the homo-interactions of Lhcb4. Taken together, these results suggest that the Lhcb4 N-terminal loop may assume multiple conformations. Upon environmental light changes, phosphorylation of Lhcb4 might induce a conformational change in its N-terminal loop that “unties the knot”, allowing the redistribution of Lhcb4 from PSII-LHCII supercomplexes to PSII dimers and



**Figure 3.** “Hinge” and “knot” connections cross the stromal gap of paired  $C_2S_2M$  supercomplexes. **(a)** Top view towards the luminal surface, showing the cryo-EM density map of paired  $C_2S_2M$  supercomplexes (grey mesh), with the stromal “hinge” density in magenta and the stromal “knot” density plus adjoining supercomplex region in red. Fitting of the high-resolution structure of Lhcb4 from spinach<sup>26</sup> (PDB 3JCU; chain r; upper Lhcb4 in lighter red, lower Lhcb4 in darker red) as in Fig. 2 is shown. **(b)** Side and end views within the membrane plane showing the connections between the paired supercomplexes and the “hinge” (connecting supercomplex density in violet mesh). The hinge connects one PSII monomer of the upper supercomplex to one PSII monomer of the lower supercomplex (PDB 3JCU: chains A, C and D, corresponding to the subunits D1, CP43 and D2; upper PSII core in cyan, lower PSII core in blue)<sup>26</sup>. **(c)** Side and end views within the membrane plane showing the connections between the paired supercomplexes and the “knot”, which links an Lhcb4 of the upper supercomplex with an Lhcb4 of the lower supercomplex. Fitting of the high-resolution structure of Lhcb4 from spinach<sup>26</sup> as in panel (a). **(d)** Enlarged views of the “knot” density plus adjoining supercomplex region fit with

two Lhcb4 subunits, as in panel (c). The inset shows the N-terminal loop of the Lhcb4 structure by Wei *et al.*<sup>26</sup>, composed of motif I (Pro12–Lys41) in black and motif II (Pro42–Phe87) in grey and red. Motif II is subdivided into the short hairpin (Sh, Ala73–Phe87, grey), the proximal part of the long hairpin that fits within the “knot” cryo-EM density (Lh P, Pro42–Gln47 and Ile66–Ser72, red) and the distal part of the long hairpin that protrudes from the “knot” density (Lh D, Thr48–Ile65, grey). The flexible Ile65 residue, which separates the proximal and distal segments of the long hairpin, is indicated.

monomers<sup>45</sup>, while potentially enabling dynamic modulation of PSII-LHCII supercomplex pairing and ultimately the degree of grana staking.

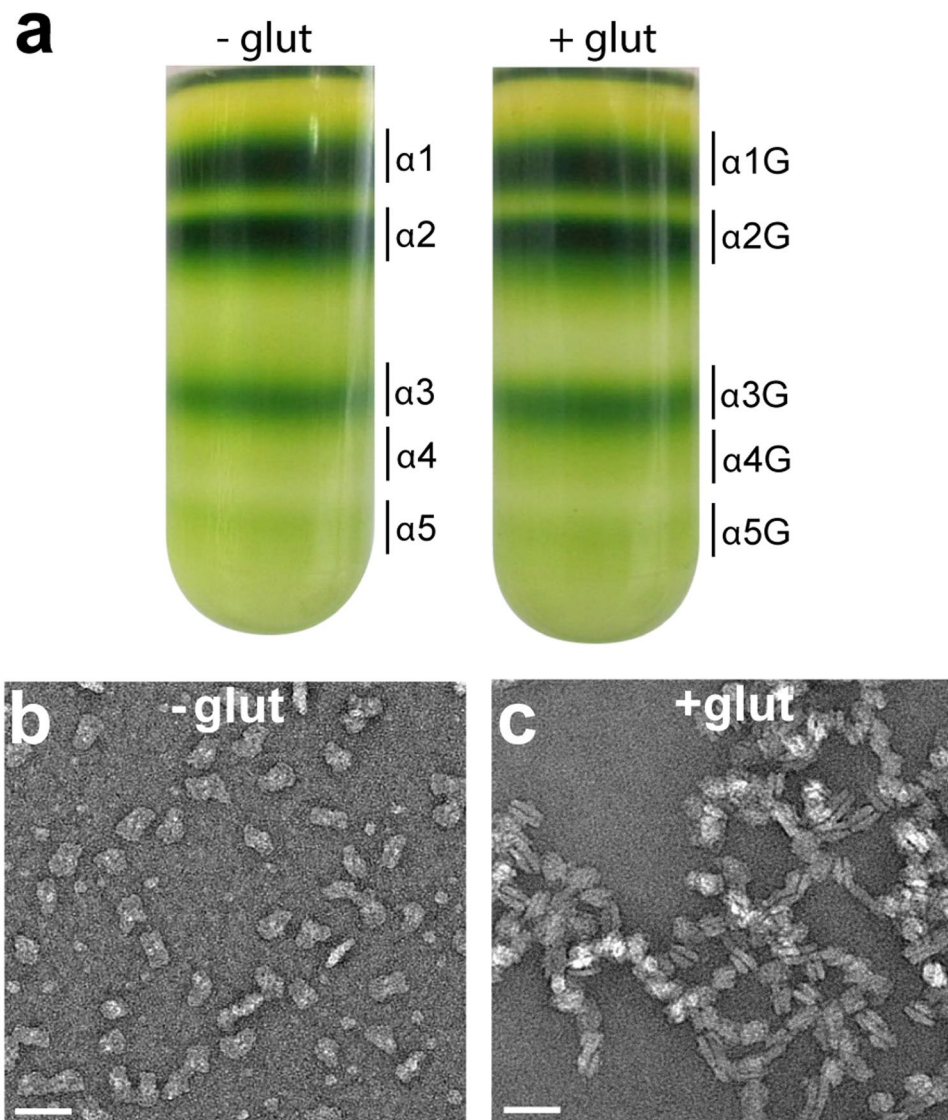
**Cations mediate the physical connection between paired PSII-LHCII supercomplexes.** Despite the well-established requirement of shielding cations for the maintenance of grana stacking<sup>11,12</sup>, the role of cations in PSII organization within grana remains unclear. Therefore, we assessed the potential function of cations in mediating the pairing of isolated PSII-LHCII supercomplexes. First, we isolated PSII-LHCII supercomplexes with sucrose density gradients that either contained (+glut) or lacked (-glut) glutaraldehyde, a cross-linking reagent often used to obtain information on protein interactions and the oligomerization state of complexes<sup>46,47</sup>. Similar density gradient profiles of solubilized thylakoids were obtained for both conditions (Fig. 4a). PSII-LHCII supercomplexes, present in bands  $\alpha 3$  and  $\alpha 3G$ , migrated to the same position on the sucrose gradient (Fig. 4a) and showed similar spectroscopic properties (Supplementary Fig. 5a). EM analysis, in cryo conditions for the  $\alpha 3$  particles (Supplementary Fig. 1a) and by negative stain for the  $\alpha 3G$  particles (Supplementary Fig. 5b), showed that both samples contained paired supercomplexes. Next, buffer exchange was used to remove salt from both the + glut and - glut samples. Negative stain EM revealed that the majority of the supercomplexes detached from their paired partner upon salt removal (Fig. 4b), while the glutaraldehyde fixation was effective in maintaining the stromal connections between paired supercomplexes (Fig. 4c). Native polyacrylamide gel electrophoresis (PAGE) of the isolated supercomplexes (Supplementary Fig. 5c) confirmed that + glut particles ( $\alpha 3G$ ) were in a paired oligomerization state, whereas - glut particles ( $\alpha 3$ ) were unpaired. These results strongly support the hypothesis that the stromal connections between apposing supercomplexes are mediated by ionic bonds, which can be easily lost by changing the ionic strength of the buffers used during the isolation procedure. Notably, previous studies that reported the purification of paired PSII-LHCII supercomplexes also used cations in their isolation buffers<sup>2,25</sup>. Recently, we observed a higher-order PSII-LHCII supercomplex oligomerization state consisting of two PSII-LHCII supercomplexes sitting side-by-side in the membrane plane (i.e., megacomplex), interacting with another two side-by-side supercomplexes across the stromal gap<sup>48</sup>. These paired megacomplexes correspond to band  $\alpha 5$  of the sucrose gradient shown in Fig. 4a. Native PAGE showed that the removal of cations disrupted the stromal connections between facing megacomplexes ( $\alpha 5$ ) unless these megacomplexes had been crosslinked with glutaraldehyde ( $\alpha 5G$ ) (Supplementary Fig. 5c). In contrast, the lateral interactions between two supercomplexes were maintained without crosslinking, suggesting a specific role of salts in mediating the stromal connections between supercomplexes in adjacent thylakoid membranes.

**Paired PSII-LHCII supercomplexes are energetically coupled.** Having assessed the requirement of cations for maintaining the structural connections between paired supercomplexes, we next sought to understand the function of supercomplex pairing by analyzing the fast chlorophyll *a* fluorescence induction curve<sup>49</sup>, which can be used to provide an estimation of the excitonic connectivity between PSII units<sup>50</sup>.

The fast phase of the chlorophyll *a* fluorescence induction curve is known as the OJIP transient, where O corresponds to the minimal fluorescence  $F_0$ , J and I are inflections, and P is the peak corresponding to the maximal fluorescence  $F_m$ . Three rise phases of the  $F_0$  to  $F_m$  kinetics are distinguishable: O–J (0–2 ms), J–I (2–30 ms), and I–P (30–1000 ms)<sup>49</sup>. In the analysis of the OJIP transient, the major inflection points of the fast fluorescence induction curve are used for the calculation of several parameters characterizing the structure and photochemical activity of a photosynthetic sample<sup>50</sup>.

OJIP curves were recorded for isolated PSII-LHCII supercomplexes that had been diluted 1,000-fold either in the standard sucrose gradient buffer containing divalent cations or in a similar buffer that lacked salts (Fig. 5a). Compared to supercomplex pairs, unpaired supercomplexes in salt-free buffer showed increased values of  $F_0$  and reduced values of  $F_m$ , with a concomitant strong reduction of the OJIP curve. The differences in OJIP curve shape and  $F_0/F_m$  ratio may reflect different redox states of the plastoquinone  $Q_A$  and differences in macromolecular organization. In particular, the rise of the  $F_0$  under low ionic conditions might be due to the disconnection of some LHCII from the supercomplexes, as suggested by negative stain EM that showed the detachment of LHCII trimers from PSII-LHCII particles upon salt removal (Fig. 4b).

Since the shape of the induction curve is influenced by excitation energy transfer between PSII units, commonly referred to as PSII connectivity<sup>51</sup> or grouping<sup>52</sup>, the sigmoidicity of the initial phase of the fast fluorescence transient can be used to estimate the degree of PSII connectivity<sup>51</sup>. To estimate the energetic PSII connectivity in paired and unpaired PSII-LHCII supercomplexes, we evaluated the O–J phase of the OJIP transient using the method developed by Strasser and Stirbel<sup>53</sup>. In this approach, the degree of the PSII connectivity can be measured and compared between different samples by estimating the sigmoidicity of the chlorophyll *a* fluorescence induction curve during the first few microseconds. For the first 300  $\mu$ s of the OJIP curves recorded in Fig. 5a, we calculated the normalized O–J phase (labelled W) and the theoretical exponential curve corresponding to an unconnected system (labelled  $W_E$ , for definitions and calculation of these parameters see Supplementary Table 2). The increase in sigmoidicity from the theoretical unconnected curve  $W_E$  to the experimental curve W was greater for the cation-containing sample than for the sample in salt-free buffer (Fig. 5b). The greater increase



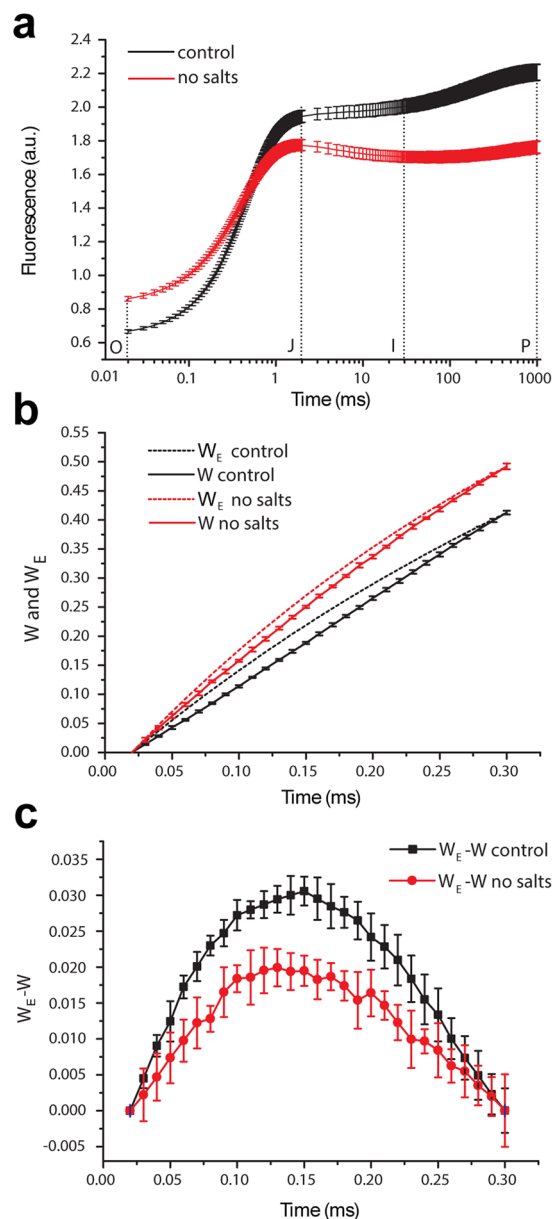
**Figure 4.** Salt removal causes unpairing of purified pea PSII-LHCII supercomplexes. **(a)** Isolation of PSII-LHCII supercomplexes from stacked pea thylakoid membranes by sucrose density gradients either containing (+ glut) or lacking glutaraldehyde (–glut). **(b,c)** Electron micrographs of uranyl acetate-stained PSII-LHCII supercomplexes contained in sucrose gradient bands  $\alpha_3$  **(b)** and  $\alpha_3G$  **(c)** upon salt removal. Scale bar, 50 nm.

in sigmoidicity for the paired PSII-LHCII supercomplexes was further emphasized by plotting the difference between  $W_E$  and  $W$  (Fig. 5c).

The O–J phase of the OJIP curve was also used to estimate the parameters of connectivity between PSII units<sup>49–53</sup> (for definitions see Supplementary Table 2). The O–J curvature parameter  $C$  and the connectivity among PSII units parameter  $p$  were over two times higher in paired PSII-LHCII supercomplexes ( $C \approx 0.35$ ,  $p \approx 0.30$ ) than in unpaired supercomplexes treated with salt-free buffer ( $C \approx 0.14$ ,  $p \approx 0.13$ ) (Table 1). In addition, the overall grouping probability  $p2G$  and the probability of connectivity between PSII units  $\omega$  (for definitions see Supplementary Table 2), indicated a significantly higher connectivity in paired PSII-LHCII supercomplexes ( $p2G \approx 0.21$ ,  $\omega \approx 0.20$ ) than in unpaired supercomplexes ( $p2G \approx 0.15$ ,  $\omega \approx 0.06$ ) (Table 1).

Despite the limitations of calculations based on OJIP curve measurements, these results support the conclusion that excitation energy migrates between supercomplex pairs that interact across the stromal gap. This suggests that energetic coupling is possible between two reaction center cores that are not members of the same PSII dimer. However, at the limited resolution of our cryo-EM map, we cannot deduce the precise route for this excitation energy diffusion. A plausible hypothesis is that energy crosses the  $\sim 20$  Å stromal gap via interactions between overlapping LHCII trimers, which may be precisely positioned to face each other by the more robust “knot” and “hinge” physical connections.

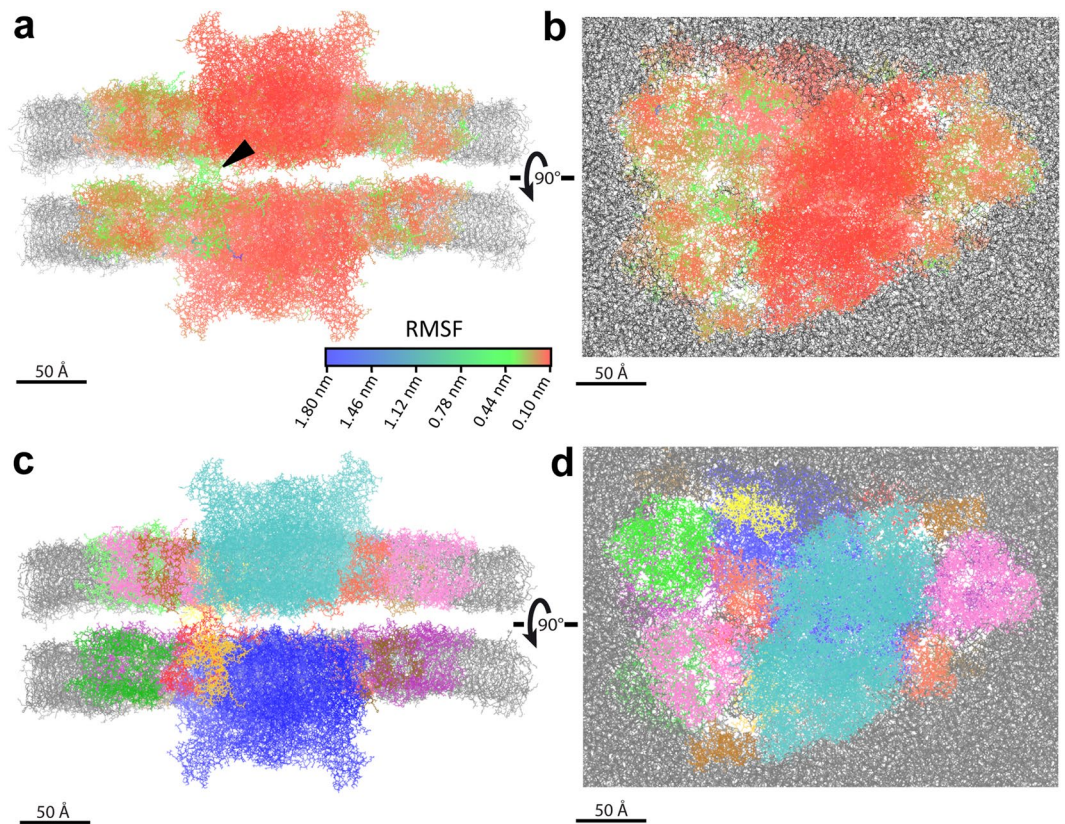




**Figure 5.** Estimation of the PSII excitonic connectivity in paired PSII-LHCII supercomplexes. (a) Chlorophyll *a* fluorescence induction curves (i.e., the OJIP transient) plotted on a logarithmic time scale (up to 1 s) for the paired PSII-LHCII supercomplexes prepared at  $1.25 \text{ mg mL}^{-1}$  in the sucrose gradient buffer and diluted 1,000-fold either in the sucrose gradient buffer (control, black line) or in a similar salt-free buffer (no salts, red line). The excitation light intensity was  $2,400 \mu\text{mol photons m}^{-2} \text{ s}^{-1}$ , with an emission peak at 630 nm. (b) The initial 300  $\mu$ s from the OJIP transient in panel (a) replotted on a linear time scale. The normalized O–J phase of the OJIP curve is labeled W (solid line), and the theoretical exponential curve corresponding to the unconnected system is labeled  $W_E$  (dashed line). For calculation of W and  $W_E$  see Supplementary Table 2. (c) Plot of  $W_E - W$  using the values from panel (b). Graphs display the mean values  $\pm$  standard deviations of six replicates.

	C	$p_{2G}$	p	$\omega$
control	$0.35 \pm 0.03$	$0.21 \pm 0.02$	$0.30 \pm 0.02$	$0.20 \pm 0.01$
no salts	$0.14 \pm 0.03$	$0.15 \pm 0.03$	$0.13 \pm 0.02$	$0.06 \pm 0.01$

**Table 1.** Selected parameters derived from fast fluorescence kinetic measurements to estimate the energetic connectivity between PSII units. Parameters for the paired PSII-LHCII supercomplexes prepared at  $1.25 \text{ mg mL}^{-1}$  in the sucrose gradient buffer and diluted 1,000-fold either in the sucrose gradient buffer (control) or in a similar salt-free buffer (no salts) were calculated according to methods previously reported<sup>49–53</sup> as defined in Supplementary Table 2.



**Figure 6.** Molecular dynamics simulations. **(a,b)** Root mean square fluctuations (RMSF) for the backbone atoms of the paired  $C_2S_2M$  PSII-LHCII supercomplexes at the end of the 7 ns molecular dynamics simulation. Supercomplexes are colored according to RMSF values, and thylakoid membranes are grey. Side view within the membrane plane **(a)** and top view towards the luminal surface **(b)** of the paired  $C_2S_2M$  supercomplexes, with interactions that match the position of the “knot” connection in the stromal gap indicated with a black triangle. **(c,d)** Corresponding side view **(c)** and top view **(d)** of the paired  $C_2S_2M$  supercomplexes colored by subunit identity as in the Supplementary Video and Fig. 2.

**Molecular dynamics simulations predict additional transient interactions between overlapping LHCII trimers.** To further investigate the interactions of paired  $C_2S_2M$  PSII-LHCII supercomplexes across the stromal gap, we ran molecular dynamics simulations using an atomic model generated from the fit of our cryo-EM density map (Fig. 2). After running the simulation for 7 ns, the backbone atoms had an average root mean square fluctuation (RMSF) of 0.23 nm compared to the reference time-averaged structure, with the central PSII core showing higher rigidity than the peripheral LHCII antennae system (Fig. 6 and Supplementary Fig. 6). Within several ps of starting the simulation, visible connections formed between the two apposing supercomplexes embedded within the appressed thylakoid membranes. Although the amino acid residues that form these connections varied during the simulation, they belong to the LHCII trimers and the monomeric Lhcb subunits (see Supplementary Video). These results suggest that flexible physical connections might occur between the antenna systems of PSII-LHCII supercomplexes that face each other across the stromal gap. The stromal N-terminal loops of Lhcb4 subunits positioned at the “knot” density seen in our cryo-EM map immediately formed connections that lasted throughout the entire timeframe of the simulation; conversely, the connections observed between the overlapping LHCII trimers were transient (see Supplementary Video). No strong connections formed at the position of the “hinge” density seen in the cryo-EM map, in agreement with the hypothesis that the proteins comprising this structure are not present in our fitted atomic model. The results of these molecular dynamics simulations are consistent with our cryo-EM study and may explain why only the most durable connections are well resolved in our cryo-EM map of paired  $C_2S_2M$  supercomplexes (Fig. 3).

## Conclusions

The structure reported here for paired  $C_2S_2M$  PSII-LHCII supercomplexes is representative of the predominant form of PSII in pea plants grown at moderate light intensity. The interaction observed between apposing supercomplexes is mediated by physical connections that span the stromal gap, one of which is likely provided by the mutual interaction of two Lhcb4 subunits, possibly through salt bridges between negatively-charged amino acids in their N-termini. In addition, there is a specific overlap between facing LHCII trimers that may be due to the formation of salt bridges between highly conserved negatively-charged residues of the Lhcb subunits in facing LHCII trimers. The preservation of the PSII-LHCII supercomplex paired architecture requires the maintenance of physiological cation concentrations in all the buffers used during the purification procedure. This suggests that

cation salt bridges mediate both the physical connections across the stromal gap as well as the transient interactions between facing LHCII trimers.

Migration of excitation energy occurs between supercomplexes that interact across the stromal gap. Molecular dynamics simulations of the paired PSII-LHCII supercomplexes suggest different roles for the physical connections and the overlapping LHCII trimers; the former likely serve a structural function to hold the supercomplexes together, whereas the latter may enable energy transfer between PSII cores embedded in adjacent thylakoid membranes. Thus, the stromal gap-spanning interactions between PSII-LHCII supercomplexes likely enable both grana stacking and efficient energy transfer between adjacent thylakoids.

It remains to be tested how the physical connections and energetic coupling we observed in isolated paired  $C_2S_2M$  supercomplexes compare to PSII-LHCII supercomplexes within intact thylakoid membranes. Exploration of this interplay between supercomplex organization and thylakoid architecture will benefit from an extensive visualization of supercomplexes within the native thylakoid membranes using *in situ* cryo-electron tomography<sup>54</sup>. Until now, evidences of contacts between PSII-LHCII supercomplexes occurring in adjacent thylakoid membranes were provided by cryo-electron tomography<sup>20</sup>, revealing interactions mediated by the stromal surfaces of both the LHCII and the PSII reaction center dimers. In this study, the geometry observed for the interaction of supercomplexes in two adjacent membranes was different from that seen in our study, but this may be due to the specific arrangement in 2D arrays of the supercomplexes observed in the tomograms. Moreover, *Arabidopsis* mutants respond to a lack of conventional LHCII trimers by generating Lhcb5-containing trimers<sup>18, 55</sup>, and to a lack of PSII cores by accumulating large amounts of LHCII organized into three dimensional structures<sup>56</sup> to ultimately maintain their grana architecture. Therefore, the functional interactions we characterized for purified pairs of  $C_2S_2M$  PSII-LHCII supercomplexes are likely one of the several possible mechanisms occurring *in vivo* within adjacent thylakoid membranes that plants adopted to secure a functional grana stacking among different flexible strategies evolved to thrive in ever-changing light conditions.

## Methods

**Plant growth and isolation of thylakoid membranes.** Pea plants (*Pisum sativum* L., var. Palladio nano) were grown in a growth chamber (SANYO MLR-351H) at 20 °C and 60% humidity, with a 8 h light/16 h dark photoperiod under 150  $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ .

Thylakoid membranes were isolated from three-week-old leaves using the following procedure that maintains the stacked granal organization. These samples are hereafter referred to as stacked thylakoid membranes. Pea leaves were disrupted by grinding with a blender in 50 mM HEPES pH 7.5, 300 mM sucrose and 5 mM  $\text{MgCl}_2$ . The suspension was filtered through four cotton cloth layers, and the filtrate was centrifuged at 1,500 g for 10 min. The pellet was washed once by centrifugation in the same buffer and then homogenized in 5 mM  $\text{MgCl}_2$  and diluted 1:1 with 50 mM MES pH 6.0, 400 mM sucrose, 15 mM NaCl and 5 mM  $\text{MgCl}_2$ , followed by 10 min centrifugation at 3,000 g. The resulting pellet of stacked thylakoid membranes was washed once by centrifugation in 25 mM MES pH 6.0, 10 mM NaCl and 5 mM  $\text{MgCl}_2$ , and then suspended and stored in 25 mM MES pH 6.0, 10 mM NaCl, 5 mM  $\text{MgCl}_2$  and 2 M glycine betaine. When necessary, stacked thylakoids were flash frozen in liquid nitrogen and stored at  $-80^\circ\text{C}$ .

**Purification of pea PSII-LHCII supercomplexes.** PSII-LHCII supercomplexes were isolated according to a previously optimized protocol<sup>32</sup>. Briefly, stacked thylakoid membranes at a Chl concentration of 1  $\text{mg mL}^{-1}$  were treated with 50 mM n-dodecyl- $\alpha$ -D-maltoside ( $\alpha$ -DDM) for 1 min at 4 °C in the dark. Phenylmethylsulphonyl fluoride (500 mM) was present during the solubilization to inhibit protease activity. After centrifuging at 21,000 g for 10 min at 4 °C, 700  $\mu\text{l}$  of the supernatant were added to the top of a linear sucrose gradient, prepared by freezing and thawing ultracentrifuge tubes filled with a buffer consisting of 0.65 M sucrose, 25 mM MES pH 5.7, 10 mM NaCl, 5 mM  $\text{CaCl}_2$  and 0.03% (w/v) of  $\alpha$ -DDM (sucrose gradient buffer). For the crosslinking experiment with glutaraldehyde, 0.1% (v/v) glutaraldehyde was added to the sucrose gradient buffer. Centrifugation was carried out at 100,000 g for 12 h at 4 °C (Surespin 630 rotor, Thermo Scientific). The sucrose band containing PSII-LHCII supercomplexes was harvested and, if necessary, concentrated to a Chl concentration above 1.25  $\text{mg mL}^{-1}$  by membrane filtration with an Amicon Ultra 100 kDa cutoff device (Millipore) and then stored at  $-80^\circ\text{C}$ . Remaining glutaraldehyde was inactivated after fractionation by adding Tris-HCl pH 8.0 to a final concentration of 80 mM.

**Spectroscopic analyses.** Absorption spectra in native conditions were recorded using a Lambda25 spectrophotometer (Perkin Elmer) at 12 °C.

**Gel electrophoresis.** Separation of thylakoid membrane protein complexes and PSII-LHCII complexes was performed using large pore blue native polyacrylamide gel electrophoresis<sup>57</sup> (lpBN-PAGE), adopting the protocol previously described<sup>48</sup>.

**Mass spectrometry analyses.** PSII-LHCII supercomplexes were centrifuged at 20,000 g for 10 min at 4 °C, and the resulting pellet was rinsed in 10 mM HEPES pH 7.5. To remove the adhered pigments, proteins were precipitated in ice-cold acetone overnight at  $-20^\circ\text{C}$ . The extracting solution was centrifuged at 20,000 g for 20 min at 4 °C, and the denatured proteins were subsequently re-solubilized in a buffer consisting of 50 mM Tris-HCl pH 8.0, 7 M urea and 2 M thiourea. Insoluble material was removed by centrifuging at 15,000 g for 10 min. Protein concentration was determined using the Bradford assay<sup>58</sup>. Proteins at a concentration of 0.5  $\text{mg mL}^{-1}$  were reduced with 10 mM DTT for 30 min at 37 °C and alkylated with 20 mM iodoacetamide for 30 min at room temperature in the dark. The protein in-solution digestion was conducted by adding Trypsin/Lys-C Mix (Promega, WI, USA) at a final protein:protease ratio of 25:1 (w/w), followed by overnight incubation at 37 °C.

Peptide desalting was conducted by solid phase extraction<sup>59</sup> using 30 mg Oasis HLB cartridges (Waters, MA, USA). Peptides were dried and then dissolved in 30  $\mu$ L of LC-MS/MS mobile phase A (water containing 0.1% (v/v) formic acid).

LC-MS/MS analyses were performed using a micro-LC Eksigent Technologies (Dublin, USA) system, with a Halo Fused C18 column (0.5  $\times$  100 mm, 2.7  $\mu$ m; Eksigent Technologies Dublin, USA) as the stationary phase. The mobile phase was a mixture of 0.1% (v/v) formic acid in water (A) and 0.1% (v/v) formic acid in acetonitrile (B), eluting at a flow-rate of 15.0  $\mu$ L min<sup>-1</sup> and at an increasing concentration of solvent B from 2% to 40% over 30 min. The injection volume was 4.0  $\mu$ L. The LC system was interfaced with a 5600 + TripleTOF<sup>TM</sup> system (AB Sciex, Concord, Canada) equipped with DuoSpray<sup>TM</sup> Ion Source and CDS (Calibrant Delivery System). The mass spectrometer was operated in information dependent acquisition (IDA) mode. Peptide profiling was performed using a mass range of 100–1600 Da (TOF scan with an accumulation time of 100.0 ms), followed by a MS/MS product ion scan from 200 to 1250 Da (accumulation time of 5.0 ms) with the abundance threshold set at 30 cps. The ion source parameters in electrospray positive mode were set as follows: curtain gas (N<sub>2</sub>) at 25 psig, nebulizer gas GAS1 at 25 psig, and GAS2 at 20 psig, ion spray floating voltage (ISFV) at 5000 V, source temperature at 450 °C and declustering potential at 25 V.

MS data were acquired with Analyst TF 1.7 (AB SCIEX, Concord, Canada). Raw files were processed with the search engine ProteinPilot<sup>TM</sup> v.5.0.1.0, 4895 (AB Sciex, Concord, Canada) using the Paragon algorithm v.5.0.1.0, 4874. The following sample parameters were used: Trypsin/Lys-C digestion, cysteine alkylation set to carbamidomethylation and no special factors. Processing parameters were set to “Biological modification”. All data files were searched, thorough ID search effort, using a UniProtKB/TrEMBL database containing Viridiplantae proteins (version 2016.09.02, with a total of 3,825,803 sequences), or a *de novo* database obtained by generating open reading frames from all six frames of each transcript in the *P. sativum* transcriptome (p.sativum\_csfl\_reftransV1 downloaded from <https://www.coolseasonfoodlegume.org/organism/Pisum/sativum/reftrans/v1>) concatenated with a reversed “decoy” version of the “forward” database. After searching, we accepted protein IDs that had a ProteinPilot Unused Score of at least 1.3 (equivalent to a 95% confidence interval) as a cutoff threshold and an estimated local false discovery rate (FDR) not higher than 1% according to Rardin *et al.*<sup>60</sup>.

**Sequence analysis.** Clustal Omega<sup>61</sup> and Jalview<sup>62</sup> software were used for multiple sequence alignment and conservation analysis, respectively. A phylogenetic tree was generated with ClustalW2.0 software<sup>63</sup> using the UPGMA method and PAM250 substitution matrix.

**Cryo-EM data collection.** PSII-LHCII supercomplexes were prepared at 1 mg mL<sup>-1</sup> in the sucrose gradient buffer. 4  $\mu$ L of sample were applied to a glow-discharged lacey carbon grid (200 Cu mesh, Quantifoil) within the chamber of a Vitrobot (mark 3, FEI). After 60 s incubation at 100% humidity and 21 °C, a quick wash with 4  $\mu$ L of 10 mM Hepes pH 7.5 was performed to remove sugar, excess solution was blotted from both sides for 4 s, and the grid was plunge-frozen in a liquid ethane/propane mixture. Samples were exposed to only dim green light during the grid preparation procedure. Data collection was performed on a Titan Krios microscope (FEI) operated at 300 kV using EPU automated acquisition software (FEI). Spanning a defocus range –1 to –3  $\mu$ m, 6,834 micrographs were recorded on a Falcon II direct electron detector (FEI) at 59,000 magnification (image pixel size of 1.4 Å), with a total dose of 47.5 e<sup>-</sup>/Å<sup>2</sup> fractionated over 7 frames (1.5 s exposure, dose rate of 30 e<sup>-</sup>/Å<sup>2</sup>/s).

**Data processing and 3D reconstruction.** Beam-induced motion was corrected by aligning the image frames with in-house developed software (<https://github.com/dtegunov/k2align>). Aligned images were used for single particle analysis. All image processing steps were performed using the Scipion platform (<http://scipion.cnb.csic.es>), which is an image processing framework that integrates several software packages into a unified interface<sup>64</sup>. For estimating the objective lens defocus parameters (i.e., contrast transfer function/CTF) in the transmission electron micrographs, we used the CTFIND4 program<sup>65</sup>. For this project, a total of 33,729 cryo-EM particles were manually extracted from 6,834 micrographs using the particle picking tool of Xmipp<sup>66</sup>. These particles were classified in 2D using Xmipp<sup>67</sup> and then in 3D using Relion<sup>68</sup>. Using e2initialmodel.py of EMAN2.1<sup>69</sup> and Ransac of Xmipp<sup>70</sup>, averages assigned to the PSII-LHCII supercomplex were used to generate several initial models and subsequently select among them the one with the highest score to be used as an unbiased low-resolution 3D template for refinement and classification. This template was used as an initial model for Relion 3D classification<sup>68</sup>. After 3D classification, 14,291 particles were identified as paired C<sub>2</sub>S<sub>2</sub>M supercomplexes, 7,021 particles as paired C<sub>2</sub>S<sub>2</sub> supercomplexes and 5,641 particles as unpaired C<sub>2</sub>S<sub>2</sub>M supercomplexes. The paired C<sub>2</sub>S<sub>2</sub>M particles were processed further with 3D refinement, using the Relion auto-refine algorithm<sup>68</sup> and D2 point group symmetry. After refinement, we used the multireference alignability method, recently implemented in Scipion, to select the “best particles” in the dataset according to the reference map and provide the alignment accuracy for each particle used in the reconstruction<sup>71</sup>. The final cryo-EM model was validated using the soft-alignment validation approach recently described by Vargas *et al.*<sup>71</sup>.

**Fitting of atomic models into cryo-EM maps.** UCSF Chimera<sup>72</sup> software was used to model atomic structures into the 3D cryo-EM reconstruction. Surface-rendered views of the EM density were calculated at a threshold of 2.5  $\sigma$ . Atomic co-ordinates from spinach<sup>26</sup> (PDB ID: 3JCU) were used for modeling both the central dimeric PSII core (removing chains P, Q and U, corresponding to the extrinsic subunits PsbP, PsbQ and PsbTn, respectively) and the monomeric Lhcb4 and Lhcb5 subunits. Atomic coordinates from pea<sup>36</sup> (PDB ID: 2BHW) were used for fitting the LHCII S- and M-trimers. For the monomeric Lhcb6, we fitted the 3D structure predicted by the PHYRE2 algorithm<sup>39</sup> for the *P. sativum* Lhcb6 protein sequence derived from the corresponding transcript (p.sativum\_csfl\_reftransV1\_0079196\_5/148-357). Local fitting and adjustment of the Lhcb6 subunit and

the LHCII trimers in the cryo-EM maps were performed by visual inspection, aided in the case of the S-trimer by its known localization in the available  $C_2S_2$  atomic structure<sup>26</sup>.

**Negative stain EM.** PSII-LHCII supercomplexes were prepared at  $1 \text{ mg mL}^{-1}$  in the sucrose gradient buffer. To remove salts from the sample in a centrifugation step, we used Amicon Ultra 100 kDa cutoff buffer exchange spin columns (Millipore) and a salt-free buffer containing 25 mM MES pH 5.7 and 0.03% (w/v)  $\alpha$ -DDM at a final ratio 1:60 (v/v). Concentrated samples were diluted to a final Chl concentration of  $\sim 30 \mu\text{g mL}^{-1}$  either in the sucrose gradient buffer or in the buffer with no salts. Samples were applied to glow discharged carbon-coated copper grids, washed quickly with distilled water and negatively stained with 2% (w/v) uranyl acetate. A FEI Tecnai F20-ST transmission electron microscope, equipped with a field emission gun (FEG) operated at 200 kV, was used for acquisition of micrographs, recorded at 38,000 magnification on a Gatan Orius 4.0 K  $\times$  2.7 K CCD camera.

**Chlorophyll a fluorescence induction measurements.** PSII-LHCII supercomplexes prepared at  $1.25 \text{ mg mL}^{-1}$  in the sucrose gradient buffer were used for the experiments. Samples were diluted to a final Chl concentration of  $1.25 \mu\text{g mL}^{-1}$  in the sucrose gradient buffer or in a similar buffer devoid of any salt. The fluorescence induction OJIP transient was measured at room temperature, with a FL3500 double modulation fluorometer (Photon Systems Instruments). Dark-adapted samples were illuminated for 1 s with continuous actinic light ( $2,400 \mu\text{mol photons m}^{-2} \text{ s}^{-1}$ , emission peak at 630 nm). The first reliable point of the transient is measured at  $t_0 = 0.02 \text{ ms}$  after the onset of illumination. After recording fast fluorescence transients, the OJIP curves were analyzed according to methods previously reported<sup>49–53</sup> to estimate the energetic connectivity between PSII units. Results shown in graphs and tables are presented as the mean value  $\pm$  standard deviation of six replicates.

**Molecular dynamics simulations.** The atomic model generated by fitting the 3D cryo-EM map of the paired  $C_2S_2M$  supercomplexes (Fig. 2) was embedded within a pair of thylakoid membranes constituted by phosphatidylglycerol (PG), digalactosyldiacylglycerol (DGDG), monogalactosyldiacylglycerol (MGDG) and sulfoquinovosyl-diacylglycerol (SQDG) molecules, according to the detailed thylakoid membrane composition of higher plants proposed by van Eerden *et al.*<sup>73</sup>. The starting structure for the molecular dynamics simulations was the atomic coordinates of the protein components modeled into the cryo-EM 3D map, with the exclusion of any bound cofactors, and the predicted structure of Lhcb6. Prior to the insertion of the paired PSII-LHCII supercomplexes into the pair of membranes, the supercomplexes and the membranes were separately energy minimized and equilibrated. Subsequently, the supercomplex pair was embedded into the thylakoid membranes and solvated with water molecules and  $\text{Na}^+$  and  $\text{Cl}^-$  ions according to van Eerden *et al.*<sup>73</sup> in order to reach a globally uncharged system.

A 7 ns fine-grained simulation was performed with the GROMACS 4.6.6 package<sup>74</sup>, with the system interactions modeled by the GROMOS G53a6 force field<sup>75</sup>. The system was simulated using the Nosé-Hoover thermostat to control the temperature<sup>76,77</sup>. By setting the thermal bath temperature at 300 K, the paired PSII-LHCII supercomplex, the pair of thylakoid membranes and the solvent were independently coupled every 0.5 ps. The pressure was semi-isotropically coupled every 2 ps to a reference pressure of 1 bar using the Parrinello-Rahman barostat<sup>78</sup>, and a compressibility of  $4.6 \times 10^{-5} \text{ bar}^{-1}$  was applied to the system. Electrostatic and Van der Waals interactions were calculated using the Particle Mesh Ewald potential with a 1.4 nm cutoff. The mobility of the PSII-LHCII supercomplex pair was determined by calculating the root mean square fluctuations (RMSF) of backbone atoms (N,  $C_\alpha$  and C atoms) during the molecular dynamics simulations with respect to the reference time-averaged structure:

$$\text{RMSF}_i = \sqrt{\langle (r_i - \langle r_i \rangle)^2 \rangle}$$

where  $r_i$  is the  $i$ -th backbone atoms and  $\langle \rangle$  denotes the time-average.

**Data availability.** The data that support the findings of this study are available from the corresponding author upon request. The cryo-EM map was deposited at the Electron Microscopy Data Bank (EMDB) with accession code EMD-3825.

**Addendum.** After the completion of this manuscript, a paper was published by van Bezouwen *et al.*<sup>79</sup>, which reported a 5.3 Å structure of the higher plant unpaired  $C_2S_2M_2$  supercomplex. This PSII-LHCII supercomplex was isolated from *Arabidopsis* and its structure determined by single particle cryo-EM. In this work, the supramolecular organization of the subunits in the  $C_2S_2M_2$  is described in particular with respect to the location and orientation of trimeric and monomeric LHCII, and the location of all the chlorophylls in the subunits and their role in the energy flow from the peripheral light harvesting complexes to the PSII core is discussed. It is worth noting that, despite the lower resolution of our cryo-EM map, the fitting of the  $C_2S_2M$  components, and in particular of the LHCII trimers and the Lhcb6 subunit, is perfectly in accordance with that proposed by van Bezouwen *et al.*<sup>79</sup>. It is also worth noting that most of the chlorophylls responsible for energy connections observed in this recent structure are on the stromal side of the LHCII complexes. This evidence may suggest their involvement in the energy exchange between PSII cores located in adjacent thylakoid membranes *in vivo*, in accordance to the increase of PSII connectivity measured in our isolated paired PSII-LHCII supercomplexes.

## References

1. Barber, J. Photosystem II: an enzyme of global significance. *Biochem. Soc. Trans.* **34**, 619–631 (2006).
2. Boekema, E. J. *et al.* Supramolecular structure of the photosystem II complex from green plants and cyanobacteria. *Proc. Natl. Acad. Sci.* **92**, 175–179 (1995).
3. Hankamer, B. *et al.* Isolation and characterization of monomeric and dimeric photosystem II complexes from spinach and their relevance to the organisation of photosystem II *in vivo*. *Eur. J. Biochem.* **243**, 422–429 (1997).

4. Danielsson, R. *et al.* Dimeric and monomeric organization of photosystem II. Distribution of five distinct complexes in the different domains of the thylakoid membrane. *J. Biol. Chem.* **281**, 14241–14249 (2006).
5. Pagliano, C., Saracco, G. & Barber, J. Structural, functional and auxiliary proteins of photosystem II. *Photosynth. Res.* **116**, 167–188 (2013).
6. Ballottari, M., Girardon, J., Dall'Osto, L. & Bassi, R. Evolution and functional properties of Photosystem II light harvesting complexes in eukaryotes. *Biochim. Biophys. Acta - Bioenerg.* **1817**, 143–157 (2012).
7. Nevo, R. *et al.* Architecture of thylakoid membrane networks. *Lipids Photosynth. Essent. Regul. Funct.* **30**, 295–328 (2009).
8. Kim, E.-H., Chow, W. S., Horton, P. & Anderson, J. M. Entropy-assisted stacking of thylakoid membranes. *Biochim. Biophys. Acta* **1708**, 187–195 (2005).
9. Borodich, A., Rojdestvenski, I. & Cottam, M. Lateral heterogeneity of photosystems in thylakoid membranes studied by Brownian dynamics simulations. *Biophys. J.* **85**, 774–789 (2003).
10. Kirchhoff, H. Significance of protein crowding, order and mobility for photosynthetic membrane functions. *Biochem. Soc. Trans.* **36**, 967–970 (2008).
11. Izawa, S. & Good, N. E. Effect of salts and electron transport on the conformation of isolated chloroplasts. II. *Electron Microscopy. Plant Physiol.* **41**, 544–552 (1966).
12. Barber, J. Influence of Surface Charges on Thylakoid Structure and Function. *Annu. Rev. Plant Physiol.* **33**, 261–295 (1982).
13. Albertsson, P. A quantitative model of the domain structure of the photosynthetic membrane. *Trends Plant Sci.* **6**, 349–358 (2001).
14. Dekker, J. P. & Boekema, E. J. Supramolecular organization of thylakoid membrane proteins in green plants. *Biochim. Biophys. Acta - Bioenerg.* **1706**, 12–39 (2005).
15. Anderson, J. M., Chow, W. S. & De Las Rivas, J. Dynamic flexibility in the structure and function of photosystem II in higher plant thylakoid membranes: the grana enigma. *Photosynth. Res.* **98**, 575–587 (2008).
16. Mullet, J. E. The amino acid sequence of the polypeptide segment which regulates membrane adhesion (grana stacking) in chloroplasts. *J. Biol. Chem.* **258**, 9941–9948 (1983).
17. Day, D. A., Ryrrie, I. J. & Fuad, N. Investigations of the role of the main light-harvesting chlorophyll-protein complex in thylakoid membranes. Reconstitution of depleted membranes from intermittent-light-grown plants with the isolated complex. *J. Cell Biol.* **98**, 163–172 (1984).
18. Andersson, J. *et al.* Absence of the Lhcb1 and Lhcb2 proteins of the light-harvesting complex of photosystem II - effects on photosynthesis, grana stacking and fitness. *Plant J.* **35**, 350–361 (2003).
19. Pribil, M., Labs, M. & Leister, D. Structure and dynamics of thylakoids in land plants. **65**, 1955–1972 (2014).
20. Daum, B., Nicastro, D., Austin, J., McIntosh, J. R. & Kühlbrandt, W. Arrangement of Photosystem II and ATP Synthase in chloroplast membranes of spinach and pea. *Plant Cell* **22**, 1299–1312 (2010).
21. Garab, G. & Mustardy, L. Role of LHClI-containing macrodomains in the structure, function and dynamics of grana. *Aust. J. Plant Physiol.* **27**, 648–658 (1999).
22. Kirchhoff, H. *et al.* Structural and functional self-organization of Photosystem II in grana thylakoids. *Biochim. Biophys. Acta - Bioenerg.* **1767**, 1180–1188 (2007).
23. Caffarri, S. *et al.* Functional architecture of higher plant photosystem II supercomplexes. *EMBO J.* **28**, 3052–3063 (2009).
24. Kouřil, R., Dekker, J. P. & Boekema, E. J. Supramolecular organization of photosystem II in green plants. *Biochim. Biophys. Acta - Bioenerg.* **1817**, 2–12 (2012).
25. Nield, J. *et al.* 3D map of the plant photosystem II supercomplex obtained by cryoelectron microscopy and single particle analysis. *Nat. Struct. Biol.* **7**, 44–47 (2000).
26. Wei, X. *et al.* Structure of spinach photosystem II-LHClI supercomplex at 3.2 Å resolution. *Nature* **1**, 1–18 (2016).
27. Boardman, N. K. Comparative photosynthesis of sun and shade plants. *Annu. Rev. Plant Physiol.* **28**, 355–377 (1977).
28. Anderson, J. M. Photoregulation of the composition, function, and structure of thylakoid membranes. *Annu. Rev. Plant Physiol.* **37**, 93–136 (1986).
29. Jia, H., Liggins, J. R. & Chow, W. S. Acclimation of leaves to low light produces large grana: the origin of the predominant attractive force at work. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **367**, 3494–3502 (2012).
30. Rumak, I. *et al.* 3-D modelling of chloroplast structure under (Mg<sup>2+</sup>) magnesium ion treatment. Relationship between thylakoid membrane arrangement and stacking. *Biochim. Biophys. Acta - Bioenerg.* **1797**, 1736–1748 (2010).
31. Schröppel-Meier, G. & Kaiser, W. M. Ion homeostasis in chloroplasts under salinity and mineral deficiency: II. Solute distribution between chloroplasts and extrachloroplastic space under excess or deficiency of sulfate, phosphate, or magnesium. *Plant Physiol.* **87**, 828–832 (1988).
32. Barera, S., Pagliano, C., Pape, T., Saracco, G. & Barber, J. Characterization of PSII-LHClI supercomplexes isolated from pea thylakoid membrane by one-step treatment with  $\alpha$ - and  $\beta$ -dodecyl-D-maltoside. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **367**, 3389–3399 (2012).
33. Albanese, P. *et al.* Dynamic reorganization of photosystem II supercomplexes in response to variations in light intensities. *Biochim. Biophys. Acta - Bioenerg.* **1857**, 1651–1660 (2016).
34. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).
35. Kirchhoff, H. *et al.* Dynamic control of protein diffusion within the granal thylakoid lumen. *Proc. Natl. Acad. Sci.* **108**, 20248–20253 (2011).
36. Standfuss, J. T., van Scheltinga, A. C., Lamborghini, M. & Kühlbrandt, W. Mechanisms of photoprotection and nonphotochemical quenching in pea light-harvesting complex at 2.5 Å resolution. *EMBO J.* **24**, 919–928 (2005).
37. Pagliano, C. *et al.* Proteomic characterization and three-dimensional electron microscopy study of PSII-LHClI supercomplexes from higher plants. *Biochim. Biophys. Acta* **1837**, 1454–1462 (2014).
38. Pan, X. *et al.* Structural insights into energy regulation of light-harvesting complex CP29 from spinach. *Nat. Struct. Mol. Biol.* **18**, 309–315 (2011).
39. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
40. Wan, T. *et al.* Crystal structure of a multilayer packed major Light-Harvesting Complex: implications for grana stacking in higher plants. *Mol. Plant* **7**, 916–919 (2014).
41. Boekema, E. J., van Roon, H., van Breemen, J. F. L. & Dekker, J. P. Supramolecular organization of photosystem II and its light-harvesting antenna in partially solubilized photosystem II membranes. *Eur. J. Biochem.* **266**, 444–452 (1999).
42. Roose, J. L., Frankel, L. K., Mummadisetti, M. P. & Bricker, T. M. The extrinsic proteins of photosystem II: update. *Planta* **243**, 889–908 (2016).
43. Allahverdiyeva, Y. *et al.* Arabidopsis plants lacking PsbQ and PsbR subunits of the oxygen-evolving complex show altered PSII super-complex organization and short-term adaptive mechanisms. *Plant J.* **75**, 671–684 (2013).
44. Shabestari, M. H., Wolfs, C. J. A. M., Spruijt, R. B., van Amerongen, H. & Huber, M. Exploring the structure of the 100 amino-acid residue long N-terminus of the plant antenna protein CP29. *Biophys. J.* **106**, 1349–1358 (2014).
45. Fristedt, R. & Vener, A. V. High light induced disassembly of Photosystem II supercomplexes in Arabidopsis requires STN7-dependent phosphorylation of CP29. *PLoS One* **6**, e24565 (2011).
46. Fadoulglou, V. E., Kokkinidis, M. & Glykos, N. M. Determination of protein oligomerization state: Two approaches based on glutaraldehyde crosslinking. *Anal. Biochem.* **373**, 404–406 (2008).

47. Kastner, B. *et al.* GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55 (2008).
48. Albanese, P. *et al.* Isolation of novel PSII-LHCII megacomplexes from pea plants characterized by a combination of proteomics and electron microscopy. *Photosynth. Res.* **130**, 19–31 (2016).
49. Stirbet, A. & Govindjee. On the relation between the Kautsky effect (chlorophyll a fluorescence induction) and Photosystem II: Basics and applications of the OJIP fluorescence transient. *J. Photochem. Photobiol. B Biol.* **104**, 236–257 (2011).
50. Stirbet, A. Excitonic connectivity between photosystem II units: what is it, and how to measure it? *Photosynth. Res.* **116**, 189–214 (2013).
51. Joliot, A. & Joliot, P. Etude cinétique de la réaction photochimique libérant l'oxygène au cours de la photosynthèse. *Comptes Rendus l'Académie des Sci.* **258**, 4622–4625 (1964).
52. Strasser, R. J. In *Chloroplast Development* (eds Argyroudi-Akoyunoglou, J. & Akoyunoglou, G.) Elsevier Biomedical, Amsterdam, 513–538 (1978).
53. Strasser, R. J. & Stirbet, A. D. Estimation of the energetic connectivity of PS II centres in plants using the fluorescence rise O-J-I-P - Fitting of experimental data to three different PS II models. *Math. Comput. Simul.* **56**, 451–461 (2001).
54. Asano, S., Engel, B. D. & Baumeister, W. *In Situ* Cryo-Electron Tomography: A Post-Reductionist Approach to Structural Biology. *J. Mol. Biol.* **428**, 332–343 (2016).
55. Ruban, A. V. *et al.* Plants lacking the main light-harvesting complex retain photosystem II macro-organization. *Nature* **421**, 648–652 (2003).
56. Belgio, E., Ungerer, P. & Ruban, A. V. Light-harvesting superstructures of green plant chloroplasts lacking photosystems. *Plant Cell Environ.* **38**, 2035–2047 (2015).
57. Järvi, S., Suorsa, M., Paakkarinen, V. & Aro, E.-M. Optimized native gel systems for separation of thylakoid protein complexes: novel super- and mega-complexes. *Biochem. J.* **439**, 207–214 (2011).
58. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
59. Guo, X. & Kristal, B. S. The use of under-loaded C18 solid-phase extraction plates increases reproducibility of analysis of tryptic peptides from unfractionated human plasma. *Anal. Biochem.* **426**, 86–90 (2012).
60. Rardin, M. J. *et al.* MS1 peptide ion intensity chromatograms in MS2 (SWATH) Data Independent Acquisitions. Improving post acquisition analysis of proteomic experiments. *Mol. Cell. Proteomics* **14**, 2405–2419 (2015).
61. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539–539 (2014).
62. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
63. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
64. de la Rosa-Trevín, J. M. *et al.* Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195**, 93–99 (2016).
65. Rohou, A. & Grigorieff, N. CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
66. Abrishami, V. *et al.* A pattern matching approach to the automatic selection of particles from low-contrast electron micrographs. *Bioinformatics* **29**, 2460–2468 (2013).
67. Sorzano, C. O. S. *et al.* A clustering approach to multireference alignment of single-particle projections in electron microscopy. *J. Struct. Biol.* **171**, 197–206 (2010).
68. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
69. Bell, J. M., Chen, M., Baldwin, P. R. & Ludtke, S. J. High resolution single particle refinement in EMAN2.1. *Methods* **100**, 25–34 (2016).
70. Vargas, J., Alvarez-Cabrera, A.-L., Marabini, R., Carazo, J. M. & Sorzano, C. O. S. Efficient initial volume determination from electron microscopy images of single particles. *Bioinformatics* **30**, 2891–2898 (2014).
71. Vargas, J., Melero, R., Gomez-Blanco, J., Carazo, J.-M. & Sorzano, C. O. S. Quantitative analysis of 3D alignment quality: its impact on soft-validation, particle pruning and homogeneity analysis. *Sci. Rep.* **7**, 6307 (2017).
72. Pettersen, E. F. *et al.* UCSF Chimera-a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
73. Van Eerden, F. J., De Jong, D. H., De Vries, A. H., Wassenaar, T. A. & Marrink, S. J. Characterization of thylakoid lipid membranes from cyanobacteria and higher plants by molecular dynamics simulations. *Biochim. Biophys. Acta - Biomembr.* **1848**, 1319–1330 (2015).
74. Hess, B., Kutzner, C., Van Der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).
75. Oostenbrink, C., Villa, A., Mark, A. E. & Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656–1676 (2004).
76. Hoover, W. G., Ladd, A. J. C. & Moran, B. High-strain-rate plastic flow studied via nonequilibrium molecular dynamics. *Phys. Rev. Lett.* **48**, 1818–1820 (1982).
77. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 (1984).
78. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
79. van Bezouwen, L. S. *et al.* Subunit and chlorophyll organization of the plant photosystem II supercomplex. *Nat. Plants.* **3**, 17080 (2017).

## Acknowledgements

This work was supported by the Italian Ministry of Education, University and Research, “Futuro in Ricerca 2013” program RBFR1334SB to C.P. and P.B. The authors acknowledge the support of employees and the use of experimental resources of Instruct, through Instruct Access Projects PID:1112 and PID:1363; Prof. Wolfgang Baumeister and Prof. Jürgen Plitzko for access to cryo-EM instrumentation and Gunter Pfeifer for assistance with cryo-EM data collection. Computational resources were provided by HPC@POLITO (The HPC Polito Project. Available: <http://www.hpc.polito.it>). We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Quadro GP100 GPU used for this research. The authors thank Dr. Andrea Fantuzzi for helpful discussion.

## Author Contributions

C.P. conceived the study. C.P., J.M.C., G.Z., E.M., and G.S. supervised the study. P.A. isolated PSII-LHCII supercomplexes. B.D.E. performed the preparation of cryo-EM grids and data acquisition. P.A., P.B. and A.G.

manually picked particles from cryo-EM dataset. R.M., J.V. and C.O.S.S. performed the image processing. P.A., C.P., P.B., A.G., and G.Z. performed the model fitting. P.A. prepared samples for mass spectrometry and analysed mass spectrometry data. M.M. performed mass spectrometry data acquisition. C.P. and P.A. prepared the EM grids in negative stain and A.C. performed image acquisition. P.A. performed gel electrophoresis, sequence alignments and phylogenetic analysis. P.A. and C.P. performed fluorescence analyses. A.G. performed molecular dynamics simulations. C.P. wrote the article. B.D.E. critically read and revised the manuscript. All authors reviewed the results and approved the final version of the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-10700-8](https://doi.org/10.1038/s41598-017-10700-8)

**Competing Interests:** The authors declare that they have no competing interests.

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017