High-resolution model of Arabidopsis Photosystem II reveals the consequences of digitonin-extraction

*André T. Graça,[a], Michael Hall[a], Karina Persson[a] and Wolfgang P. Schröder[a] **

[a] Department of Chemistry, Umeå University, SE-901 87 Umeå, Sweden

* Corresponding author

Abstract

In higher plants, the photosynthetic process is performed and regulated by Photosystem II (PSII). *Arabidopsis thaliana* was the first higher plant with a fully sequenced genome, conferring it the status of a model organism; nonetheless, a high-resolution structure of its Photosystem II is missing. We present the first Cryo-EM high-resolution structure of Arabidopsis PSII supercomplex with average resolution of 2.79 Å, an important model for future PSII studies. The digitonin extracted PSII complexes demonstrate the importance of: the LHG2630-lipid-headgroup in the trimerization of the light-harvesting complex II; the stabilization of the PsbJ subunit and the CP43-loop E by DGD520-lipid; the choice of detergent for the integrity of membrane protein complexes. Furthermore, our data shows at the anticipated Mn_4CaO_5-site a single metal ion density as a reminiscent early stage of Photosystem II photoactivation.

Keywords: Photosystem II; cryo-electron microscopy; digitonin; Mn_4CaO_5; lipids

Introduction

Photosynthesis is the most important natural process existing on our planet, in which solar radiation is converted into biomass[1,2]. It provides food, oxygen to breathe and renewable, sustainable energy. It is also one of the few processes that reduce amounts of greenhouse gas carbon dioxide in the air[3]. The weed plant *Arabidopsis thaliana* (thale cress or mouse-ear cress) became a model organism for plant research due to several traits such as size, generation time, accessibility, manipulation, and genetics[4]. The use of Arabidopsis as a model organism was enhanced as its full genome was sequenced and published in the year 2000[5].

Photosystem II (PSII)—a key player in photosynthesis—is a multi-subunit light-driven water:plastoquinone oxidoreductase[6] located in the thylakoid membrane of chloroplasts[7]. Functional PSII has a dimeric core, surrounded by light-harvesting complexes (LHCII), and thus constitutes one of the largest protein complexes in nature with more than 60 protein subunits and a molecular mass exceeding 1400 kDa[8]. In its monomeric composition, Photosystem II reaction centre is built of the essential D1-D2 heterodimer, that hosts several cofactors responsible for the light-induced charge separation and water oxidation[6,8]. These core proteins are connected to an internal core chlorophyll *a* binding antenna, composed of CP43 and CP47 proteins[9]. Chlorophyll *a/b* binding minor antenna proteins (CP24, CP26, CP29) link the major chlorophyll *a/b* binding LHCII to the PSII core[10,11]. The trimeric LHCII is the most abundant chlorophyll-binding protein assembly in eukaryotic photosynthetic organisms, and it is constituted of three polypeptides, Lhcb1Lhcb1-3[12,13]. Several of these trimeric assemblies can bind simultaneously to a PSII core complex at specific locations. Their relative location to the PSII dimeric core (C_2) and the strength of the PSII-LHCII interaction creates4 distinct classes[14]: strong (S), medium (M), loose (L) and naked (N). The N-LHCII type, recently discovered in the green algae *Chlamydomonas reinhardtii*[15–17]; and the less characterized L-LHCII type can be found in land plants[18], although trimers are not abundantly associated with PSII under optimal light conditions.

The lumen side of land plants PSII complex harbours four extrinsic proteins (PsbO, PsbP, PsbQ and PsbTn). Additionally, there are at least 10 low molecular mass proteins (< 10 kDa) at various sites, between major proteins and in the outskirts of the complex[19]. An immense network of pigments, lipids and other essential cofactors are known to bind to the different

proteins of the complex[8,20–22]. The pigments—chlorophylls and carotenoids—are essential for light absorption while the lipids main function is the ligation of different subunits at several oligomerization interfaces and the binding of some pigments[23].

Structural studies of Photosystem II were for an extended period restricted to X-ray crystallography methods which demand high concentrations of pure homogeneous and stable complexes and the capability of symmetric assembly to form crystals[24,25]. Thus, structural studies of bacterial PSII from *Thermosynechococcus elongates* and *Thermosynechococcus vulcanus* have been commonly performed due to their high stability and accessible large batch extraction. Such attempts resulted in several high- to near atomic resolution structures, between 1.9-3.8 Å average resolution[21,26–29]. The bacterial PSIIcc is expected to be very similar to the dimeric PSII core of higher plants (PSII C_2), however there are several distinct differences[30]: i. bacterial PSII lacks chlorophyll *b*; ii. the light-harvesting antenna is integrated into the membrane in higher plants, while it is peripheral in bacteria (phycobilisomes); iii. the extrinsic proteins differ between bacteria and higher plants; iv. the pigments and the lipids not always share conserved binding-sites between species of different kingdoms.

Higher plant PSII supercomplexes have traditionally been extracted from thylakoid membranes using dodecylmaltoside (α- or β- isoforms) and digitonin detergents[31–33]. Other detergents have been screened for the purpose, but results show that PSII complexes do not remain stable in a wide variety of detergents[31]. The high concentration of detergents, and the different sizes of PSII complexes extracted from the thylakoid membrane, are two factors that hindered the formation of stable crystal assemblies[34,35]. A new opportunity to study the higher plants PSII structure came with the recent developments of single-particle cryo-electron microscopy (EM)[36]. This method allows acquisition of structural information from large protein complexes at resolutions similar to X-ray crystallography. As the sample is plunge frozen and the imaged particles are segregated into classes upon computational analyses, the sample does not need to be as pure and enriched as for crystallisation and their level of stability is only relevant up to the moment of plunge freezing[36]. The first higher plant PSII Cryo-EM structures were recently reported for; *Spinacia oleracea* (Spinach) at 3.2 Å resolution (PDB: 3JCU)[22]; *Pisum Sativum* (Pea), with a resolution of 2.7 Å and 3.2 Å (PDB: 5XNL and 5XNM, respectively)[20], the former consisting of a smaller C_2S_2 complex. However, concerning

Arabidopsis thaliana, the best structural model of PSII (PDB: 5MDX)[33]—extracted from a mutant phenotype—has a resolution of 5.3 Å.

Photosystem II oxygen production concurrently exposes the complex to oxidative stress, leading to damage of the D1 protein[37]. To cope with this, plants have developed various protective mechanisms, such as the ability to repair the PSII complex by replacing the D1 protein[38]. Under high-light conditions, i.e. when plants absorb more light than they can use for photosynthesis the D1 protein is replaced every 20-30 minutes, which means that the assembly state of PSII is in a constant switch between functional and repair/inactive state. To become functional, PSII needs to correctly assemble the site for water oxidation; a process driven by light and known as photoactivation[37,39]. At distinct steps of photoactivation, manganese ions (Mn^{2+}) interact with specific residues of the D1 protein, while being oxidized[39]. One calcium ion (Ca^{2+}) and five oxygen atoms bridging the metal ions are indispensable to form the functional high-valence manganese cluster, Mn_4CaO_5. When the full water oxidation complex is formed, Mn_4CaO_5 is coordinated to several D1 and CP43 amino acids[26]. Even though most of the photoactivation steps are still to be understood, there is evidence that manganese and calcium ions compete to bind at the high-affinity Mn-binding site (HAS)[39]—involving at least the D1 residue Asp170—in an early step of photoactivation. The following steps are the oxidation of the first-bound Mn^{2+} and the so-called 'dark rearrangement', proposed as a light-independent dynamic where the C-terminal of the D1 protein adopts a new structural conformation like the observed in the Mn_4CaO_5 binding pocket of mature PSII[39]. In an attempt to gain new information of this activation process, Gisriel and colleagues[40] isolated monomeric PSII from *Synechocystis* sp. PCC 6803 that were EDTA-washed and buffer exchanged. This recent study showed that the Mn_4CaO_5-binding pocket undergoes major rearrangements upon assembly of the OEC, before the formation of a fully functional PSII.

Here we present the first high-resolution structure of Arabidopsis PSII $C_2S_2M_2$ supercomplex. This model shows that the absence of the extrinsic PsbP and PsbQ proteins leads to a disordered C-terminus of the reaction centre D1 protein which we speculate to be important for the assembly process of the manganese water splitting complex of PSII. Furthermore, new

knowledge of protein-protein, protein-ligand and ligand-ligand interactions is attained by accessing the consequences of the digitonin molecules bound to PSII.

Results

The overall structure of the Arabidopsis Photosystem II complex

The imaged Photosystem II $C_2S_2M_2$-type complexes were isolated from wild-type Arabidopsis BBY membrane fragments[41] solubilised with a mixture of digitonin and β-DDM detergents. Visual inspection of the obtained micrographs suggested a high level of sample homogeneity. At the same time, the implemented *ab initio* 3D reconstruction strategy allowed us to further filter the dataset to a more homogeneous subset of 100 712 PSII particles. The 3D reconstruction, with imposed C2 symmetry, yielded a map at an overall resolution of 2.79 Å for the C_2S_2 supercomplex, according to $FSC_{0.143}$ golden standard (Appendix, Fig S1 and Table S1). However, the resolution was found to be heterogeneous, with a higher resolution of 2.5 Å throughout the core region and a lower resolution of 6.0 Å in the outer regions of the complex (Appendix, Fig S2). An additional data collection using a Volta phase plate was performed on the same grid, yielding a map at 3.6 Å global resolution ($C_2S_2M_2$) that helped to resolve certain peripheric regions of the complex such as additional amino acids of stromal terminal regions of certain subunits (e.g.: N-terminal of the D1 protein).

The Arabidopsis PSII $C_2S_2M_2$ supercomplex has a two-fold symmetry structure (Appendix, Fig S3), with dimensions estimated to be 280 x 180 x 107 Å (LWH). Our model revealed the presence of 17 different protein subunits in the PSII core complex (Appendix, Table S2), including PsbA (D1), PsbB (CP47), PsbC (CP43), PsbD (D2), PsbE, PsbF, PsbH, PsbI, PsbK PsbL, PsbM, PsbO, PsbTc, PsbTn, PsbW, PsbX and PsbZ. An overview of the modelled proteins and their location in the Arabidopsis PSII monomer is shown in **Fig 1**. Comparing this model to the Pea and Spinach models, the spatial distribution of the core proteins did not reveal any dramatic change, which suggests that the PSII core is structurally and compositionally very well conserved among higher plants. The light-harvesting proteins CP24, CP26, CP29 and four LHCII trimers were identified and modelled (Appendix, Table S3). The highly resolved features of the map, especially of the core complex proteins, allowed

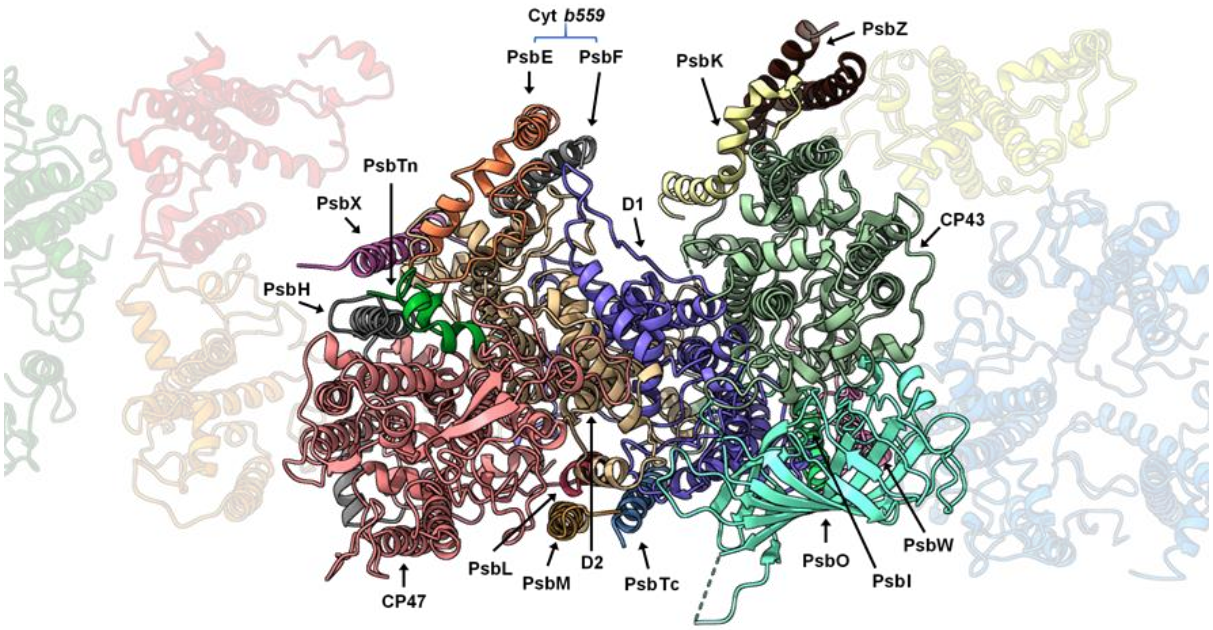

Figure 1. Localization of the Photosystem II core proteins. An entire PSII monomer is displayed in first plane, as seen from the lumen side of the thylakoid membrane, with the peripheral antennae in the background.

that several new structural observations could be made including: 35 N-terminal amino acids of the D1 protein, lacking on the previous medium-resolution Arabidopsis PSII model (PDB: 5MDX), were fully modelled; similarly, N- and C-termini of D2 were modelled to a greater extent when compared to 5MDX; the nuclear-encoded extrinsic protein PsbTn; per PSII monomer, one chloride (Cl^-) ion was found bound to the complex (Appendix, Fig S4) in the vicinity of the OEC; the vast carotenoid pigment network and the lipids present in PSII; and 141 water molecules were modelled.

The main four core proteins—D1, D2, CP47 and CP43—bind together 73 cofactors of which 35 are chlorophylls *a* (Appendix, Fig S5). The current resolution enabled us to model the correct spatial orientation of chlorophylls' chlorin rings and often the full extension of their hydrophobic tails. Unobserved in the previous medium-resolution map of Arabidopsis PSII, a density for CP29 chlorophyll *a* at position 616 is clearly visible in our EM map.

In total, 162 chlorophyll *a*, 60 chlorophyll *b*, four pheophytin *a*, and two heme molecules were modelled. Additionally, 60 lipids (**Fig 2A**), 42 carotenoids (Appendix, Fig S6), and two bicarbonate molecules were visible in our EM map and for the first time described in an Arabidopsis PSII model (Appendix, Table S4). Besides the cofactors mentioned above, 22

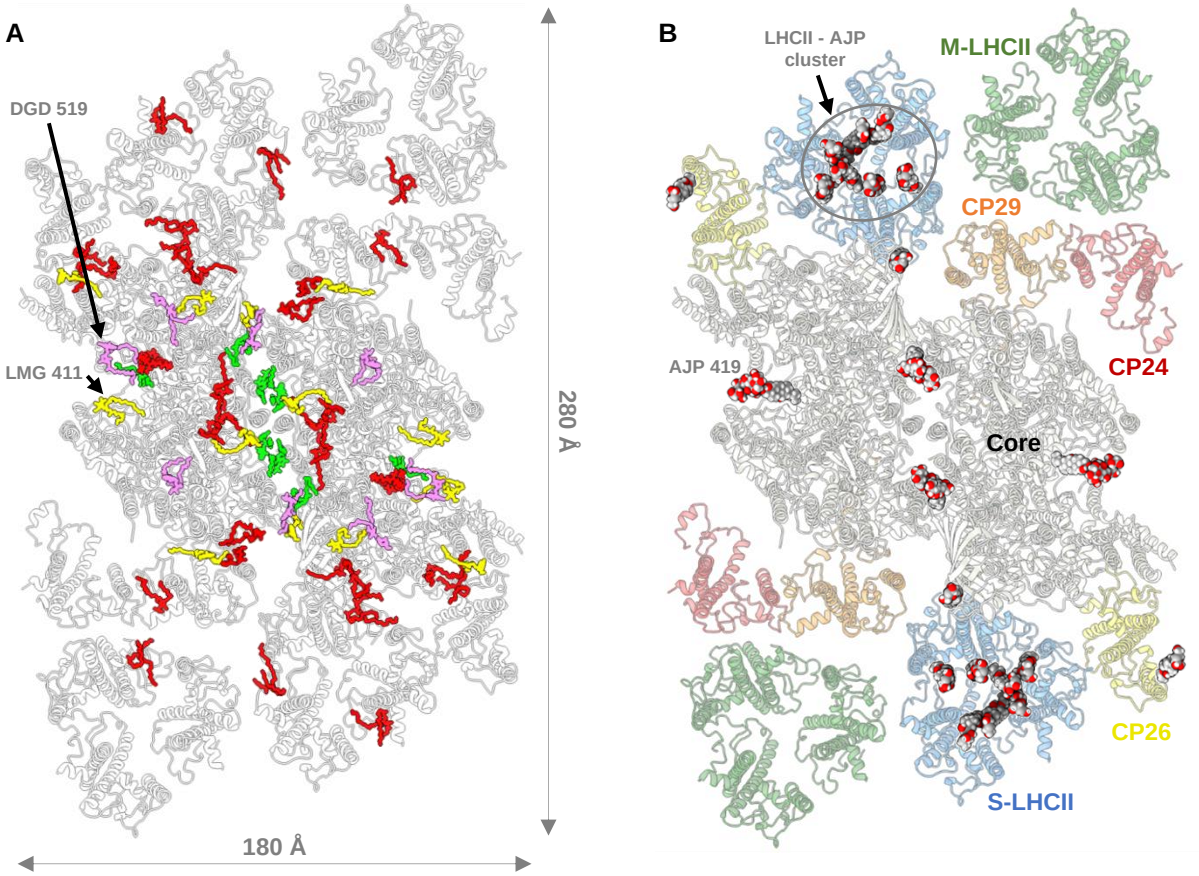

Figure 2. A, Distribution of the lipids in Arabidopsis PSII, as seen from the lumenal side: monogalactosyldiacylglycerol (MGDG/LMG, in yellow), digalactosyldiacylglycerol (DGDG/DGD, in violet), sulfoquinovosyldiacylglycerol (SQDG/SQD, in green) and phosphatidylglycerol (PG/LHG, in red). **B**, Distribution of the modelled digitonin molecules (ligand code: AJP) in our PSII supercomplex, as seen from the lumenal side. The secondary structure of the protein backbones is represented in cartoon style and the detergent molecules represented by spherical atomic representation.

digitonin molecules were identified and modelled (**Fig 2B**). The well-defined densities of these detergent molecules and those of the several PSII cofactors present in our model, demonstrate the quality of the obtained EM map (Appendix, Fig S7).

Of the four extrinsic proteins (PsbO, PsbP, PsbQ, and PsbTn), PsbO and PsbTn were detected in our EM map. The Arabidopsis genome encodes two isoforms of the PsbO protein, referred to as PsbO1 and PsbO2, that share 90% sequence identity. PsbO1 is expressed at a higher level than its counterpart and it is suggested to be bound to PSII and involved in its activity[42]. With a high-resolution structure of PSII we can differentiate between some of the non-conserved residues between the two isoforms. The densities corresponding to PsbO residues at positions 128 and 187, fit better to a proline and an asparagine as in PsbO1, than the corresponding

alanine and lysine residues of PsbO2. Therefore, we have modelled PsbO in accordance with PsbO1 sequence (Appendix, Table S2). Nevertheless, the local resolution of our EM corresponding to PsbO is lower than the PSII core, suggesting that the entire protein has some flexibility or that a lesser percentage of PSII supercomplexes may indeed bind the PsbO2 isoform.

The analysis of individual 2D and 3D classes, the detection of PsbTn with low occupancy, and the absence of PsbP, PsbQ and the neighbouring transmembrane protein PsbR, suggest that the binding of these proteins is highly sensitive to protocol variations. PsbP, PsbQ and PsbR subunits were also reported missing in the previous Arabidopsis PSII model and in the unstacked model of Pea PSII. The absence of these proteins is probably owing to thylakoid membrane solubilizations performed at pH 7.5 using high concentrations of α-DDM or lower concentrations of DDM with additional digitonin detergents (Appendix, Table S5). We do not disregard that the strength of the binding of PsbP and PsbQ to the PSII complex may differ between land plants. This argument is supported by the presence of these subunits (in substoichiometric amount) in the Pea PSII electrostatic potential map, extracted at pH 7.5[22].

In contrast to the previous medium-resolution Arabidopsis PSII cryo-EM structure from PsbS knock-out mutant[33], the hereby presented atomic model is the first PSII structure derived from wild-type phenotype of *Arabidopsis thaliana*. Although the PsbS protein was present in the starting BBY sample, we were not able to detect any density corresponding to this protein in any of the investigated 2D projections or the final EM map. Thus, the location of the enigmatic energy quenching[43] PsbS protein remains to be determined.

Of the major antennae complexes, we identify two strongly bound LHCII (S-LHCII) and two moderately bound (M-LHCII). S-LHCII contains three monomers A, B and C ($S\text{-}LHCII_A$, $S\text{-}LHCII_B$ and $S\text{-}LHCII_C$); these are within exciton energy transfer (EET) distance (~20Å) to other PSII subunits at four different places, $S\text{-}LHCII_B$ to CP26, $S\text{-}LHCII_A$ to CP43 and CP29, and $S\text{-}LHCII_C$ to the adjacent $M\text{-}LHCII_1$ (Appendix, Fig S5 and Table S6), which probably explains the strongly bound features of this LHCII subunit.

In general, the region around M-LHCII trimers and CP24 could not be easily resolved due to the: higher flexibility of this region in relation to the core of the complex; diminished

occupancy arising from the heterogeneity of the sample with an unquantifiable number of C_2S_2 particles present in our dataset; destabilisation caused by the used detergent mix. The low resolution of the M-LHCII-CP24 region, did not allow to determine which of the *lhcb*1, *lhcb*2 or *lhcb*3 gene product[44] correspond to which LHCII monomer. M-LHCII trimers—monomers modelled according to Lhcb3 (in contact with CP24) and two Lhcb1.4 protein sequences—and associated CP24 protein were placed into their respective densities. Besides their relative location to the core, detailed structural information cannot be inferred from the current EM map thus, ligands were only modelled for unequivocal densities. We note that divergences at the sequence level may occur since our EM map does not present resolved side-chain densities for the M-trimer Lhcb proteins.

The peripheral low molecular mass proteins of the PSII core

Three low molecular mass proteins, PsbW, PsbH and PsbZ, were found located in-between the PSII core and the antenna complexes (**Fig 1**). The PsbW protein links the LHCII monomer A to CP43[45]; meanwhile, the PsbH is bridging the CP29 to CP47, and the PsbZ bridging the CP26 to the CP43 protein. As none of them has been shown to bind any pigment, likely, they are not directly involved in energy transfer from the antenna to the reaction centre. Thus, suggesting that their primary function is to support the association of the antennae to the core of the PSII complex.

The well-resolved side chains densities for the PsbW protein enabled us to fully fit the N-terminal tail and model its remaining residues; and address the unknown structure of its C-terminus (**Fig 3**). We concluded that the N-terminal tail was misplaced in the previous medium-resolution PSII model and identified the amino acids PsbW-Asn99, PsbW-Trp103, and LHC_A-Asn122 (Appendix, Fig S8), as being the potential motifs that anchor the S-LHCII trimer to PsbW through electrostatic interactions.

PsbW is suggested to interact with the PsbH subunit of the opposite PSII monomer and in this way participate in the stabilisation of the PSII dimer[46]. This interaction of electrostatic nature is supported by our finding that the negatively charged C-terminal tail of the PsbW protein and the positively charged N-terminal tail of PsbH are both on the stromal side close to each

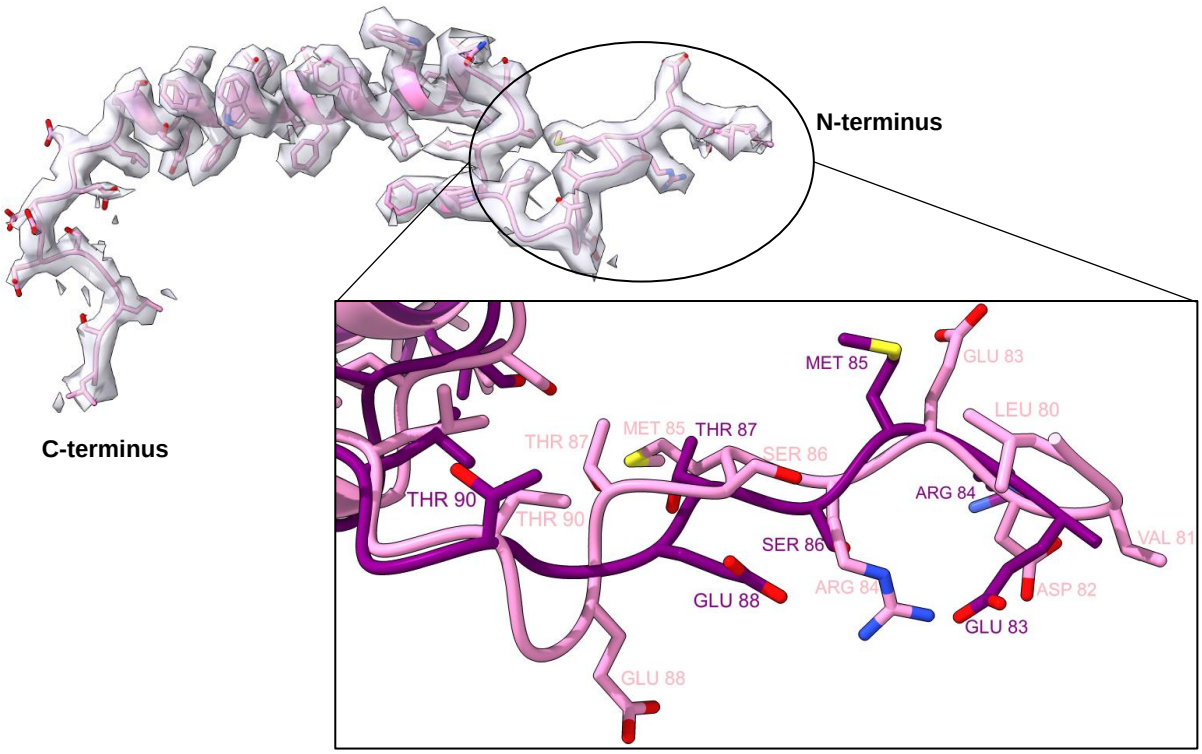

Figure 3. The low molecular weight protein PsbW fit into its respective density. Inset: Dissimilar modelling of the N-terminus of PsbW between #??? and 5MDX (purple) Arabidopsis PSII models

other. Phosphorylation of PsbH would in this case hinder such interaction, triggering the disassembly of PSII and preparing it for repair[47]. Although the C-terminal tail of PsbW adopts a conformation that follows in the opposite direction of PsbH, this possibility cannot be excluded since no cryo-EM map of higher plant PSII model was able to resolve the 12 amino acids of PsbH N-terminal tail.

We used the sequence motif PSLK (Pro81-Lys84) of the PsbX protein and the corresponding side chains densities, to correctly place the subunit and resolve four additional amino acids of its N-terminus (Appendix, Fig S9). Our placement of PsbX is shifted four to five amino acids (≈ 7.9 Å, alpha-carbon shift) towards the stromal side compared to the previous medium-resolution Arabidopsis PSII model.

Digitonin impact several oligomerization interfaces between PSII proteins

The superposition of Pea and Spinach PSII structures to our PSII model reveals that the distance between the two PSII monomeric cores is longer in Arabidopsis PSII. This is probably explained due to the presence of two digitonin molecules at the monomer-monomer interface, affecting the intermonomer distance (**Fig 4** and Appendix, Fig S10-11). However, the stability of the dimer does not seem to be affected as the purification of a homogenous sample of dimerised PSII was possible.

Digitonin molecules have a hexacyclic structure—a lipophilic steroidal sapogenin[48]—that grants partial rigidity to the molecule and result in a high-level of occupancy in the EM map (Appendix, Fig S7). In contrast, the hydrophilic oligosaccharide chains of digitonin stick outside the membrane plane leading to a higher degree of flexibility becoming practically unresolved.

The broadly accessible luminal pocket at the centre of each S-LHCII trimer contains at least seven digitonin molecules (Appendix, Fig S12) that destabilise the natural configuration of the LHCII complexes. We found that the LHCII trimers lack the violaxanthin pigments—ligand code XAT1622—reported to be located at the interface between the monomers[20,22]. In each Lhcb monomer, a digitonin molecule was found to disrupt the interaction between the violaxanthin terminal ring located on the inner side of the LHCII trimer and the chlorophyll *b* CHL607. The opposite terminal ring of the violaxanthin molecule—facing the outside-side of the LHCII trimer—is known to interact with the phosphatidylglycerol (PG) lipid LHG2630. Studies show that LHG2630 plays an essential role in the trimerization of Lhcb proteins[49]. Our EM map indicates that the hydrophobic chains of LHG2630 adopt a new conformation while the molecule remains tightly bound to the complex (Appendix, Fig S12). This interaction is sustained by hydrogen bonds and an ionic interaction between the PG head group and the amino acids Tyr78 and Lys217, respectively. Our structural data confirms that this phospholipid interacts with chlorophyll *a* molecule CLA611 through coordination between its phosphodiester group and the chlorophyll central magnesium atom. Studies of PSII dimerization show that the trans-hexadecanoic fatty acid chains of PG and not its phosphodiester group are the essential factor at the oligomerisation interface[50]. Our finding suggests that one should not infer the same conclusions for PG lipids at other positions of the

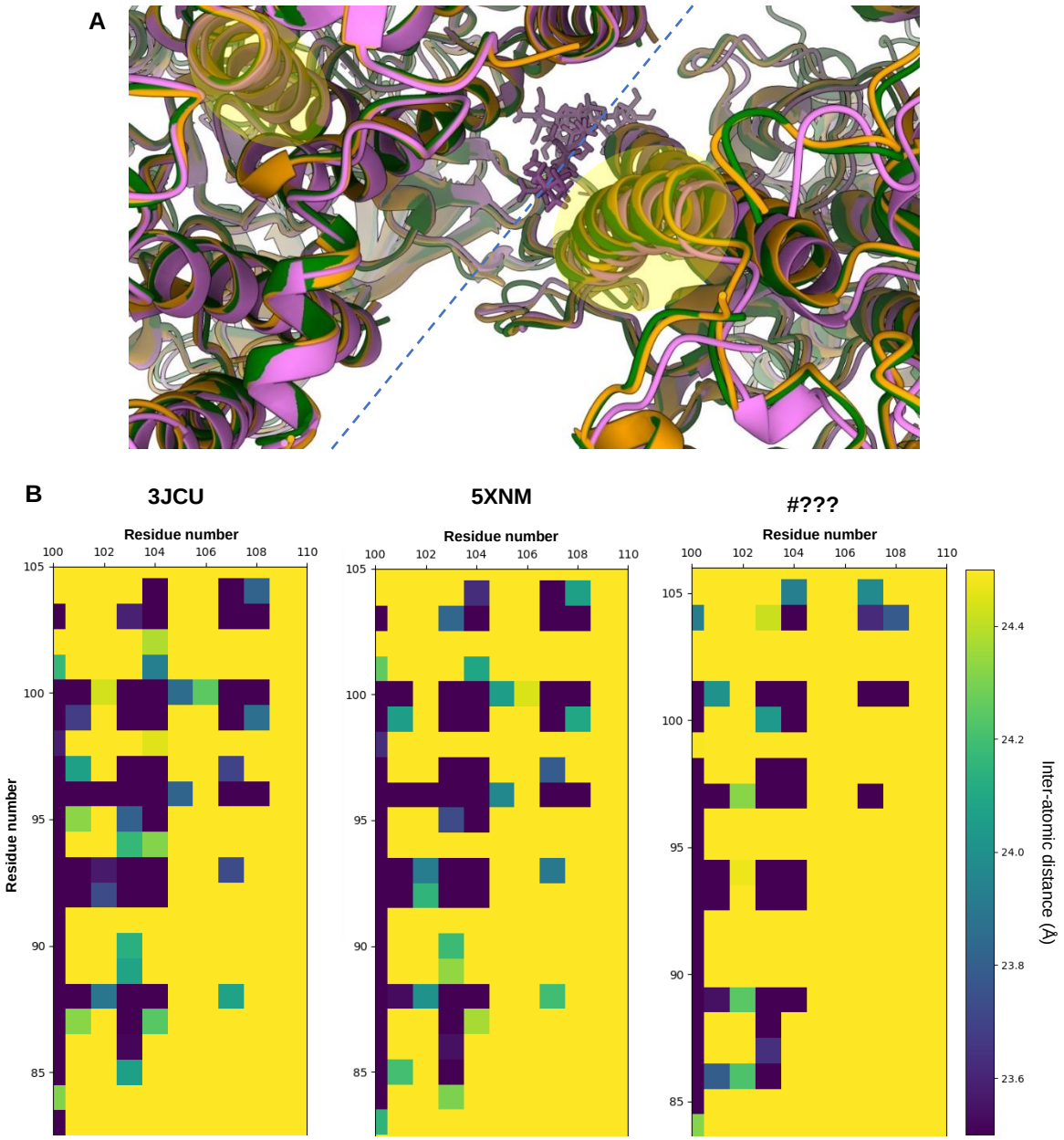

Figure 4. **A**, Dimerization interface between monomeric PSII cores: 3JCU in green, 5XNM in orange, #??? in violet. At the dimerization axis (represented in blue dashed line) a digitonin molecule can be found on model #???. Models were aligned to the D1 protein of the same monomer (to the left of the dimerization interface) **B**, Heatmaps reporting the calculated intermonomer Cα-Cα atomic distances, for each high-resolution higher plant PSII model, between α-helix B of the D1 protein (xx axis) and α-helix B of the CP47 protein from the opposite PSII monomer (yy axis)—highlighted in **A**.

PSII complex. The head group of LHG2630 lipids is crucial to stabilise the monomer-monomer interfaces in LHCII and ligates CLA611. CHL601, which in Pea and Spinach PSII structures is also in contact with XAT1622, remains bound to PSII.

Further investigation of our EM map allowed us to observe that a digitonin molecule occupies the binding site of the digalactosyldiacylglycerol (DGDG) lipid DGD520 found in Pea and Spinach PSII (**Fig 2B**). Due to the loss of interaction with DGD520, the neighbouring galactolipid DGD519 has a high degree of flexibility. In Pea PSII, DGD519 hydrogen bonds with DGD520 and the CP43 loop E (Leu401-Ala416), which explains the flexibility of this loop (**Fig 5**). Furthermore, the digitonin-replaced lipid, DGD520, is responsible to mediate the connection between the PsbJ C-terminus and the PSII core through hydrogen bond interactions. This structural finding can explain the loss of PsbJ subunit in our structure.

It should be noted that the cofactors LHG2630, XAT1622, DGD519 and DGD520 were not observed in the 5MDX model.

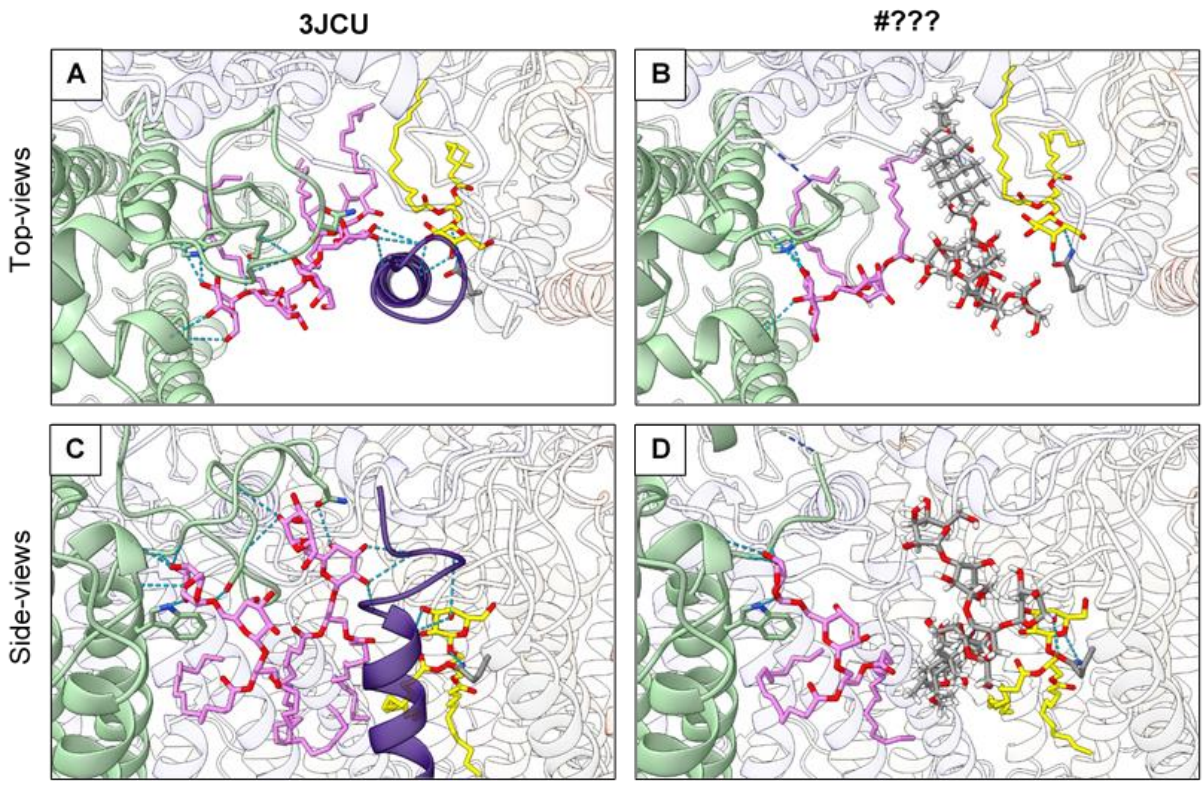

Figure 5. Hydrogen-bonding between loop E of CP43 (in light green) and PsbJ (in purple) subunit are mediated by MGDG (in yellow) and PG (not represented) lipids and disrupted by a digitonin molecule. A and B are top-views, while C and D represent views along the membrane plane, centered around the DGD520 lipid (in pink) present in 3JCU, but replaced by a digitonin molecule in #???. Hydrogen bonds are represented by dashed blue lines and the interrupted loop E of CP43 is represented by a dashed-line half-coloured in blue (N-terminus) and light green (C-terminus).

The PsbO short-loop is disordered as result of the digitonin bound between PSII monomers

The extrinsic protein PsbO has two loops—here termed short-loop (Ala154 to Pro156) and long-loop (Ser239 to Ala279)—to which structural information have been difficult to obtain for Arabidopsis PSII (Appendix, Fig S13). In Spinach PSII[22], which retains the extrinsic PsbP and PsbQ proteins, both loop regions are well structured. The unstacked PSII structure from Pea (PDB: 5XNM) has the short-loop modelled but the long-loop, which is contiguous to PsbP in Spinach PSII and in Pea stacked PSII (PDB: 5XNL) models, is absent.

The increased quality of our EM map revealed three additional amino acids in the short-loop N-terminal with a conformation dissimilar to their counterpart amino acids in Pea and Spinach models. The amino acid sequence of the short-loop peptide is the poorest conserved region of the PsbO protein among higher plants (Appendix, Fig S14), suggesting that this flexible region is of lower functional relevance. Nevertheless, it is conserved to a higher degree in Spinach and Pea, while being the least conserved in Arabidopsis PsbO, justifying the diverged structural arrangement (Asn153-Glu157, Appendix, Fig S15) of the peptide backbone including the newly modelled amino acids. The hydrophilic sugar chains of the digitonin molecules harboured between PSII monomers are within no distance to interact with PsbO short-loop, and thus could be the reason for its disorder.

To note that we do not exclude the possibility that an even larger number of detergent molecules are bound to the complex, even if they cannot be observed due to their high flexibility.

The disordered PsbO long-loop is related to several structural changes at the OEC

In the here presented PSII structure, the C-terminus of the D2 protein was found to be twisted (Appendix, Fig S16) in a similar manner to what was reported for the cyanobacterial *Synechocystis* Apo-PSII[40]—making the last C-terminal residues occupy the region where the otherwise ordered PsbO long-loop would be located.

303 The PsbO-Asp243 (Asp158, according to 5XNM sequence numbering) is a residue of PsbO
304 long-loop known to participate in a hydrogen-bonding network that connects the Mn_4CaO_5 to
305 the lumen and stabilise the binding of PsbP and PsbQ proteins[51]. Interestingly, the 5XNM
306 model presents the long-loop modelled to a higher extent (six additional amino acids at the
307 C-terminus, including Asp158). Our analyses of hydrogen-bonding protein-protein
308 interactions of the 5XNM model show that Asp158 and Leu157 hydrogen-bond residues D1-
309 Arg334 and D1-Asn335, respectively. These interactions are responsible to maintain the
310 structure of the D1 C-terminal and consequently keep the Mn_4CaO_5 in its correct position.

311 Considering that the long-loop region of PsbO is highly conserved in almost all higher plants,
312 we can assume that the observed disorder in Arabidopsis PSII models is linked to the loss of
313 the extrinsic proteins PsbP and PsbQ. Nevertheless, it is unclear why the C-terminus of the
314 D2 protein in our PSII structure is twisted and why the unstacked PSII from Pea has a more
315 ordered PsbO long-loop, and a structured D1 C-terminus capable of binding the inorganic
316 cluster. We hypothesize that both observations are correlated to each other, and a possible
317 consequence of the use of digitonin as a solubilization agent.

318

319 The OEC in the absence of PsbP and PsbQ proteins

320 The Mn_4CaO_5 cluster is coordinated by crucial amino acids of the D1 protein and the core
321 antennae protein CP43, often referred to as "first shell" residues. Amino acids from the D2
322 protein and the inner antenna CP47 are essential "second shell" residues that: sustain the
323 structure of the OEC and form the channels which deliver H_2O molecules to the water
324 oxidation cluster. The extrinsic subunits are known to stabilise the hydrophilic pocket that

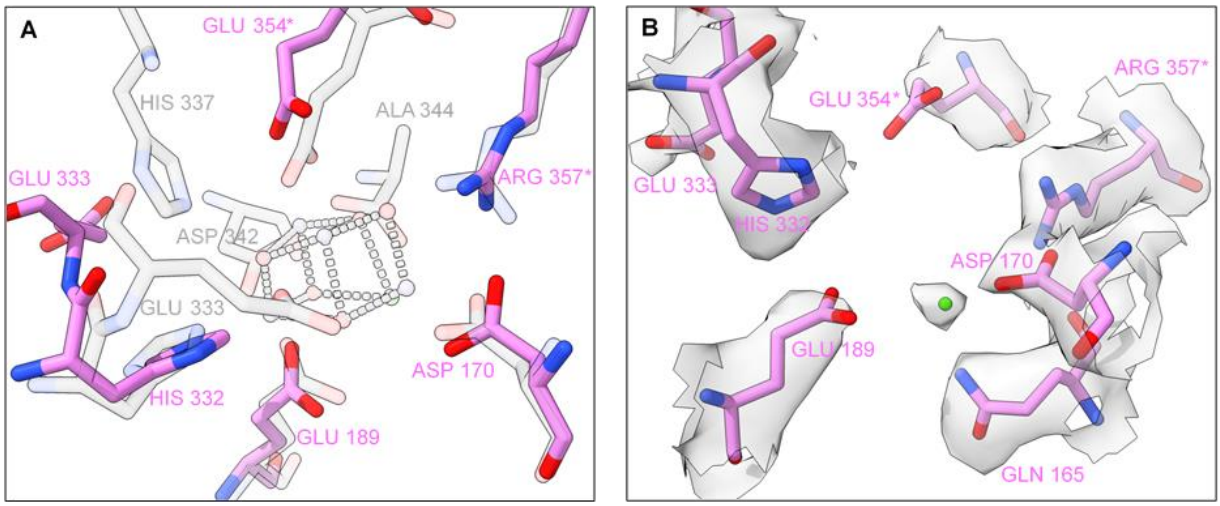

Figure 6. The manganese cluster binding site. A, Amino acid ligands to Mn_4CaO_5: Spinach PSII (3JCU, transparent grey) superimposed to Arabidopsis PSII (#???, pink). To avoid double labelling, the overlapping residues of 3JCU are not identified for clarity. B, Arabidopsis amino acid ligands to Mn_4CaO_5 fit into their respective densities and an extra density attributed to a positive metal ion (in our structure modelled as manganese). *Identifies the amino acids belonging to the CP43 protein, while the remaining apply to the D1 subunit.

hosts the inorganic cluster. Comparing our model to the 3JCU mature PSII structure containing the extrinsic proteins and the Mn cluster, we found that in the absence of PsbP and PsbQ proteins, the Mn_4CaO_5 binding pocket undergoes conformational changes. Without the presence of the Mn_4CaO_5 cluster, the C-terminal tail of the D1 protein (amino acids His337-Ala344) was found to be disordered and therefore not visible in our EM map. Out of the seven amino acid that ligate to Mn_4CaO_5, only two, D1-Asp170 and D1-Glu189, retain their relative positions to the mature Spinach PSII structure (**Fig 6A**). D1-Glu329 adopts an unusual conformation, where it is found to hydrogen bond with D1-His332. Glutamate 333 of the D1 protein, one of the coordinating residues of the manganese cluster, is consequently in a different conformation (**Fig 6B**) and hydrogen-bonded to D1-Glu329. The C-terminus of the D1 protein (amino acids 337-344) is disordered and not visible in our EM map.

The core antenna protein CP43 consists of six transmembrane helices connected by short loops (loop A-D) and a very large extrinsic domain known as loop E. Glutamate 354 and arginine 357, also ligands of the Mn_4CaO_5, are part of loop E. Both amino acids nearly conserve their position relative to mature PSII (**Fig 6A**). The PsbO seems to play an essential role in stabilising the CP43 loop E—connected by 10 interprotein hydrogen bonds—that otherwise would have high degree of flexibility. Finally, near the water oxidation cluster binding site,

342 loop E is stabilised via hydrogen bonds between CP43-Arg357 and D1-Asp170. A careful
343 analysis reveals that CP43-Glu354 is slightly receded from its coordinating position to Mn2
344 and Mn3 of the Mn_4CaO_5, but pre-organised to receive and coordinate the manganese ions
345 (**Fig 6A**). At the expected position of Ca^{2+} ion on a mature PSII structure, we find a density
346 that can be attributed to a positively charged ion (**Fig 6B**), potentially coordinated by the D1
347 residues Gln165 (also retaining its position related to mature PSII), Asp170 and Glu189. EM
348 data by itself offers no signifying conclusion regarding the identity of the metal ion, which
349 could be a Ca^{2+} ion, or the first manganese ion that binds to PSII to form the Mn_4CaO_5. If the
350 latter is true, this binding site might well be the long sought high-affinity site of PSII oxygen-
351 evolving complex.

352

353 # Discussion

354 In this work, we presented the first high-resolution structure of Photosystem II from wild-
355 type *Arabidopsis thaliana*, allowing a detailed comparison to other higher plants' PSII
356 structures, while fairer comparison to existing high-resolution cyanobacterial PSII models.
357 The quality of the obtained EM map enabled a better and more comprehensive modelling of
358 most of the PSII core proteins; a detailed analysis of the positioning of chlorophyll, carotenoid,
359 and lipid molecules; and the placement of water molecules, of which some compose the water
360 channels delivering substrate to the catalytic site of this water-splitting enzyme.

361 To isolate PSII, we solubilised thylakoid membranes with a detergent mixture containing β-
362 DDM—a harsher isoform of the more commonly used α-DDM[32]—and digitonin. Even though
363 β-DDM can be a more disruptive detergent hindering the obtention of large PSII
364 supercomplexes, we did not find evidence of any structural changes caused by it. Instead, we
365 found that digitonin molecules disrupt several regions of PSII by replacing lipids that keep
366 the structural integrity of the membrane protein complex. Exploiting the digitonin effects on
367 the major antennae, we found that the PG phosphodiester group is the essential lipid element
368 to stabilise the trimerization of LHCII complexes. Furthermore, we found that digitonin
369 molecules are responsible for: the loss of the PsbJ subunit; the consequent instability of a CP43
370 loop; the disordered D1 C-terminus and the resultant fall of the Mn_4CaO_5 cluster. Besides

revealing new knowledge, our model is a representation of the disruptive mechanisms that digitonin molecules can have in the stability of protein complexes, leading to a dramatic loss of function.

The PsbJ protein is one of the most hydrophobic proteins in the thylakoid membrane and has been suggested to be involved in the electron flow within PSII[52] even though it does not seem to bind any pigment or cofactor. It is found in Pea and Spinach structures[20,22] bound to the PSII complex through hydrogen bonding with lipids associated with CP43 protein, but neither found in our Arabidopsis PSII structure, nor in the 5MDX model. Regarding the PsbY protein, we do not detect any densities suggesting the presence of this low molecular mass subunit—as for the PSII models of Spinach (3JCU) and Pea (5XNM and 5XNL)—even when western-blotting analysis show that the protein is present in various PSII samples from higher plants including Arabidopsis[53]. So far, the structure of the PsbY subunit has only been observed on cyanobacterial PSII models[29], which in higher plants $C_2S_2M_2$-type PSII would be located on the outskirts of the complex (considering a conserved position). The small proteins PsbY and PsbJ are both hydrophobic, have a single transmembrane helix and are suggested to exist in the outskirts of the complex, thus likely to be dissociated from the complex in the presence of detergents.

As previously suggested[54], the low molecular mass protein PsbJ might be required to stably bind PsbR, which in turn is responsible for protecting PsbP and PsbQ from proteolytic degradation. PsbO is known to be the most stable extrinsic protein, weakly sensitive to the depletion of the neighbouring PsbP and PsbQ proteins. We anticipate that the PsbO long-loop requires PsbP to maintain its stability, which is correlated with the loss of the transmembrane subunits PsbJ and PsbR, and the extrinsic proteins PsbP and PsbQ. Furthermore, the oxygen-evolving proteins PsbP and PsbQ are pH-sensitive[20], and both proteins bind the strongest to the PSII lumen side around pH 6[31]. Unfortunately, PSII complexes tend to aggregate at pH 6, obstructing single-particle cryo-EM studies. The isolation of PSII complexes at pH 6 using alternatives to detergents, such as amphiphilic[17] or styrene-maleic acid (SMA) copolymers forming membrane proteins embedded in nanodiscs, might offer a solution to these challenges. Not less crucial, such copolymers would protect PSII from the undesirable effects of detergents and allow to extract PsbJ (and possibly PsbR and PsbY) bound to PSII.

In comparison to other PSII structures, new knowledge on the assembly of the OEC from higher plants can possibly be derived from our model which exhibits similar characteristics to PSII still in assembly stage, prior to the binding of the extrinsic PsbP and PsbQ: the monomeric *Synechocystis* apo-PSII[40] presents analogous features to the high-resolution Arabidopsis PSII, particularly at the Mn_4CaO_5 binding-site.

These findings allow us to step forward in PSII research and bring new research questions: Could the detergent molecules anchored inside the moiety of isolated chlorophyll-binding proteins alter their spectroscopic properties; how has that influenced the outcomes of different PSII and LHCII spectroscopical studies; and if digitonin molecules fit into the centre of the LHCII trimer, did they replace other naturally present molecules such as yet undetermined lipids at those positions?

Materials and Methods

Photosystem II supercomplex extraction

Arabidopsis thaliana (ecotype Columbia-0) plants were grown on soil in greenhouse at Umeå University (Umeå, Sweden) for 8 weeks under controlled temperature (23°C) and humidity (70%) conditions, with 8 h light / 16 h dark photoperiod. Seeds were obtained and identified by Arabidopsis Biological Resource Center (ABRC; https://abrc.osu.edu/). Local and national regulations was obeyed and the permission from Jordbruksverket - Växtkontrollenheten has been obtained (Dnr 4.6.20-6365/15).

Plants were harvested after a dark period of 16 h and BBY membranes[41] prepared according to previous protocol[55]. The extraction of PSII $C_2S_2M_2$ complexes was performed[31] with the following modifications: solubilisation was performed by adding a detergent mixture of 0.2% β-DDM and 1% digitonin, followed by a 30 min dark incubation at 4°C.

Sample preparation and electron microscopy

427 For cryo-electron microscopy, solubilised PSII sample was washed with a 10mM HEPES-KOH
428 pH 7.5 + 0.01% w/v digitonin solution (bellow CMC level) and concentrated to 1 mg/ml Chl
429 using an Amicon Ultra 30k centrifugation filter (Millipore). 4 µl concentrated sample was
430 applied to glow-discharged (30 s at 50 mA) Quantifoil R 1.2/1.3 Cu 300 grids (Quantifoil Micro
431 Tools) and grids were plunge frozen in liquid ethane using an FEI Vitrobot MkIV (Thermo
432 Fisher Scientific) at 100% humidity, 4°C, using a blot force of −5, wait time of 1 second, and
433 blotting time of 5 seconds. Automated data collection was performed using EPU software on
434 a Titan Krios G2 transmission electron microscope (Thermo Fisher Scientific) operated at 300
435 kV, equipped with a Gatan K2 Summit direct electron detector and Gatan Quantum GIF
436 energy filter at the Umeå Core Facility for Electron Microscopy, a node of the Swedish
437 National Cryo-EM facility. In total 12227 movies, fractionated in 40 frames, were collected at
438 a pixel size of 0.82 Å, total dose of 59.7 $e^-/Å^2$ and defocus range of -1.5 to -3.0 µm. On the same
439 microscope and grid, a data set was collected using Volta phase plate[56]: in total 3585 movies,
440 fractionated in 40 frames, were collected at a pixel size of 1.09 Å, total dose of 36.7 $e^-/Å^2$ and
441 defocus of -0.7 µm.

442

443 **Cryo-EM data processing, model building and refinement**

444 The collected datasets were processed using the software package cryoSPARC[57] (version
445 v2.15).

446 VPP dataset: Automated particle picking based on 116 manually picked templates was carried
447 out before the major 2D classification round. The combination of well-resolved 2D classes
448 isolated a subset of 80456 particles, out of 273202 particles automatically picked. The *ab-Initio*
449 3D reconstruction strategy segregated particles into 2 different 3D-classes, allowing us to
450 proceed with one class of a more homogeneous set of 56322 particles. Homogeneous
451 refinement of the preliminary 3D volume was carried with imposed C2 symmetry, yielding a
452 3D reconstruction with an estimated average resolution of 3.6 Å.

453 High-resolution data set: Template-free automated particle picking was carried out before the
454 major 2D classification round. From the selected 2D classes we restricted the data set to a total
455 of 110 659 particles, out of 416 262 particles automatically picked. The *ab-Initio* 3D

456 reconstruction strategy segregated particles into 4 different 3D-classes, allowing us to proceed
457 with a more homogeneous subset of 100 712 particles. The selected particles were individually
458 motion-corrected, and homogeneous refinement of the preliminary 3D volume was carried
459 with imposed C2 symmetry. After a round of per-particle CTF refinement, the final
460 homogeneous refinement yielded a 3D reconstruction with an estimated average resolution
461 of 2.79Å, using the gold standard FSC (Appendix, Fig S1).

462 Three independent post-processing methods were applied to the resultant high-resolution
463 electrostatic potential map: local sharping with Phenix Autosharpen tool[58] (Phenix v1.18.2);
464 density modification with Phenix denmod tool[59] (Phenix v1.18.2) using the outputted half-
465 maps by cryoSPARC; application of the negative of the Laplacian operator using Chimera[60]
466 (version 1.14). The three resulting post-processed maps were useful to confirm certain
467 densities, such as the presence of ions and water molecules.

468 For model building, the previous Arabidopsis atomic model 5MDX was manually docked and
469 each chain morph-fitted to the 2.79 Å map using COOT[61] (version 0.92). The new cofactors
470 and the PsbTn subunit where found using the 5XNM Pea PSII model as a template. Overall
471 the presented model was manually refined using COOT and both the high-resolution and the
472 VPP dataset, and automatically refined with Phenix Real-space refinement tool[62] (Phenix
473 v1.18.2).

474

475 **Model analysis and preparation of figures**

476 Molecular graphics and analyses were performed with UCSF ChimeraX[63] (version 1.1),
477 developed by the Resource for Biocomputing, Visualization, and Informatics at the University
478 of California, San Francisco. Surface electrostatics were calculated using the software
479 packages PDB2PQR (to assign atomic charges and radii to the protein backbone using force
480 field parameters) and APBS (to generate the electrostatic surface)[64]. We recommend ChimeraX
481 for model and surface visualisation.

Acknowledgements

The authors thank to Dr Johannes Messinger, Dr Thomas Kieselbach and Dr Domenica Farci for reviewing the manuscript. This work was supported by grants from the Carl Trygger foundation to WPS, and the Swedish Research Council (VR), grant 2016-05009 to KP. The data was collected at the Umeå Core Facility for Electron Microscopy, a node of the Cryo-EM Swedish National Facility, funded by the Knut and Alice Wallenberg, Family Erling Persson and Kempe Foundations, SciLifeLab, Stockholm University and Umeå University.

Author contributions

A.T.G. and W.P.S. designed research; A.T.G performed sample preparation, data processing, and model building; A.T.G and M.H. performed cryo-EM data collection; A.T.G and W.P.S. wrote the paper; M.H. and K.P. assisted model building and writing of the paper.

Additional information

The executed work with plant material complies with the local and national regulations.

The authors declare no competing interests.

Data availability

The raw multi-frame micrographs of the higher-resolution data set and the VPP dataset were deposited in the Electron Microscopy Public Image Archive with accession codes EMPIAR-##### and EMPIAR-#####, respectively. Electrostatic potential maps used for refinement, unsharpened EM map, half-maps and other relevant post processed maps have been deposited in the Electron Microscopy Data Bank, EM-####. The modelled structure is available in the Protein Data Bank (PDB) under the accession code #???.

References

1. Stirbet, A., Lazár, D., Guo, Y. & Govindjee, G. Photosynthesis: basics, history and modelling. *Ann. Bot.* **126**, 511–537 (2020).
2. Beadle, C. L. & Long, S. P. Photosynthesis — is it limiting to biomass production? *Biomass* **8**, 119–168 (1985).
3. Tkemaladze, G. S. & Makhashvili, K. A. Climate changes and photosynthesis. *Ann. Agrar. Sci.* **14**, 119–126 (2016).
4. Koornneef, M. & Meinke, D. The development of Arabidopsis as a model plant. *Plant J.* **61**, 909–921 (2010).
5. Poczai, P., Cernák, I., Varga, I. & Hyvönen, J. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Genet. Resour. Crop Evol.* **61**, 796–815 (2014).
6. Govindjee, Kern, J. F., Messinger, J. & Whitmarsh, J. Photosystem II. in *Encyclopedia of Life Sciences* (John Wiley & Sons, Ltd, 2010). doi:10.1002/9780470015902.a0000669.pub2
7. Simpson, D. J. & Von Wettstein, D. The structure and function of the thylakoid membrane. *Carlsberg Res. Commun.* **54**, 55–65 (1989).
8. Cao, P. *et al.* Structure, assembly and energy transfer of plant photosystem II supercomplex. *Biochim. Biophys. Acta - Bioenerg.* **1859**, 633–644 (2018).
9. Barber, J., Morris, E. & Büchel, C. Revealing the structure of the photosystem II chlorophyll binding proteins, CP43 and CP47. *Biochim. Biophys. Acta - Bioenerg.* **1459**, 239–247 (2000).
10. Caffarri, S., Kouřil, R., Kereïche, S., Boekema, E. J. & Croce, R. Functional architecture of higher plant photosystem II supercomplexes. *EMBO J.* **28**, 3052–3063 (2009).
11. Nield, J. & Barber, J. Refinement of the structural model for the Photosystem II supercomplex of higher plants. *Biochim. Biophys. Acta - Bioenerg.* **1757**, 353–361 (2006).
12. Jackowski, G., Kacprzak, K. & Jansson, S. Identification of Lhcb1/Lhcb2/Lhcb3 heterotrimers of the main light-harvesting chlorophyll a/b–protein complex of Photosystem II (LHC II). *Biochim. Biophys. Acta - Bioenerg.* **1504**, 340–345 (2001).
13. Caffarri, S., Croce, R., Cattivelli, L. & Bassi, R. A look within LHCII: Differential analysis of the Lhcb1-3 complexes building the major trimeric antenna complex of higher-plant photosynthesis. *Biochemistry* **43**, 9467–9476 (2004).
14. Boekema, E. J., Van Roon, H., Calkoen, F., Bassi, R. & Dekker, J. P. Multiple types of association of photosystem II and its light-harvesting antenna in partially solubilized photosystem II membranes. *Biochemistry* **38**, 2233–2239 (1999).
15. Tokutsu, R., Kato, N., Bui, K. H., Ishikawa, T. & Minagawa, J. Revisiting the supramolecular organization of photosystem II in Chlamydomonas reinhardtii. *J. Biol. Chem.* **287**, 31574–31581 (2012).
16. Shen, L. *et al.* Structure of a C2S2M2N2-type PSII–LHCII supercomplex from the green alga Chlamydomonas reinhardtii. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 21246–21255 (2019).
17. Burton-Smith, R. N. *et al.* Structural determination of the large photosystem II-light-harvesting complex II supercomplex of Chlamydomonas reinhardtii using nonionic amphipol. *J. Biol. Chem.* **294**, 15003–15013 (2019).
18. Nosek, L., Semchonok, D., Boekema, E. J., Ilík, P. & Kouřil, R. Structural variability of plant photosystem II megacomplexes in thylakoid membranes. *Plant J.* **89**, 104–111 (2017).
19. Shi, L. X. & Schröder, W. P. The low molecular mass subunits of the photosynthetic supracomplex, photosystem II. *Biochim. Biophys. Acta - Bioenerg.* **1608**, 75–96 (2004).
20. Su, X. *et al.* Structure and assembly mechanism of plant C2S2M2 -type PSII-LHCII supercomplex. *Science (80-.).* **357**, 815–820 (2017).
21. Guskov, A. *et al.* Cyanobacterial photosystem II at 2.9-Å resolution and the role of quinones, lipids, channels and chloride. *Nat. Struct. Mol. Biol.* **16**, 334–342 (2009).
22. Wei, X. *et al.* Structure of spinach photosystem II-LHCII supercomplex at 3.2 Å resolution. *Nature* **534**, 69–74 (2016).
23. Loll, B., Kern, J., Saenger, W., Zouni, A. & Biesiadka, J. Lipids in photosystem II: Interactions with protein and cofactors. *Biochim. Biophys. Acta - Bioenerg.* **1767**, 509–519 (2007).
24. Wang, H. W. & Wang, J. W. How cryo-electron microscopy and X-ray crystallography complement each other. *Protein Sci.* **26**, 32–39 (2017).
25. Rondeau, J. M. & Schreuder, H. *Protein Crystallography and Drug Discovery. The Practice of Medicinal Chemistry: Fourth Edition* (Elsevier Ltd, 2015). doi:10.1016/B978-0-12-417205-0.00022-5
26. Umena, Y., Kawakami, K., Shen, J. R. & Kamiya, N. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9Å. *Nature* **473**, 55–60 (2011).
27. Zouni, A. *et al.* Crystal structure of photosystem II from Synechococcus elongatus at 3.8 Å resolution. *Nature* **409**, 739–743 (2001).
28. Nakajima, Y. *et al.* Thylakoid membrane lipid sulfoquinovosyl-diacylglycerol (SQDG) is required for full functioning of photosystem II in Thermosynechococcus elongatus. *J. Biol. Chem.* **293**, 14786–14797 (2018).
29. Hellmich, J. *et al.* Native-like Photosystem II Superstructure at 2.44 Å Resolution through Detergent Extraction from the Protein Crystal. *Structure* **22**, 1607–1615 (2014).
30. Müh, F. & Zouni, A. Structural basis of light-harvesting in the photosystem II core complex. *Protein Science* **29**, 1090–1119 (2020).
31. Crepin, A., Santabarbara, S. & Caffarri, S. Biochemical and spectroscopic characterization of highly stable photosystem II supercomplexes from arabidopsis. *J. Biol. Chem.* **291**, 19157–19171 (2016).
32. Pagliano, C., Barera, S., Chimirri, F., Saracco, G. &

Barber, J. Comparison of the α and β isomeric forms of the detergent n-dodecyl-D-maltoside for solubilizing photosynthetic complexes from pea thylakoid membranes. *Biochim. Biophys. Acta - Bioenerg.* **1817**, 1506–1515 (2012).

33. Van Bezouwen, L. S. *et al.* Subunit and chlorophyll organization of the plant photosystem II supercomplex. *Nat. Plants* **3**, 1–11 (2017).
34. Adir, N. Crystallization of the oxygen-evolving reaction centre of photosystem II in nine different detergent mixtures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55**, 891–894 (1999).
35. da Fonseca, P. C. A., Morris, E. P. & Büchel, C. Electron Crystallography in Photosynthesis Research. in 125–150 (2008). doi:10.1007/978-1-4020-8250-4_7
36. Cheng, Y., Grigorieff, N., Penczek, P. A. & Walz, T. A primer to single-particle cryo-electron microscopy. *Cell* **161**, 438–449 (2015).
37. Andersson, B. & Aro, E.-M. Photodamage and D1 Protein Turnover in Photosystem II. in *Regulation of Photosynthesis* 377–393 (Kluwer Academic Publishers, 2001). doi:10.1007/0-306-48148-0_22
38. Nickelsen, J. & Rengstl, B. Photosystem II assembly: From cyanobacteria to plants. *Annu. Rev. Plant Biol.* **64**, 609–635 (2013).
39. Avramov, A. P., Hwang, H. J. & Burnap, R. L. The role of Ca2+ and protein scaffolding in the formation of nature's water oxidizing complex. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 28036–28045 (2020).
40. Gisriel, C. J. *et al.* Cryo-EM Structure of Monomeric Photosystem II from Synechocystis sp. PCC 6803 Lacking the Water-Oxidation Complex. *Joule* (2020). doi:10.1016/j.joule.2020.07.016
41. Berthold, D. A., Babcock, G. T. & Yocum, C. F. A Highly Resolved, Oxygen Evolving Photosystem II Preparation from Spinach Thylakoid Membranes. *Fed. Eur. Biochem. Soc. Lett.* **134**, 231–234 (1981).
42. Lundin, B., Hansson, M., Schoefs, B., Vener, A. V. & Spetea, C. The Arabidopsis PsbO2 protein regulates dephosphorylation and turnover of the photosystem II reaction centre D1 protein. *Plant J.* **49**, 528–539 (2007).
43. Funk, C. *et al.* The PSII-S Protein of Higher Plants: A New Type of Pigment-Binding Protein. *Biochemistry* **34**, 11133–11141 (1995).
44. Jansson, S. A guide to the Lhc genes and their relatives in Arabidopsis. *Trends Plant Sci.* **4**, 236–240 (1999).
45. García-Cerdán, J. G. *et al.* The PsbW protein stabilizes the supramolecular organization of photosystem II in higher plants. *Plant J.* **65**, 368–381 (2011).
46. Shi, L. X., Lorković, Z. J., Oelmüller, R. & Schröder, W. P. The low molecular mass PsbW protein is involved in the stabilization of the dimeric photosystem II complex in Arabidopsis thaliana. *J. Biol. Chem.* **275**, 37945–37950 (2000).
47. Shi, L. X., Lorković, Z. J., Oelmüller, R. & Schröder, W. P. The low molecular mass PsbW protein is involved in the stabilization of the dimeric photosystem II complex in Arabidopsis thaliana. *J. Biol. Chem.* **275**, 37945–37950 (2000).
48. Korchowiec, B. *et al.* Impact of two different saponins on the organization of model lipid membranes. *Biochim. Biophys. Acta - Biomembr.* **1848**, 1963–1973 (2015).
49. Hobe, S., Foerster, R., Klingler, J. & Paulsen, H. N-Proximal Sequence Motif in Light-Harvesting Chlorophyll a/b-Binding Protein Is Essential for the Trimerization of Light-Harvesting Chlorophyll a/b Complex. *Biochemistry* **34**, 10224–10228 (1995).
50. Kruse, O. *et al.* Phosphatidylglycerol Is Involved in the Dimerization of Photosystem II. *J. Biol. Chem.* **275**, 6509–6514 (2000).
51. Zhu, Q. *et al.* Function of PsbO-Asp158 in photosystem II: effects of mutation of this residue on the binding of PsbO and function of PSII in Thermosynechococcus vulcanus. *Photosynth. Res.* **146**, 29–40 (2020).
52. Regel, R. E. *et al.* Deregulation of Electron Flow within Photosystem II in the Absence of the PsbJ Protein. *J. Biol. Chem.* **276**, 41473–41478 (2001).
53. Von Sydow, L. *et al.* The PsbY protein of Arabidopsis Photosystem II is important for the redox control of cytochrome b559. *Biochim. Biophys. Acta - Bioenerg.* **1857**, 1524–1533 (2016).
54. Suorsa, M. *et al.* PsbR, a missing link in the assembly of the oxygen-evolving complex of plant photosystem II. *J. Biol. Chem.* **281**, 145–150 (2006).
55. Chen, Y. E. *et al.* The low molecular mass photosystem II protein PsbTn is important for light acclimation. *Plant Physiol.* **179**, 1739–1753 (2019).
56. Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J. M. & Baumeister, W. Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15635–15640 (2014).
57. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
58. Terwilliger, T. C., Sobolev, O. V., Afonine, P. V. & Adams, P. D. Automated map sharpening by maximization of detail and connectivity. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 545–559 (2018).
59. Terwilliger, T. C., Ludtke, S. J., Read, R. J., Adams, P. D. & Afonine, P. V. Improvement of cryo-EM maps by density modification. *Nat. Methods* **17**, 923–927 (2020).
60. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
61. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
62. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 531–544 (2018).
63. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **8**, 70–82 (2020).
64. Unni, S. *et al.* Web servers and services for electrostatics calculations with APBS and PDB2PQR. *J. Comput. Chem.* **32**, 1488–1491 (2011).