Check for updates

Million-atom molecular dynamics simulations reveal the interfacial interactions and assembly of plant PSII-LHCII supercomplex†

Ruichao Mao, [ID] [a] Han Zhang, [a] Lihua Bie, [a] Lu-Ning Liu [ID] *[bc] and Jun Gao [ID] *[a]

Protein–protein interface interactions dictate efficient excitation energy transfer from light-harvesting antennas to the photosystem II (PSII) core. In this work, we construct a 1.2 million atom-scale model of plant C_2S_2-type PSII-LHCII supercomplex and perform microsecond-scale molecular dynamics (MD) simulations to explore the interactions and assembly mechanisms of the sizeable PSII-LHCII supercomplex. We optimize the nonbonding interactions of the PSII-LHCII cryo-EM structure using microsecond-scale MD simulations. Binding free energy calculations with component decompositions reveal that hydrophobic interactions predominantly drive antenna–core association and the antenna– antenna interactions are relatively weak. Despite the positive electrostatic interaction energies, hydrogen bonds and salt bridges mainly provide directional or anchoring forces for interface binding. Analysis of the roles of small intrinsic subunits of PSII suggests that LHCII and CP26 first interact with small intrinsic subunits and then bind to the core proteins, whereas CP29 adopts a one-step binding process to the PSII core without the assistance of other factors. Our study provides insights into the molecular underpinnings of the self-organization and regulation of plant PSII-LHCII. It lays the framework for deciphering the general assembly principles of photosynthetic supercomplexes and possibly other macromolecular structures. The finding also has implications for repurposing photosynthetic systems to enhance photosynthesis.

Introduction

Photosynthesis is the primary source of energy for most life on Earth. It is estimated that photosynthesis produces more than 100 billion tons of dry biomass annually, which is equivalent to 100 times the weight of the current total human population on Earth and equal to an average energy storage rate of about 100 TW.[1] The high-efficiency energy conversion is governed by the physiological coordination and arrangement of photosynthetically active chromoprotein complexes, which were generated during more than 3.5 billion years of evolution.[2] Understanding the assembly and energy transfer mechanisms of photosynthetic protein complexes is pivotal for advancing our knowledge of the fundamentals of photosynthesis and underpinning the development of artificial photosynthesis to enhance and modulate energy conversion.[3]

Photosystem II (PSII) is the only enzyme that catalyzes water splitting in oxygenic photosynthesis and has been a primary target in designing biomimetic photosynthetic systems.[4] Our knowledge about the assembly principles of the PSII core has been greatly improved by studies on high-resolution structures (up to 1.9 Å) and functions.[5–7] Moreover, substantial developments have been made recently in PSII-based hybrid systems and the utilization of PSII assemblies for photocatalytic water splitting.[3,8,9] For example, a hybrid natural-artificial photosynthetic platform was constructed by wiring a PSII core and a silicon photoelectrochemical cell, which could perform the water-splitting process under solar irradiation.[8]

Plant PSII dimeric core associates with membrane-spinning light-harvesting antenna complexes (LHCII) to form PSII-LHCII supercomplexes. The dynamic interactions between LHCII and PSII have been the major challenge in the high-resolution structural characterization of PSII-LHCII supercomplexes.[10] This has been drastically improved by the technological advances of cryo-electron microscopy (cryo-EM), demonstrating the power in revolving large biomolecular assemblies.[11–18] However, the resolved 3D structures reflect the static states of multiprotein supercomplexes under specific experimental conditions.[19] They could not delineate dynamic protein associations and the assembly process of photosynthetic PSII-LHCII supercomplexes. Indeed, cryo-EM structures have suggested

[a]Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan 430070, Hubei, China. E-mail: gaojun@mail.hzau.edu.cn

[b]Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK. E-mail: luning.liu@liverpool.ac.uk

[c]Frontiers Science Center for Deep Ocean Multispheres and Earth System, College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d2ra08240c

a higher degree of flexibility of atomic positions at the antenna-PSII core interface and dynamic associations of LHCII with the PSII core,[11] which are fundamental for state transitions and PSII repair.[20–22] Moreover, PSII-LHCII undertakes a stepwise assembly pathway, and the attachment of LHCII to the dimeric PSII core appears to occur at the final step of PSII-LHCII assembly.[23] Some protein subunits, including PsbW, PsbZ, and PsbH, were suggested to be involved in the antenna-PSII core assembly.[11,23–26] The detailed protein–protein interactions and assembly mechanisms that govern the association of LHCII with the PSII core remain elusive.

Molecular dynamics (MD) simulations provide a powerful toolkit to investigate *in silico* the dynamics of biomolecular assemblies with atomic resolutions at a time scale ranging from femtoseconds to milliseconds.[19,27] Large-scale MD simulations have been applied in studying the structures and functions of photosynthetic macromolecular complexes and organelles at different scales.[17–19,28–30] Here, we construct a million-atom scale model of plant C_2S_2-type PSII-LHCII supercomplex embedded in the solvated membrane, based on the cryo-EM structure,[11] and perform microsecond-scale MD simulations to study the protein interactions and assembly process of the sizeable PSII-LHCII supercomplex in a near cellular environment. Our results indicate the general binding principles and atomistic details of the PSII-LHCII assembly.

Methods

Initial model

The cryo-EM structure of spinach PSII-LHCII (PDB ID: 3JCU)[11] was selected as the initial model for simulating the interactions between LHCII proteins and the PSII core. According to the previous studies,[31–33] the PSII-LHCII complex was embedded in a pre-equilibrated lipid bilayer consisting of single-component 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) to mimics native thylakoid membranes. The system was then neutralized by the addition of counter ions (202 sodium ions) and solvated in a double-shell water box containing 277 727 TIP3P water molecules. Our simulation box (240.5 Å × 350.0 Å × 155.5 Å) contains a total of 1 159 431 atoms, and the solvated complexes with lipid bilayers are shown in Fig. S1.†

Histidine residues in the system were singly protonated on Nε, except those coordinating to non-heme FE, HEM or CLA *via* Nε (and were thus protonated on Nδ). In addition, the two disulfide bonds contained in each monomer (C112–C135 of PsbO and C17–C26 of PsbTn) were explicitly considered. The AMBER ff14SB force field parameter set[34] was selected for standard amino acids residues. For the ten cofactors (PHO, BCR, PL9, LHG, SQD, LMG, DGD, LUT, XAT, NEX) (full names are shown in Table S1†), the generalized Amber force field (GAFF) parameter set was adopted.[35] The parameters of CLA and CHL were taken from those developed by Ceccarelli *et al.* for bacteriochlorophyll *a*.[36] The parameters of HEM were taken from the AMBER parameter database.[37] The parameters of POPC were taken from previous work.[38] For the non-heme Fe clusters, we obtained the parameters using the MCPB.py program.[39] The atomic charges of the cofactors were

determined by fitting the electrostatic potential around these molecules by using the RESP model.[40] The atomic charges of the OEX complexes were assigned according to the redox states of its atoms in the dark-adapted (S1) state as follows: Mn1–Mn2, +3; Mn3–Mn4, +4; O1–O5, −2, Ca, +2. According to Ogata *et al.*,[41] the equilibrium bond lengths, bond angles, and torsion angles were set as the average value of OEX in the two monomers in the cryo-EM structure. The force constants for bond lengths, bond angles, and torsion angles were appropriately set at 1500 kcal mol^{-1} Å^{-2}, 500 kcal mol^{-1} Å^{-2}, and 100 kcal mol^{-1} Å^{-2}, respectively, to maintain the configuration of OEX during the simulations.

Molecular dynamic simulations

Energy minimization of the whole system was performed with a three-step procedure: (1) the system was minimized with 20 000 steps with the steepest descent algorithm by freezing the PSII-LHCII complex to smooth the contact between the phospholipid membrane and the protein. (2) 10 000 steps of energy minimization were performed for the whole system with restraints on the protein backbone and heavy atoms of cofactors (100 kcal mol^{-1} Å^{-2}) (3) 10 000 steps of energy minimization was performed without any restraints. The whole system was then slowly heated to 300 K within 60 ps under the restraint (10 kcal mol^{-1} Å^{-2}) to the protein backbone and heavy atoms of cofactors. After these optimization and heating procedures, the constraints were released, and 10 ns equilibrium simulations were performed under the *NPT* ensemble at 300 K to balance the dimensions and density of the system. Finally, a 1 µs production MD simulation was performed using the GPU implementation of PMEMD from the AMBER16 software package.[42] Atomic coordinates of all atoms were recorded every 1 ps. Temperature is controlled here using the Langevin thermostat[43] while pressure is controlled using the anisotropic Berendsen barostat.[44] Covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm.[45] A non-bonding cutoff of 10 Å was applied to van der Waals interactions, and the particle mesh Ewald (PME) method[46] was used to deal with long-range electrostatics.

Binding free energy analysis and identification of hot spots

The structure and function of proteins are extremely sensitive to the surrounding environment. For biomolecular in aqueous environments, a general molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) approach is used, which replaced the explicit solvent with an implicit continuous solvent. However, for membrane proteins, the effect of the membrane environment must be properly considered to ensure accuracy. In this work, the implicit membrane model developed by Luo *et al.*[47] was used to examine the binding free energy between the antennas and the PSII core complex. Unfortunately, the implicit membrane model does not support energy decomposition. So, the classical implicit solvent model approach was also applied.

The MM/PBSA method calculates the binding free energy of the complex (ΔG_{bind}) by eqn (1). In eqn (1), $-T\Delta S$ is the

contribution of entropy to the system's free energy, which mainly involves the energy changes caused by conformational changes such as translation and rotation of the protein. Given that there is no significant change in protein subunits conformation, and the analysis mainly focuses on the electrostatic and hydrophobic contributions of the system rather than the absolute binding free energy, the contribution of the entropy ($-T\Delta S$) was not calculated as that in recent literature.[48–50] ΔG_{gas} is the sum of the electrostatic interaction energy ΔG_{elec} and the van der Waals interaction energy ΔG_{vdW} in vacuum (eqn (2)), representing the contribution of the molecular potential energy. The solvation energy includes the electrostatic solvation energy ΔG_{PB} and the non-polar solvation energy ΔG_{np} (eqn (3)).

$$\Delta G_{bind} = \Delta G_{gas} + \Delta G_{sol} - T\Delta S \tag{1}$$

$$\Delta G_{gas} = \Delta G_{elec} + \Delta G_{vdW} \tag{2}$$

$$\Delta G_{sol} = \Delta G_{PB} + \Delta G_{np} \tag{3}$$

$$G_{np} = \gamma \times SASA + b \tag{4}$$

For the solvated PSII-LHCII complexes with lipid bilayers, the binding free energy of the protein–protein interfaces was calculated using the MM/PBSA approach with the MMPBSA.py module in AmberTools20. For the implicit membrane model, the heterogeneous dielectric membrane model is used here because it describes the membrane environment more accurately than the single dielectric membrane model.[47] The spline fitting was adopted, and the implicit membrane thickness was obtained by calculating the average explicit membrane thickness of the last 800 ns trajectory. This was accomplished by calculating the location of the center of mass of the N and P atoms in the phosphatidylcholine headgroups at the top of the membrane, performing the same calculation on the bottom of the membrane, and taking the difference between them. Periodic boundary conditions were used here so that the value of the electrostatic solvation energy ΔG_{PB} is always 0, as it is part of the vacuum electrostatic interaction energy ΔG_{elec}. Other PB settings are consistent with Greene *et al.*[47] The internal dielectric constant settings for protein in the implicit membrane model and the implicit solvent model will be discussed in the next section. ΔG_{np} is calculated by eqn (4), where the values of constant γ and constant b are set to 0.005 kcal mol^{-1} Å^{-2} and 0.92 kcal mol^{-1}, respectively. Solvent accessible surface area (SASA) was calculated using a water probe radius of 1.4 Å.

The free energy contributions of protein residues could be divided into polar (ΔG_{polar}) and nonpolar ($\Delta G_{nonpolar}$) interactions according to eqn (5) where each part is the sum of two energy terms, as shown in eqn (6) and (7). In the following analysis, ΔG_{polar} is considered as the contribution of electrostatic interactions, while $\Delta G_{nonpolar}$ is considered as the contribution of hydrophobic interactions.[48,49]

$$\Delta G_{bind} = \Delta G_{polar} + \Delta G_{nonpolar} \tag{5}$$

$$\Delta G_{polar} = \Delta G_{elec} + \Delta G_{PB} \tag{6}$$

$$\Delta G_{nonpolar} = \Delta G_{vdW} + \Delta G_{np} \tag{7}$$

According to the previous literature,[51] the hot spots are identified as those whose absolute free energy value is larger than 1.5 kcal mol^{-1}. 8000 snapshots were extracted at intervals of 100 ps along the trajectory. It should be mentioned that since the PSII-LHCII complex is a dimer, the binding free energy values are discussed using the average of the two monomers.

Results

Construction of the structural model of PSII-LHCII for MD simulations

The structures of several PSII-LHCII complexes from different species have been solved.[12–14,52] In this work, the cryo-EM structure from spinach PSII-LHCII (PDB ID: 3JCU) was selected as the initial model for simulating the interactions between antenna proteins and the PSII core.[11] As shown in Fig. 1A, the spinach C_2S_2-type PSII-LHCII supercomplex forms a homodimer with two-fold symmetry. Each monomer comprises a PSII core, a major peripheral light-harvesting complex (LHCII), and two minor chlorophyll-binding proteins of 26 and 29 kDa (CP26, CP29). The PSII core contains four large intrinsic subunits (D1, D2, CP43 and CP47), four extrinsic subunits (PsbO, PsbP, PsbQ, PsbTn) on the luminal surface (Fig. S1†), and twelve small membrane-spanning subunits (PsbE, PsbF, PsbH, PsbI, PsbJ, PsbK, PsbL, PsbM, PsbTc, PsbW, PsbX and PsbZ). Three light-harvesting complexes are bound to the core complex through the LHCII–core, CP26–core, and CP29–core interfaces, which display different protein contacts. Here, we define the three interfaces as $S_{LHCII/core}$, $S_{CP26/core}$, and $S_{CP29/core}$, respectively. In addition, LHCII also forms interfaces with CP26 and CP29, defined as $S_{LHCII/CP26}$ and $S_{LHCII/CP29}$, respectively. This PSII-LHCII structure was embedded in a pre-equilibrated lipid bilayer consisting of single-component 1-palmitoyl-2-oleoyl-*sn-glycero*-3-phosphocholine (POPC) to mimic native thylakoid membranes (Fig. S1†).[31–33] The model construction and the MD simulation process of the PSII-LHCII complex are shown in ESI Video 1.†

The conformations from the 1 μs MD production run were compared to the cryo-EM structure to check the quality of MD simulations. The average structure was obtained over a 1 μs trajectory, and the snapshot with the smallest RMSD value from the average structure was adopted as the representative structure. We found that the representative structure has only slight differences with the cryo-EM structure (Fig. S2†). First, Ramachandran plot analysis from PROCHECK[53] revealed that the MD structure has a wider conformational distribution space and most of the backbone in the allowed regions of Ramachandran plot, compared with the cryo-EM structure (Fig. 1B and C). The number of residues in disallowed regions was reduced from 0.6% to 0.2%, and the number in generously allowed regions was reduced from 0.6% to 0.3% (Table 1). The number of residues in the most favored regions was similar (88.1% for MD structure and 88.3% for cryo-EM structure). These results confirmed the good quality of the cryo-EM structure and indicated that MD simulations only slightly improved
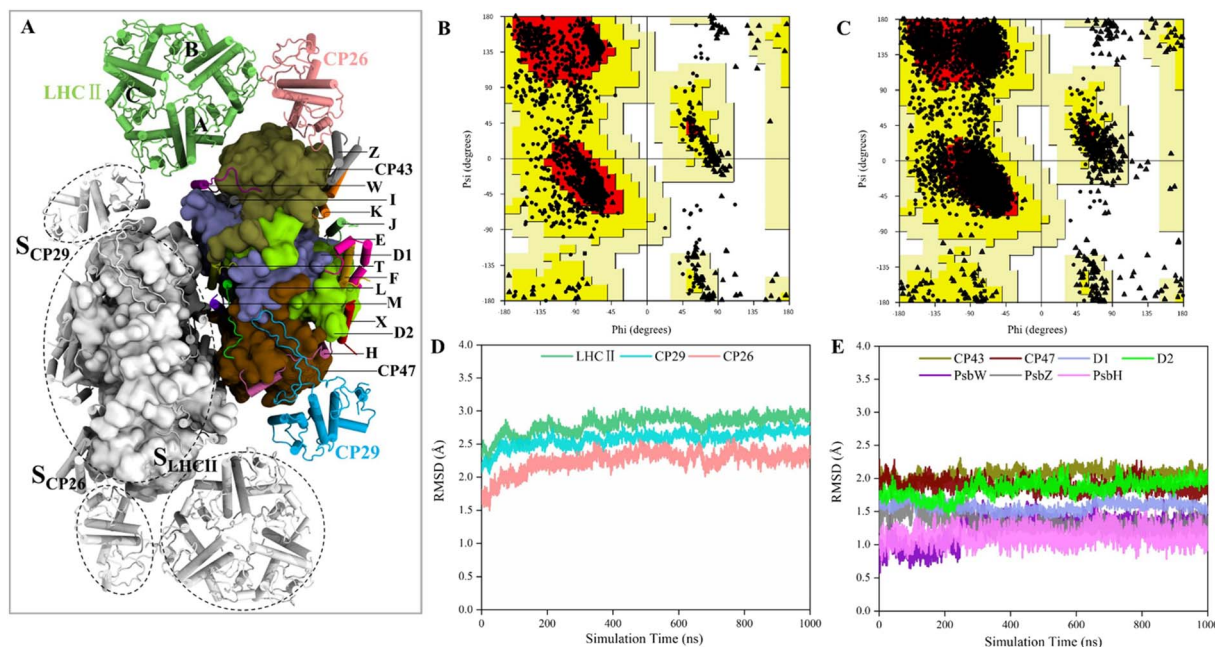
Fig. 1 Validation of the quality of MD simulation. (A) Stromal view on the PSII–LHCII dimer. For clarity, only the subunits of the right monomer are highlighted, and different subunits are colored and labeled differently (the four extrinsic subunits are not shown). Among them, the 4 large core subunits are shown using the surface model and the remaining parts are shown using the cartoon model. The positions of three antennae–core interfaces are indicated in the left monomer. (B and C) Ramachandran plots for cryo-EM and MD structures, respectively. (D and E) RMSD values of proteins associated with the antennae–core interfaces (for CP26 and PsbW, only the RMSDs of non-loop regions were considered due to the high volatility of the loop regions). The RMSDs are obtained using the average of the two monomers and are colored as in (A).

the stereochemistry of the cryo-EM structure. Then, we examined the quality of nonbonded interactions using ERRAT program.[54] The structures were evaluated based on the quality factor, which depends on the statistics of nonbonded atomic interactions in the 3-D protein structure. The results showed that the overall quality factor of the MD structure (90.2%) was greater than that of the cryo-EM structure (79.8%) (Table 1), indicating that MD simulations did improve the quality of nonbonded interactions from the PSII-LHCII cryo-EM structure. Since the antenna–core binding mainly depends on nonbonded interactions between the protein interfaces, the improvement in the accuracy of nonbonded interactions here is very important for the subsequent accurate calculation of protein interface affinities.

Conformational fluctuation analysis

To verify whether the antenna–core interfaces reach equilibrium, we calculated the RMSD values of three antennas (LHCII, CP29, CP26) and seven PSII core subunits (4 large intrinsic subunits D1, D2, Cp43, Cp47, and 3 small intrinsic subunits PsbW, PsbZ, PsbH). As shown in Fig. 1D and E, after 200 ns of simulation, the RMSD values of all the proteins at the antenna–core interfaces remained consistent, indicating that the $S_{LHCII/core}$, $S_{CP26/core}$, and $S_{CP29/core}$ interfaces reached stable states. Likewise, RMSD analysis showed that the pigments also reached equilibrium (Fig. S3†). Accordingly, the MD simulation trajectories after 200 ns were used for subsequent statistical analysis.

The RMSD values of the peripheral antennas LHCII, CP29, and CP26 (ranging from 1.50 Å to 3.07 Å, Fig. 1D) were generally more significant than those of the core proteins (ranging from 0.58 Å to 2.31 Å, Fig. 1E), suggesting that the light-harvesting antennas possess larger conformational fluctuations than the PSII core subunits, consistent with the experimentally observed high mobility of antenna proteins.[10] To further characterize the fluctuation features of protein residues, we investigated root

Table 1 Validation of cryo-EM and MD structures using PROCHECK and ERRAT program

| Structure | Ramachandran plot statistics (%) | | | | ERRAT (%) |
	Most favored	Additionally allowed	Generously allowed	Disallowed	Overall quality factor
Cryo-EM	88.3	10.5	0.6	0.6	79.8
MD	88.1	11.4	0.3	0.2	90.2

mean square fluctuation (RMSF) of the protein backbone C_α atoms around their average positions. The volatility of extrinsic subunits (average RMSF value of 1.91 Å, Fig. 2A) and antenna proteins (average RMSF value of 1.71 Å, Fig. 2B) was markedly higher than that of the core complex (average RMSF value of 1.16 Å, Fig. 2B), consistent with RMSD results (Fig. 1). In addition, the fluctuation of pigments was strongly correlated with the fluctuation of binding proteins (Fig. S4†). For individual antennas, CP26 has a notably higher RMSF value (average value: 2.22 Å) than LHCII monomer adjacent to the core (LHCII(A), average value: 1.36 Å) and CP29 (average value: 1.45 Å). The

greater fluctuation was mainly ascribed to the loop regions of CP26, which is away from the $S_{CP26/core}$ interface; likewise, the fluctuating regions of CP29 and LHCII(A) are also not at the antenna–core interfaces (Fig. 2C–E). Consequently, although the peripheral antennas showed overall conformational fluctuations, $S_{LHCII/core}$, $S_{CP26/core}$, and $S_{CP29/core}$ exhibited relatively stable conformations (Fig. 2B), presumably favoring the association of antennas with the PSII core and energy transfer. The analysis also verified the reliability of MD simulations using the constructed structural model.

Fig. 2 RMSF fluctuations of the PSII-LHCII backbone. (A) Side view and (B) stromal view on the PSII-LHCII dimer. All protein residues were colored according to the RMSF values of C_α atoms capped at 2.5 Å: red represents a high RMSF value and blue represents a small RMSF value. (C–E) RMSF values of LHCII(A), CP29 and CP26 residues, respectively. Grey boxes and arrows indicate the regions with higher fluctuation values.

Analysis of the interface binding free energies

To study the binding characteristics at different antenna–core interfaces, the binding free energies of the three antenna–core interfaces ($S_{LHCII/core}$, $S_{CP26/core}$, and $S_{CP29/core}$) were calculated using the MM/PBSA approach.[48,51,55] As the choice of internal dielectric constant values has a significant impact on the results,[56] we examined a set of values using $S_{LHCII/core}$ as an example (Table S2†) to determine an appropriate internal dielectric constant. For the implicit membrane model,[47,57] a value of 20 is appropriate when the ligand is charged, according to recommended settings in the literature.[57] Therefore, we tested values around 20, including 10, 15, 20, 25, and 30. The change in binding free energy was examined with internal dielectric constant values of 1–8, often used for the implicit solvent model.[58,59] The results showed that both models' free energy decreased with the internal dielectric constant increased, and only the electrostatic interaction energy changed (Table S2†). In addition, the electrostatic interaction energy was always positive, whereas the contribution of hydrophobic interactions remained negative. These results suggest that the choice of internal dielectric constants values does not have remarkable effects on qualitative determination of binding affinity. Therefore, the common setting, namely the internal dielectric constant value of 20 for the implicit membrane model and 1 for the implicit solvent model, was used to investigate the binding free energy of protein interfaces.

Based on the heterogeneous implicit membrane model, free energies of the three antenna–core protein interfaces were obtained from equilibrium trajectories (*i.e.*, 200–1000 ns) (Table 2). To verify the convergence of the free energy calculations, we performed convergence tests for $S_{LHCII/core}$ (-58.0 ± 7.2 kcal mol^{-1}, -57.0 ± 7.9 kcal mol^{-1}, -59.4 ± 7.2 kcal mol^{-1}, -56.4 ± 13.6 kcal mol^{-1}), $S_{CP26/core}$ (2.3 ± 5.5 kcal mol^{-1}, 1.5 ± 5.6 kcal mol^{-1}, 2.0 ± 5.6 kcal mol^{-1}, 3.2 ± 6.4 kcal mol^{-1}) and $S_{CP29/core}$ (-111.6 ± 9.6 kcal mol^{-1}, -112.7 ± 9.7 kcal mol^{-1}, -116.1 ± 9.4 kcal mol^{-1}, -116.1 ± 9.8 kcal mol^{-1}) using trajectories of 800–1000 ns, 600–1000 ns, 400–1000 ns and 200–1000 ns, respectively. In addition, non-equilibrium trajectories (0–50 ns) were also used to calculate the binding free energies of $S_{LHCII/core}$ (-2.7 ± 14.8 kcal mol^{-1}), $S_{CP26/core}$ (12.2 ± 5.1 kcal mol^{-1}) and $S_{CP29/core}$ (-84.5 ± 8.7 kcal mol^{-1}), the results are clearly different from the above calculations using equilibrium trajectories, indicating that a long simulation scale is necessary for accurate free energy calculations of the antenna–core interface.

$S_{CP29/core}$ (-116.1 kcal mol^{-1}) has the largest total free energy among the three interfaces, which is double of $S_{LHCII/core}$ (-56.4 kcal mol^{-1}), indicating the strong binding affinity at $S_{CP29/core}$. The electrostatic interaction energies (ΔG_{elec}) of $S_{LHCII/core}$, $S_{CP26/core}$, and $S_{CP29/core}$ were 58.5 kcal mol^{-1}, 55.9 kcal mol^{-1} and 69.8 kcal mol^{-1}. In contrast, their hydrophobic interaction energies (ΔG_{np}) were -114.9 kcal mol^{-1}, -52.7 kcal mol^{-1}, and -185.9 kcal mol^{-1}, respectively. As described in Methods (eqn (6) and (7)), the binding free energies can be divided into ΔG_{polar} and $\Delta G_{nonpolar}$ (hydrophobic interaction). Our analysis suggested that $\Delta G_{nonpolar}$ provides the dominant contribution to the binding of the antenna to the PSII core. Meanwhile, the ΔG_{polar} values of all interfaces are positive, indicating that polar or electrostatic interactions between the antenna and PSII provide repulsive forces unfavorable for binding. This conclusion was also supported by the calculation using the implicit water model (Table S3†). In contrast, the binding free energies of the two antenna–antenna interfaces, $S_{LHCII/CP26}$ and $S_{LHCII/CP29}$, were both positive values (Table S4†), indicative of the very weak antenna–antenna association (Table 2).

Free energy decomposition and hot spot analysis

Based on the identification of hydrogen bonds (HBs) and salt bridges (SBs) at the antenna–core interface in the cryo-EM structure,[11] electrostatic interactions were previously assumed to be the main driving force mediating the binding of antennae to the PSII core.[60] However, our free energy calculation indicated that the total electrostatic interactions act as repulsive forces. To address this contradiction, we examined the contributions of individual residues to the total free energies and HB/SB interactions formed by the interface residues. The free energy contributions were decomposed into polar and non-polar fractions to evaluate the contributions of electrostatic and hydrophobic interactions. Residues with an absolute value of binding free energy greater than 1.5 kcal mol^{-1} were identified as hot spots (Tables S5 and S6†), according to previous work.[51]

For $S_{LHCII/core}$, 23 amino acids were identified as hot spots (Fig. 3E and Table S5†). Among them, 20 hot spots had negative free energy values, and their total free energy contribution was -42.22 kcal mol^{-1}, accounting for 65% of the total binding free energies (-64.5 kcal mol^{-1}). Three residues in LHCII, E175, K179, and D215, have positive free energy contributions. All $\Delta G_{nonpolar}$ of residues are negative, but ΔG_{polar} of 13 residues are positive. For $S_{CP26/core}$, only 5 amino acids were identified as hot

Table 2 Binding free energy of different antenna–core interfaces (kcal mol^{-1}). ΔG_{vdW}, the van der Waals interaction energy; ΔG_{np}, the non-polar solvation energy; ΔG_{elec}, the electrostatic interaction energy; $\Delta G_{nonpolar}$, the contribution of hydrophobic interactions; ΔG_{polar}, the contribution of electrostatic interactions; ΔG_{bind}, the binding free energy

Interface	ΔG_{vdW}	ΔG_{np}	ΔG_{elec}[a]	$\Delta G_{nonpolar}$[b]	ΔG_{polar}[c]	ΔG_{bind}[d]
$S_{LHCII/core}$	-103.5 ± 12.5	-11.4 ± 1.0	58.5 ± 2.5	-114.9	58.5	-56.4 ± 13.6
$S_{CP26/core}$	-46.5 ± 5.3	-6.2 ± 0.5	55.9 ± 3.0	-52.7	55.9	3.2 ± 6.4
$S_{CP29/core}$	-167.4 ± 8.5	-18.5 ± 0.7	69.8 ± 3.8	-185.9	69.8	-116.1 ± 9.8

[a] Includes the electrostatic solvation energy, see methods. [b] Formula (6). [c] Formula (7). [d] Formula (5).
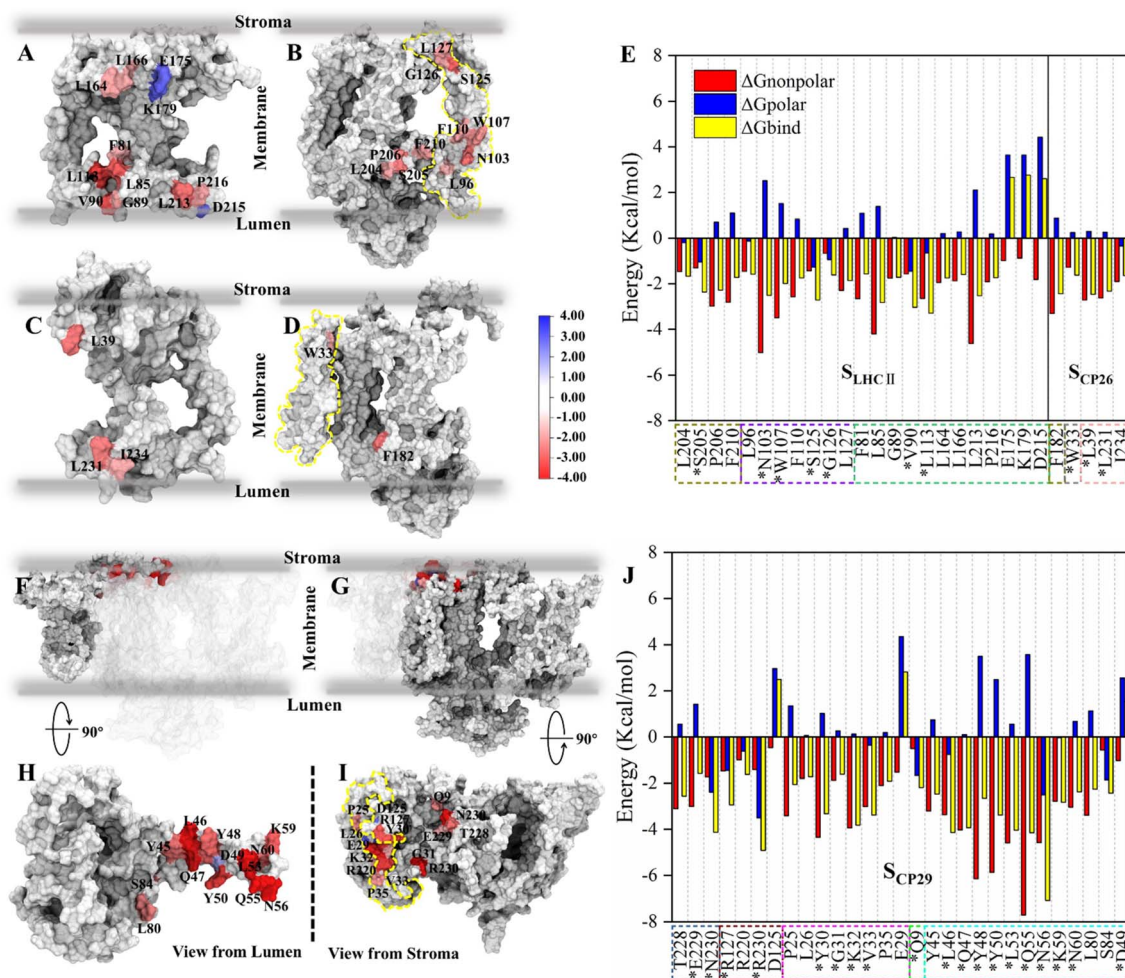
Fig. 3 Binding free energy and corresponding component contribution of hot spots on $S_{LHCII/core}$, $S_{CP26/core}$ and $S_{CP29/core}$ and their localization. (A and B) Localization of hot spots on LHCII and the core complex (at the $S_{LHCII/core}$), respectively. (C and D) Localization of hot spots on CP26 and the core complex (at the $S_{CP26/core}$), respectively. The protein surface map colored according to the free energy contribution of each hot spot. The yellow dotted lines in panels (B) and (D) correspond to the positions of the small intrinsic subunits PsbH and PsbZ, respectively. The unit of energy is kcal mol^{-1}. (E) Binding free energy and corresponding component contribution of hot spots, including hydrophobic (red), electrostatic (blue) and total (yellow) contribution. Color scheme is presented as the same as shown in Fig. 1A. The dotted box lines with different colors represent different subunits, which are CP43 (tan), PsbW (purple), LHCII (lime), PsbZ (silver) and CP26 (pink). Hot spots that participate in the formation of hydrogen bonds or salt bridges are marked with *. See Table S5† for more details. (F and H) Localization of CP29 and corresponding hot spots. (G and I) Are the localization of core complex (at the $S_{CP29/core}$) and corresponding hot spots. The yellow dotted lines in panel (i) correspond to the positions of the small intrinsic subunits PsbH. (J) Binding free energy and corresponding component contribution of hot spots. The dotted box lines with different colors represent different subunits, which are CP47 (ochre), D1 (ice blue), PsbH (mauve), PsbL (green), and CP29 (cyan), and the colors correspond to Fig. 1A. See Table S6† for more details.

spots (Fig. 3E and Table S5†), consistent with the relatively low interface affinity (Table 2). Interestingly, ΔG_{polar} of most of the residues are positive (Fig. 3E and Table S5†). A large number of hot spots at $S_{CP29/core}$ (Fig. 3J and Table S6†), compared with those at $S_{LHCII/core}$ and $S_{CP26/core}$, is consistent with its higher interface affinity (Table 2). 26 of the 29 hot spots contribute negative free energies. Their total free energy contribution was −79.54 kcal mol^{-1}, accounting for 68% of the total binding free energy. In contrast, 19 hot spots have positive ΔG_{polar} contribution. The free energy decomposition indicated that the ΔG_{polar} contribution (*i.e.* electrostatic interactions) is primarily positive, representing unfavorable interactions, consistent with our free energy analysis (Table 2). It should be noted that the

residues with hydrogen bonds may contribute negative ΔG_{polar}; we will discuss it in the next section.

We also determined the spatial distribution of hot spots. At $S_{LHCII/core}$, 12 hot spots are located in LHCII, including 4 hot spots (L164, L166, E175, K179) at the stromal side and 8 hot spots (F81, L113, L85, V90, G89, L213, P216, D215) at the luminal side. Eleven hot spots are located in the core (Fig. 3A and B), including 3 hot spots (L127, G126, and S125) at the stromal side and 8 near the luminal side. At $S_{CP26/core}$, 3 hot spots are located at CP26, and 1 hot spot is in the core complex, all of which are near the luminal side or the stromal side (Fig. 3C and D). At $S_{CP29/core}$, all hot spots are concentrated at the stromal side: 16 hot spots were localized in the core

complex, and 13 were localized at CP29 (Fig. 3F–I). Moreover, the hot spots in the core complex are widely distributed in multiple protein chains: 7 of the 16 hot spots are located at PsbH, 4 hot spots are at CP47, 3 hot spots are at D1, and the remaining one is localized at PsbL (Fig. 3I). At CP29, all the 13 hot spots are localized at Motif II (Pro42–Phe87) of the N-terminal long loop chain, implying its role of CP29–core complex binding. Collectively, the hot spots are mainly distributed at the luminal or stromal side, indicating that the main driving force in these interfaces' binding originates from the luminal or stromal side.

Hydrogen bonds/salt bridges across the antenna–core interfaces

To further determine the effects of electrostatic interactions on the interfacial affinity, HBs/SBs at the antenna–core interfaces were analyzed. The distance for determining the formation of HBs was set to 3.5 Å, and the corresponding angle was 135°.[32,61] SB analysis was performed using a homemade python script, and a cutoff distance of 6 Å was used to detect SBs between the basic nitrogen and acidic oxygens.[62] HBs/SBs with an occupancy greater than 30% during MD simulations were considered stable. In addition to the HBs/SBs characterized in the cryo-EM structure (Table S7†), 25 additional HB/SB interactions were identified between the antenna–core interfaces during the course of 1 μs MD simulations at room temperature (Tables 3 and 4). Intriguingly, many of these HBs/SBs have been reported in other PSII-LHCII structures at higher resolutions,[13,52]

confirming the reliability of the structural model achieved from MD simulations.

The total numbers and spatial locations of HBs/SBs found at $S_{LHCII/core}$, $S_{CP26/core}$, and $S_{CP29/core}$ are consistent with those of the hot spots identified. The numbers of HBs/SBs are 25/6 for $S_{CP29/core}$, 9/0 for $S_{LHCII/core}$, and 4/2 for $S_{CP26/core}$ (Table 3). Eight of the 9 HBs at $S_{LHCII/core}$ are mediated by PsbW, and 7 are distributed at the luminal side (Fig. 4A and Table 3). Three of the 4 HBs and all SBs at $S_{CP26/core}$ are mediated by PsbZ, and 3 HBs/SBs are located at the stromal side (Fig. 4B and Table 3). All the 25 HBs and 6 SBs at $S_{CP29/core}$ are distributed at the stromal side and are mediated by Motif II of CP29. Residues of Motif II form HBs/SBs with multiple protein subunits of the PSII core (PsbH: 5/1; PsbL: 6/0; CP47: 10/5; D1: 4/0) (Fig. 4C and Table 4).

The differences between average binding free energies of the HB/SB-involved hot spots and all hot spots were −0.38 kcal mol^{-1} for $S_{LHCII/core}$, −0.04 kcal mol^{-1} for $S_{CP26/core}$, and −0.66 kcal mol^{-1} for $S_{CP29/core}$ (Fig. S5†), which accounted for 26%, 1.6%, and 27% of the average binding free energies of all hot spots, respectively. This suggests that the formation of HBs/SBs predominately modulates hot spots' binding free energy contribution and enhances their binding ability. Moreover, the change in binding free energies was dominated by the reduction in electrostatic interaction energies (Fig. S5†). At $S_{LHCII/core}$, 8 of the 23 hot spots were involved in HB formation (Fig. 3A–E, residues marked with *), and 5 hot spots, including V90 and L113 (LHCII), S125 and G126 (PsbW), and S205 (CP43), provide negative electrostatic interaction energies. At $S_{CP26/core}$, 3 of the 5 hot spots (W33, L39, L231) were involved in HB

Table 3 Hydrogen bonds and salt bridges at $S_{LHCII/core}$ and $S_{CP26/core}$

	Hydrogen bond					
	Donor	Acceptor	Distance (Å)	Occupancy (%)	PSII subunits	Cryo-EM structures[a]
$S_{LHCII/core}$	VAL_90@N	ASN_103@Oδ	2.96 ± 0.16	64	PsbW	
	ASN_88@Nδ	PRO_97@O	2.88 ± 0.14	38	PsbW	
	ASN_103@Nδ	LEU_113@O	2.92 ± 0.15	95	PsbW	
	SER_205@N	ALA_214@O	2.92 ± 0.15	66	CP43	
	ASN_103@Nδ	VAL_90@O	2.94 ± 0.15	88	PsbW	
	GLY_126@N	GLU_175@Oε	3.03 ± 0.19	88	PsbW	
	TRP_107@Nε	TYR_112@Oη	3.14 ± 0.17	32	PsbW	
	SER_101@N	ASN_88@O	2.99 ± 0.16	37	PsbW	
	SER_125@Oγ	GLU_175@Oε	2.64 ± 0.11	34	PsbW	
$S_{CP26/core}$	ARG_32@Nη	SER_143@O	2.90 ± 0.17	37	CP43	
	TRP_33@Nε	GLY_38@O	2.92 ± 0.16	61	PsbZ	5XNL, 7OUI
	SER_59@Oγ	LEU_231@O	2.78 ± 0.17	31	PsbZ	3JCU, 5XNL
	LYS_37@Nζ	LEU_39@O	2.88 ± 0.15	40	PsbZ	3JCU, 5XNL
	Salt bridge					
	Acidic	Basic	Distance (Å)	Occupancy (%)	PSII subunits	Cryo-EM structures[a]
$S_{CP26/core}$	ASP_41	LYS_37	3.99 ± 0.92	47	PsbZ	5XNL
	ASP_44	LYS_37	5.05 ± 1.00	42	PsbZ	5XNL

[a] Interfacial HBs/SBs present in the cryo-EM structure of spinach (PDB ID: 3JCU, 3.2 Å), pea (PDB ID: 5XNL, 2.7 Å) and arabidopsis (PDB ID: 7OUI, 2.79 Å).

Table 4 Hydrogen bonds and salt bridges at $S_{CP29/core}$

Hydrogen bond

Donor	Acceptor	Distance (Å)	Occupancy (%)	Proteins in PSII	Cryo-EM structures[a]
TYR_48@N	GLY_31@O	2.94 ± 0.14	99	PsbH	3JCU, 5XNL
ASN_56@Nδ	GLU_8@O	2.92 ± 0.15	96	PsbL	
GLN_47@Nε	TYR_30@O	2.92 ± 0.16	81	PsbH	3JCU
GLN_55@N	ASN_7@Oδ	2.99 ± 0.18	81	PsbL	
LEU_46@N	VAL_33@O	2.94 ± 0.14	87	PsbH	5XNL, 7OUI
TYR_50@Oη	SER_132@Oγ	2.85 ± 0.16	59	CP47	
ASN_56@Nδ	ASN_5@O	2.84 ± 0.12	56	PsbL	
ASN_56@N	PRO_6@O	2.98 ± 0.16	48	PsbL	
GLN_55@Nε	THR_10@O	3.03 ± 0.19	34	CP47	3JCU, 5XNL, 7OUI
ASN_60@Nδ	ARG_476@O	2.90 ± 0.14	50	CP47	5XNL
GLN_55@Nε	ASN_230@Oδ	2.97 ± 0.17	45	D1	5XNL, 7OUI
ASN_56@Nδ	PRO_6@O	2.92 ± 0.16	33	PsbL	
SER_84@Oγ	GLU_29@Oε	2.68 ± 0.13	32	PsbH	
VAL_33@N	LEU_46@O	2.91 ± 0.14	99	PsbH	7OUI
ASN_230@Nδ	ASN_56@Oδ	2.90 ± 0.15	95	D1	5XNL, 7OUI
GLU_229@N	LEU_57@O	3.02 ± 0.17	83	D1	5XNL
GLN_9@Nε	ASN_56@Oδ	2.90 ± 0.14	88	PsbL	5XNL
ASN_230@N	GLN_55@O	3.03 ± 0.16	59	D1	3JCU, 7OUI
GLN_223@Nε	TYR_48@Oη	3.07 ± 0.18	48	CP47	
ASN_14@Nδ	LEU_53@O	2.97 ± 0.17	40	CP47	7OUI
ARG_230@Nη	ASN_60@Oδ	2.94 ± 0.16	77	CP47	
ARG_230@Nη	ASP_51@Oδ	2.80 ± 0.11	48	CP47	
LYS_137@Nζ	GLN_81@Oε	2.84 ± 0.13	86	CP47	
ARG_476@Nη	LEU_53@O	2.98 ± 0.18	30	CP47	3JCU
ARG_230@Nε	ASP_51@Oδ	3.08 ± 0.18	30	CP47	

Salt bridge

Acidic	Basic	Distance (Å)	Occupancy (%)	Proteins in PSII	Cryo-EM structures[a]
GLU_85	LYS_32	3.80 ± 0.45	95	PsbH	3JCU, 7OUI
ASP_49	LYS_130	5.34 ± 0.39	55	CP47	3JCU
GLU_71	ARG_127	4.32 ± 0.93	43	CP47	
ASP_51	ARG_230	4.47 ± 0.29	71	CP47	
ASP_477	LYS_59	5.35 ± 0.33	39	CP47	3JCU, 7OUI
ASP_483	LYS_59	4.07 ± 1.23	79	CP47	3JCU, 5XNL

[a] Interfacial HBs/SBs also present in the cryo-EM structure of spinach (PDB ID: 3JCU, 3.2 Å), pea (PDB ID: 5XNL, 2.7 Å) and arabidopsis (PDB ID: 7OUI, 2.79 Å).

formation and W33 contributes negative electrostatic interaction energies (Fig. 3C–E). At $S_{CP29/core}$, 19 of the 29 hot spots were involved in HB/SB formation (Fig. 3J, residues marked with *), and 8 hot spots, including N230 (D1), R127 and R230 (CP47), V33 (PsbH), Q9 (PsbL), as well as L46, N56, and K59 (CP29), contribute negative electrostatic interaction energy. Overall, our results revealed that HBs/SBs are favorable for the binding of the interfaces, which is consistent to a certain extent with the previous study[60] that indicated that electrostatic interactions were the main driving force to mediate the binding of antennae to the PSII core. However, our analysis showed that hydrogen bonds contribute only about 20% of the binding energies (Fig. S5†). In contrast, most of the binding energies come from $\Delta G_{nonpolar}$ (hydrophobic interaction). It is likely that hydrogen bonds mainly provide directional interactions and the anchoring of the interface.

Roles of PSII small intrinsic subunits in antenna–core association

To study the roles of small intrinsic subunits in antenna–core binding, the binding free energies at the antenna–core interface in the absence of PsbW, PsbZ, and PsbH were calculated (Fig. 5A). The binding affinities at the antenna–core interfaces were significantly weakened in the absence of PsbW, PsbZ, and PsbH (Fig. 5B). The hydrophobic interaction energies of the ΔPsbW, ΔPsbZ, and ΔPsbH systems were decreased by 66.5%, 52.3%, and 32.9%, respectively, and the changes in electrostatic interaction energies were negligible. These results indicate that the small intrinsic subunits of PSII (PsbW, PsbZ, and PsbH) contribute about 30% to 60% of the binding free energies for antenna–core binding, corroborating the experimental findings that peripheral antenna content was reduced in the absence of PsbW and PsbZ.[24,25]
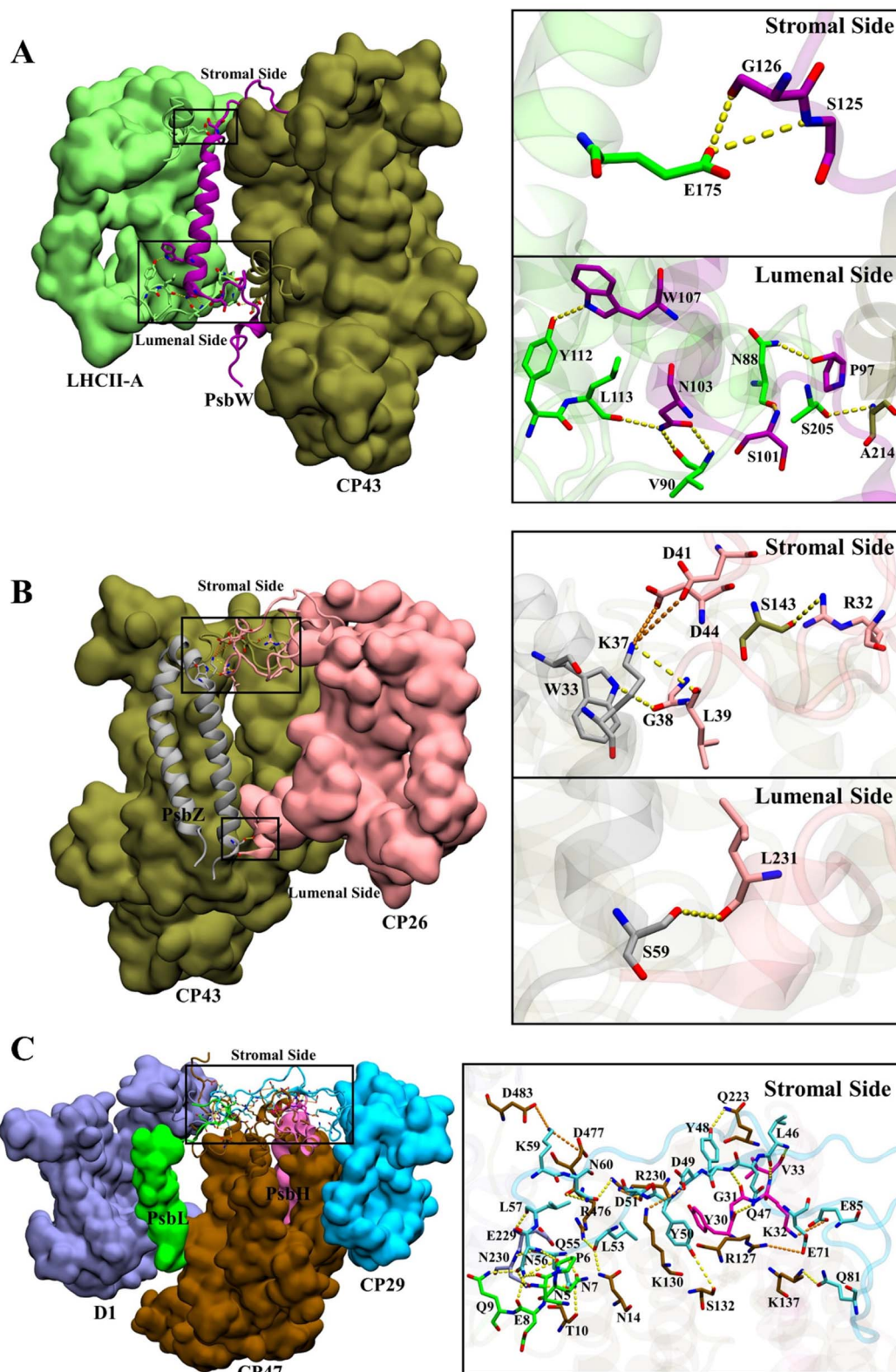
Fig. 4 Hydrogen bond and salt bridge interactions at $S_{LHCII/core}$ (a), $S_{CP26/core}$ (b), and $S_{CP29/core}$ (c). Hydrogen bonds and salt bridges are represented by yellow and orange dashed lines, respectively. Colour scheme is presented as the same as shown in Fig. 1A.

Most of the hot spots and residues involved in HB/SB formation of CP29 were identified on Motif II of CP29, which represents a long loop region (Pro42–Phe87) and is responsible for binding to the core proteins.[11] In the absence of Motif II, the binding free energy increased from -116.1 ± 9.8 kcal mol^{-1} to 56.6 ± 4.1 (Fig. 5B), implicating that CP29 can not bind to core
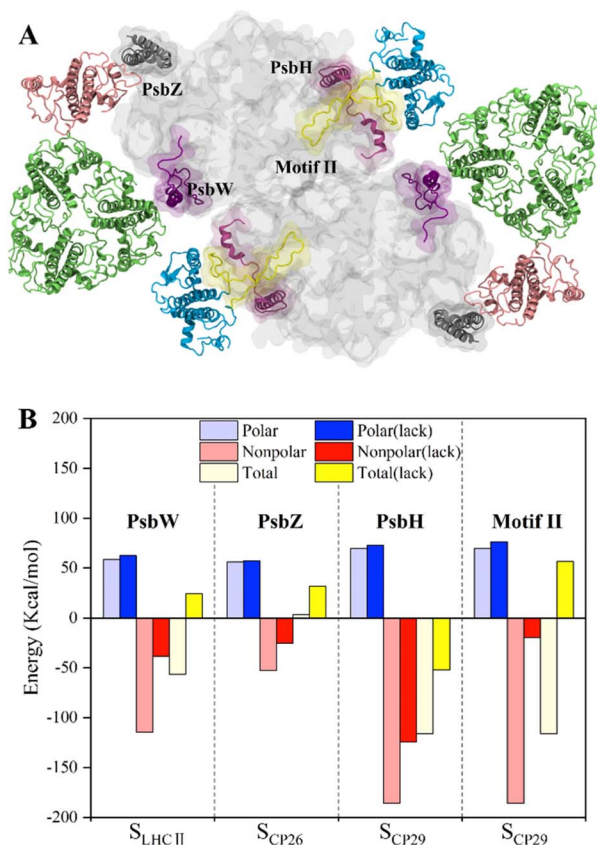
Fig. 5 Binding free energy changes caused by the deletion of key subunits and a schematic diagram of the inferred assembly process. (A) Display of the small intrinsic subunits PsbW, PsbZ, PsbH, and Motif II of the N-terminal loop region of CP29, the colors of protein subunits correspond to Fig. 1a, and the Motif II is shown in yellow. (B) Affinity changes of the antenna–core interface in the absence of PsbW, PsbZ, PsbH and Motif II.

proteins without Motif II. Moreover, CP29 interacts with multiple subunits including PsbH, CP47, D1, and PsbL. This differs from LHCII and CP26, which interact with core proteins *via* small intrinsic subunits, *i.e.*, PsbW and PsbZ. The absence of PsbH only led to a relatively small percentage (32.9%) of binding free energy lose, suggesting that the function of PsbH in binding with CP29 and the core may not be as significant as expected.

Discussion

High-resolution structures of PSII-LHCII supercomplexes have advanced our knowledge about the self-organization of photosynthetic antenna-reaction center systems. However, the relatively low resolution at highly dynamic regions in the systems, particularly the antenna–core interface, has raised many questions in understanding the PSII-LHCII assembly. Standing on the cryo-EM structure, we developed a reliable million-atom-scale model of plant C_2S_2-type PSII-LHCII supercomplex embedded in a solvated membrane using *in silico* simulations. Microsecond-scale MD simulations on the 1.1 megadalton

multi-protein ensemble allow us to comprehensively analyze the inter-protein interactions and binding free energies at the antenna–core and antenna–antenna interfaces, providing very suggestive details of the stepwise assembly of the PSII-LHCII supercomplex.

In our study, the binding free energy calculations with corresponding component decompositions revealed that the antenna–core binding is a competitive process between electrostatic energy and hydrophobic energy, in which the electrostatic interactions provide mainly repulsive energy. In contrast, the hydrophobic contributions serve as the main driving force to facilitate antenna–core association (Fig. 3). This may imply a common mechanism underlying the protein–protein interactions and assembly of membrane macromolecular complexes, in which hydrophobic interactions play an essential role. In addition, compared to the cryo-EM structure, MD simulations reveal the HBs/SBs interactions at the antenna–core interface more comprehensively (Fig. 4). Electrostatic interaction energy analysis further indicated that the formation of HBs/SBs enhances the binding affinity of hot spots and is thought to play a positioning role during antenna–core assembly.

Possible assembly pathways of PSII-LHCII, PSII-CP26 and PSII-CP29

Our MD simulations data supported a stepwise assembly of the PSII–LHCII complex (Fig. 6). The small intrinsic proteins PsbW and PsbZ bind to the periphery of the PSII core.[63] They may play critical roles in mediating the binding of the antenna to the PSII core. Their absence, especially PsbW, could significantly weaken the hydrophobic interactions at the antenna–core interface (Fig. 5). The LHCII–core binding may undertake two steps: LHCII first binds to PsbW, due to its strong interaction with PsbW, to form the LHCII–PsbW assembly intermediate, and then binds to the PSII core as one assembly unit (Fig. 6). Previous experimental results supported that PsbW and LHCII co-occur with the PSII core.[23] The CP26–core binding may also undertake two steps and form the CP26–PsbZ assembly intermediate before binding to the core. A previous experimental study has consistently shown that PsbZ could be released together with CP26 from the PSII complex.[23] For CP29, $S_{CP29/core}$ interaction analysis showed that binding to the PSII core was not strongly dependent on specific small intrinsic subunits (such as PsbH). Instead, CP29 forms extensive interactions with multiple core subunits of PSII through its Motif II to facilitate stable binding to the core. Therefore, it is more likely that the anchoring of CP29 to the PSII core represents a one-step process without the assistance of other factors and any assembly intermediate formation (Fig. 6).

Weak interactions of the antenna–antenna interfaces support an independent assembly

Our data also indicated weaker binding affinity at the antenna–antenna interfaces than at the antenna–core interfaces, suggesting independent association/disassociation of individual antenna complexes LHCII, CP29, and CP26. Interestingly, under

Fig. 6 Schematic diagram of the antenna–core assembly process of the C_2S_2-type PSII–LHCII. LHCII is represented in green, CP26 in pink, CP29 in blue, PsbW in purple, and PsbZ in gray, these colors correspond to Fig. 1A. Motif II of CP29 is shown with a yellow coil. Brown and maroon show dimerized PSII cores. Arrows with numbers represent potential assembly steps.

low-light conditions, two other antennas, M-LHCII and L-LHCII, could attach to the C_2S_2-type PSII-LHCII super-complex.[12] They do not interact directly with the PSII core but are bridged by CP29, ultimately forming the $C_2S_2M_2L_2$-type PSII-LHCII complex.[12] These periphery antennae belong to the Lhcb protein family with similar sequence and structural features, suggesting that they may have similar interaction characteristics. It is known that LHCII can migrate between PSI and PSII to balance the distribution of excitation energy between PSI and PSII in state transitions.[20,21] We speculate the interfacial affinities between M-LHCII/L-LHCII and CP29 are relatively weak, which might be essential for driving the dynamic association/dissociation of M-LHCII/L-LHCII to the PSII core in state transitions.[64,65] Given the relatively strong interaction with the core, S-LHCII might be difficult for detach from the PSII core. This is consistent with previous experimental results, which showed that only L-LHCII freely diffuse between PSI and PSII in state transitions[66] and that M-LHCII may detach from the PSII-LHCII supercomplex under specific physiological conditions.[65] The detailed mechanisms underlying the dynamic PSII-LHCII assembly in state transitions remain to be elucidated. It has been presumed that phosphorylation of LHCII N-terminus sites could alter the binding affinity of LHCII to PSII, resulting in its disassociation from PSII.[20,66,67] As electrostatic interactions provide repulsion between protein interfaces and protein residues at the stromal side of PSII-LHCII exhibits a negative electrostatic potential (Fig. S6†), phosphorylation of LHCII N-terminus likely leads to an increase in the negative electrostatic potential and the repulsive force at the protein–protein interfaces, which drive the disassociation of LHCII from PSII.

Conclusions

Our structural model and simulations approach provide the framework for characterizing the molecular mechanisms of photosynthetic complexes' structural and functional dynamics. A better understanding of the antenna–photosynthetic reaction center assembly and interactions will inform strategies to enhance photosynthetic energy transfer and design efficient artificial photosynthetic systems.

Author contributions

JG and LNL: idea conceptualization. JG and LNL: project administration, and funding acquisition. RM, LB, HZ, and JG: methodology and validation. RM, LB, JG and LNL: article writing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 J. Barber, *Chem. Soc. Rev.*, 2009, **38**, 185–196.
2 L. Wu, L. Zhang, W. Tu, R. Sun and C. Yang, *J. Plant Physiol.*, 2020, **251**, 153189.
3 M. Xuan and J. Li, *Natl. Sci. Rev.*, 2021, **8**, nwab051.
4 P. Cao, X. W. Pan, X. D. Su, Z. F. Liu and M. Li, *Curr. Opin. Struct. Biol.*, 2020, **63**, 49–57.
5 J. R. Shen, *Annu. Rev. Plant Biol.*, 2015, **66**, 23–48.
6 L. X. Shi, M. Hall, C. Funk and W. P. SchröDer, *Biochim. Biophys. Acta*, 2016, **1817**, 13–25.
7 C. Pagliano, G. Saracco and J. Barber, *Photosynth. Res.*, 2013, **116**, 167–188.
8 W. Wang, H. Wang, Q. Zhu, W. Qin, G. Han, J. R. Shen, X. Zong and C. Li, *Angew. Chem.*, 2016, **128**, 9375–9379.
9 O. Yehezkeli, R. Tel-Vered, J. Wasserman, A. Trifonov, D. Michaeli, R. Nechushtai and I. Willner, *Nat. Commun.*, 2012, **3**, 742.

10 F. Azadi-Chegeni, M. E. Ward, G. Perin, D. Simionato, T. Morosinotto, M. Baldus and A. Pandit, *Biophys. J.*, 2021, **120**, 270–283.

11 X. Wei, X. Su, P. Cao, X. Liu, W. Chang, M. Li, X. Zhang and Z. Liu, *Nature*, 2016, **534**, 69–74.

12 X. Sheng, A. Watanabe, A. J. Li, E. Kim, C. H. Song, K. Murata, D. F. Song, J. Minagawa and Z. F. Liu, *Nat. Plants*, 2019, **5**, 1320–1330.

13 X. Su, J. Ma, X. Wei, P. Cao, D. Zhu, W. Chang, Z. Liu, X. Zhang and M. Li, *Science*, 2017, **357**, 815–820.

14 P. Cao, X. Su, X. Pan, Z. Liu, W. Chang and M. Li, *Biochim. Biophys. Acta*, 2018, **1859**, 633–644.

15 R. Croce and H. van Amerongen, *Science*, 2020, **369**, eaay2058.

16 X. Sheng, Z. Liu, E. Kim and J. Minagawa, *Plant Cell Physiol.*, 2021, **62**, 1108–1120.

17 P. Cao, L. Bracun, A. Yamagata, B. M. Christianson, T. Negami, B. Zou, T. Terada, D. P. Canniffe, M. Shirouzu, M. Li and L. N. Liu, *Nat. Commun.*, 2022, **13**, 1977.

18 L. Bracun, A. Yamagata, B. M. Christianson, T. Terada, D. P. Canniffe, M. Shirouzu and L. N. Liu, *Sci Adv*, 2021, **7**, eabf8864.

19 N. Liguori, R. Croce, S. J. Marrink and S. Thallmair, *Photosynth. Res.*, 2020, **144**, 273–295.

20 A. V. Vener, *Biochim. Biophys. Acta*, 2007, **1767**, 449–457.

21 J. F. Allen, *Physiol. Plant.*, 2017, **161**, 28–44.

22 S. JäRvi, M. Suorsa and E. M. Aro, *Biochim. Biophys. Acta*, 2015, **1847**, 900–909.

23 A. Rokka, M. Suorsa, A. N. Battchikova and E. M. Aro, *Biochem. J.*, 2005, **388**, 159–168.

24 J. G. Garcia-Cerdan, L. Kovacs, T. Toth, S. Kereiche, E. Aseeva, E. J. Boekema, F. Mamedov, C. Funk and W. P. Schroder, *Plant J.*, 2011, **65**, 368–381.

25 M. Swiatek, *Plant Cell*, 2001, **13**, 1347–1368.

26 Y. Mohammadi, A. R. Haghighi and S. G. Sehrigh, *Int. J. BIOautom.*, 2016, **20**, 351–358.

27 L. Nierzwicki and G. Palermo, *Front. Mol. Biosci.*, 2021, **8**, 641208.

28 J. R. Perilla, B. C. Goh, C. K. Cassidy, B. Liu, R. C. Bernardi, T. Rudack, H. Yu, Z. Wu and K. Schulten, *Curr. Opin. Struct. Biol.*, 2015, **31**, 64–74.

29 A. Singharoy, C. Maffeo, K. H. Delgado-Magnero, D. J. K. Swainsbury, M. Sener, U. Kleinekathofer, J. W. Vant, J. Nguyen, A. Hitchcock, B. Isralewitz, I. Teo, D. E. Chandler, J. E. Stone, J. C. Phillips, T. V. Pogorelov, M. I. Mallus, C. Chipot, Z. Luthey-Schulten, D. P. Tieleman, C. N. Hunter, E. Tajkhorshid, A. Aksimentiev and K. Schulten, *Cell*, 2019, **179**, 1098–1111.

30 M. Faulkner, I. Szabó, S. L. Weetman, F. Sicard, R. G. Huber, P. J. Bond, E. Rosta and L.-N. Liu, *Sci. Rep.*, 2020, **10**, 17501.

31 L. Zhang, D. A. Silva, H. Zhang, A. Yue, Y. Yan and X. Huang, *Nat. Commun.*, 2014, **5**, 4170.

32 R. Mao, X. Wang and J. Gao, *Front. Chem.*, 2021, **9**, 764107.

33 N. Sakashita, H. C. Watanabe, T. Ikeda, K. Saito and H. Ishikita, *Biochemistry*, 2017, **56**, 3049–3057.

34 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.

35 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

36 M. Ceccarelli, P. Procacci and M. Marchi, *J. Comput. Chem.*, 2003, **24**, 129–142.

37 D. A. Giammona, PhD thesis, University of California, 1984.

38 P. M. Kasson, E. Lindahl and V. S. Pande, *PLoS Comput. Biol.*, 2010, **6**, e1000829.

39 P. F. Li and K. M. Merz, *J. Chem. Inf. Model.*, 2016, **56**, 599–604.

40 C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.

41 K. Ogata, M. Hatakeyama, Y. Sakamoto and S. Nakamura, *J. Phys. Chem. B*, 2019, **123**, 6444–6452.

42 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.

43 S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, *J. Chem. Phys.*, 1995, **103**, 4613–4621.

44 H. J. C. P. Berendsen, J. P. M. V. Postma, W. F. V. Gunsteren, A. D. Dinola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

45 S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952–962.

46 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

47 D. Greene, R. Qi, R. Nguyen, T. Qiu and R. Luo, *J. Chem. Inf. Model.*, 2019, **59**, 3041–3056.

48 L. Zhang, R. Tang, S. Bai, N. K. Connors, L. H. Lua, Y. P. Chuan, A. P. Middelberg and Y. Sun, *PLoS One*, 2014, **9**, e107313.

49 L. Zhang, R. Tang, S. Bai, N. K. Connors, L. H. Lua, Y. P. Chuan, A. P. Middelberg and Y. Sun, *J. Phys. Chem. B*, 2013, **117**, 5411–5421.

50 R. Mao, L. Bie, M. Xu, X. Wang and J. Gao, *Phys. Chem. Chem. Phys.*, 2021, **23**, 12549–12558.

51 V. Lafont, M. Schaefer, R. H. Stote, D. Altschuh and A. Dejaegere, *Proteins*, 2007, **67**, 418–434.

52 A. T. Graca, M. Hall, K. Persson and W. P. Schroder, *Sci. Rep.*, 2021, **11**, 15534.

53 R. A. Laskowski, M. W. Macarthur, D. S. Moss and J. M. Thornton, *J. Appl. Crystallogr.*, 1993, **26**, 283–291.

54 C. Colovos and T. O. Yeates, *Protein Sci.*, 1993, **2**, 1511–1519.

55 H. J. Bai, W. Z. Ma, S. Y. Liu and L. H. Lai, *Proteins*, 2008, **70**, 1323–1331.

56 W. Wang and P. A. Kollman, *J. Mol. Biol.*, 2000, **303**, 567–582.

57 L. Xiao, J. Diao, D. Greene, J. Wang and R. Luo, *J. Chem. Theory Comput.*, 2017, **13**, 3398–3412.

58 G. Archontis and T. Simonson, *Biophys. J.*, 2005, **88**, 3888–3904.

59 S. Genheden and U. Ryde, *Expert Opin. Drug Discovery*, 2015, **10**, 449–461.

60 H. J. Liu and R. E. Blankenship, *Biochim. Biophys. Acta*, 2019, **1860**, 148079.

61 F. Guerra, M. Siemers, C. Mielack and A. N. Bondar, *J. Phys. Chem. B*, 2018, **122**, 4625–4641.

62 M. Gur, E. Taka, S. Z. Yilmaz, C. Kilinc, U. Aktas and M. Golcuk, *J. Chem. Phys.*, 2020, **153**, 075101.

63 Y. Lu, *Front. Plant Sci.*, 2016, **7**, 168.

64 K. Qin, A. R. Fernie and Y. Zhang, *Biomolecules*, 2021, **11**, 1839.

65 L. s. Kovács, J. Damkjær, S. Kereïche, C. Ilioaia, A. V. Ruban, E. J. Boekema, S. Jansson and P. Horton, *Plant Cell*, 2006, **18**, 3106–3120.

66 P. Galka, S. Santabarbara, T. T. H. Khuong, H. Degand, P. Morsomme, R. C. Jennings, E. J. Boekema and S. Caffarri, *Plant Cell*, 2012, **24**, 2963–2978.

67 S. Rinalducci, M. R. Larsen, S. Mohammed and L. Zolla, *J. Proteome Res.*, 2006, **5**, 973–982.

© 2023 The Author(s). Published by the Royal Society of Chemistry