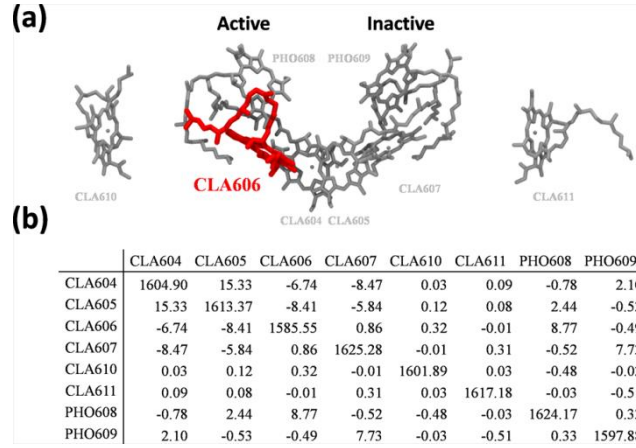
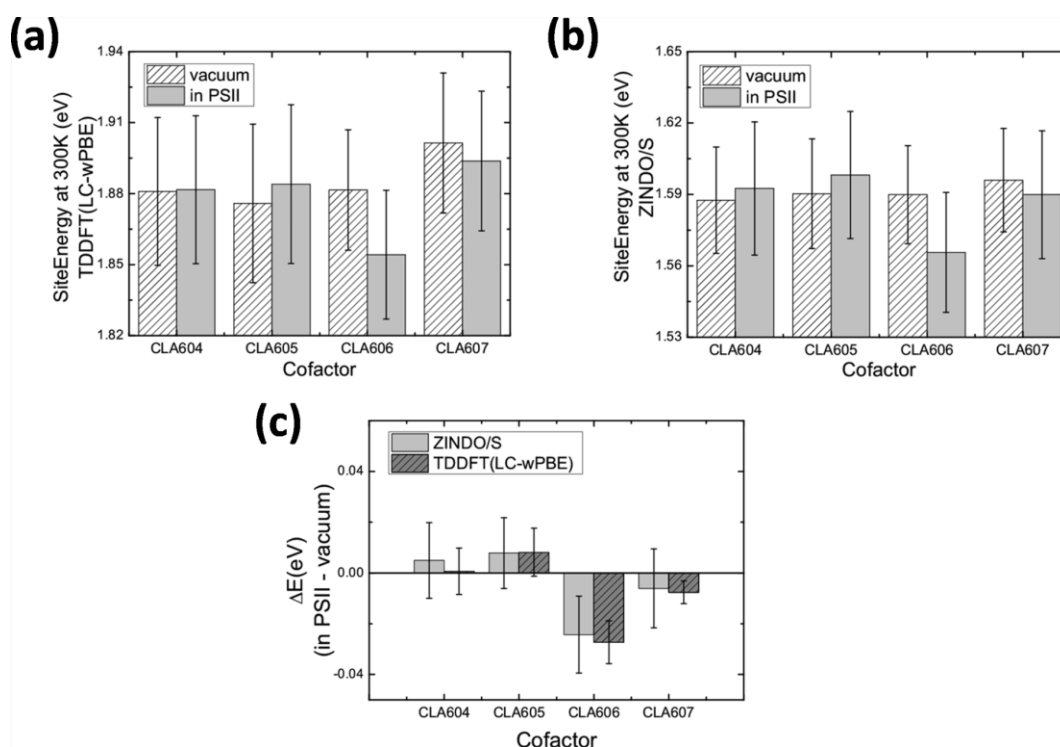


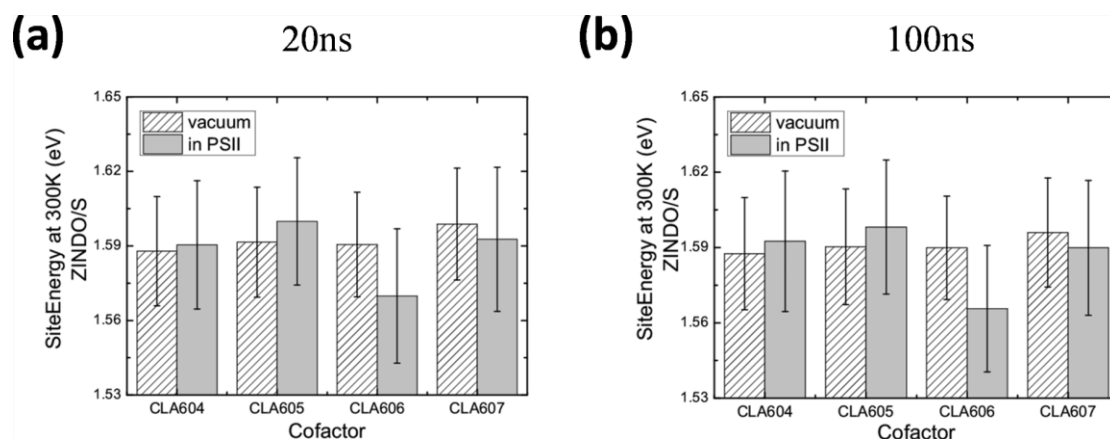
Supplementary Figure 1. RMSDs of the protein (C- α atoms) and cofactors (heavy atoms of the Pi-electron conjugate rings) in the reaction center as a function of simulation time at 300K are shown in (a) and (b) respectively. RMSDs at 77K are shown in (c)-(d).



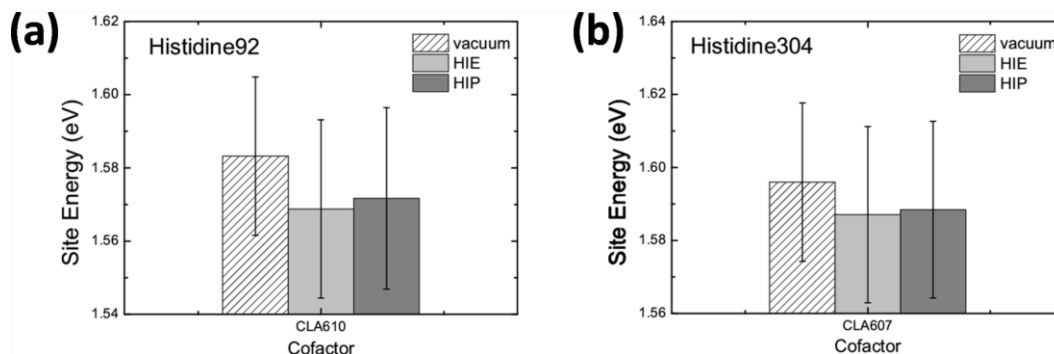
Supplementary Figure 2. (a) As revealed by the exciton model, CLA606 is the most probable (~70%) site for the excitation of the reaction center at 77K. The probabilities (P_m) are calculated as a weighed sum of each cofactor's contributions to all the excited states (M) obtained by the exciton model ($P_m = \sum_M f(M) |c_m^{(M)}|^2$). (b) Exciton matrix (unit: meV) built from 77K MD simulations. The diagonal terms are the average site energies for the cofactors and the off-diagonal terms correspond to the coupling strengths.



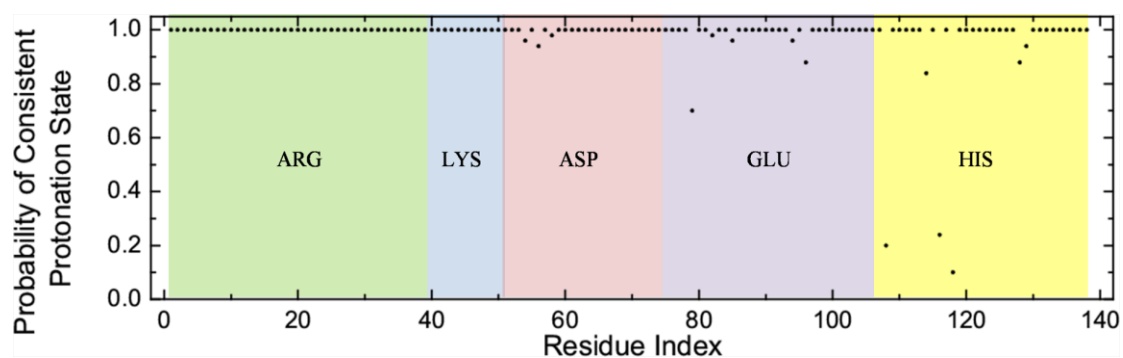
Supplementary Figure 3. (a) Comparison of site energies of the four CLAs in vacuum with those in PSII calculated by ZINDO/S; (b) The same as (a) except that the TDDFT (LC-wPBE) method is used for the calculations; (c) The site energy shift due to the presence of the PSII environment ($\Delta E = E_{\text{PSII}} - E_{\text{vacuum}}$) obtained ZINDO/S and TDDFT (LC-wPBE) are compared.



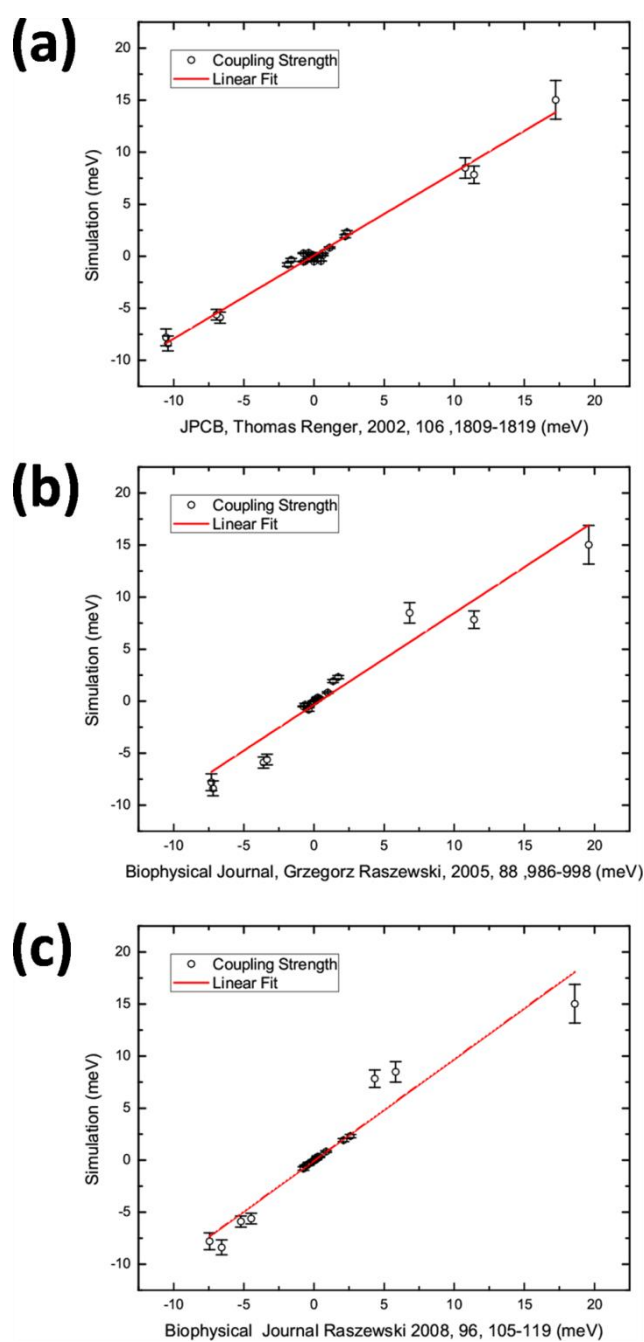
Supplementary Figure 4. (a) Site energies of the four CLAs in vacuum (striped bar) and in PSII (grey bar) calculated by ZINDO/S using five 20ns MD simulations; (b) the same as (a) except that two 100ns MD simulations are used for the calculation.



Supplementary Figure 5 (a) Comparison of site energies of CLA610 in vacuum (striped bar), in PSII with HIE92 (gray bar), and HIP92 (black bar). (b) The same as (a) except that PSII complex with different protonation states of His304 are investigated.



Supplementary Figure 6. Probability of the consistent protonation states adopted in our MD simulations and those predicted by the MCCE software obtained from 50 conformations. Please refer to Supplementary Table 4 for the residue index.



Supplementary Figure 7. The coupling strength calculated from the 300K MD simulations in this study is consistent with previous models^{35,52,53}.

Supplementary Table 1. The coordinated protein residues/waters for the Chlorophyll-a (CLA) molecules in the PSII reaction center.

CLA ResID	Coordinated Molecule	Coordinated Atom	CLA ResID	Coordinated Molecule	Coordinated Atom
604	His198	NE2	624	His26	NE2
605	His197	NE2	625	His9	NE2
606	Water	OW	626	His142	NE2
607	Water	OW	627	His114	NE2
610	His118	NE2	628	His237	NE2
611	His117	NE2	629	His430	NE2
612	CLA612	OW	630	His118	NE2
613	His201	ND1	631	Water	OW
614	His202	NE2	632	His441	NE2
615	His455	NE2	633	His251	NE2
616	His100	NE2	634	Water	OW
617	His157	ND1	635	His444	NE2
618	Water	OW	636	His53	NE2
619	His466	NE2	637	His56	NE2
620	His216	NE2	638	Asn39	OD1
621	Water	OW	639	His164	NE2
622	His469	NE2	640	His132	NE2
623	His23	NE2			

Supplementary Table 2. Calculated site energies of the eight cofactors in the PSII reaction centers in vacuum and in PSII at 300K.

Molecules	Site Energy (eV)			
	Vacuum		PSII	
	Average	STD	Average	STD
CLA604	1.588	0.022	1.593	0.028
CLA605	1.590	0.023	1.598	0.027
CLA606	1.590	0.021	1.566	0.025
CLA607	1.596	0.022	1.590	0.027
CLA610	1.583	0.022	1.578	0.029
CLA611	1.594	0.021	1.592	0.029
PHO608	1.532	0.027	1.619	0.042
PHO609	1.533	0.029	1.584	0.044

Supplementary Table 3. Calculated site energies of the eight cofactors in the PSII reaction centers in vacuum and in PSII at 77K.

Molecules	Site Energy (eV)			
	Vacuum		PSII	
	Average	STD	Average	STD
CLA604	1.611	0.014	1.605	0.018
CLA605	1.601	0.014	1.613	0.017
CLA606	1.602	0.014	1.586	0.022
CLA607	1.616	0.014	1.625	0.020
CLA610	1.597	0.014	1.602	0.022
CLA611	1.600	0.014	1.617	0.020
PHO608	1.542	0.019	1.624	0.035
PHO609	1.540	0.020	1.598	0.031

Supplementary Table 4. Comparison between protonation states adopted in our MD simulations and those predicted by the MCCE software. <pKa> are obtained by averaging predictions made by the MCCE software using 50 conformations.

index	Residues	<pKa>	predicted state	used state	index	Residues	<pKa>	predicted state	used state	index	Residues	<pKa>	predicted state	used state
1	A0016	12.24	ARG	ARG	47	I0005	12.67	LYS	LYS	93	D0219	1.14	GLU	GLU
2	A0027	14.00	ARG	ARG	48	O0069	11.16	LYS	LYS	94	D0242	3.38	GLU	GLU
3	A0064	14.00	ARG	ARG	49	O0160	13.95	LYS	LYS	95	D0302	0.00	GLU	GLU
4	A0129	13.78	ARG	ARG	50	V0134	14.00	LYS	LYS	96	D0312	2.25	GLU	GLU
5	A0136	14.00	ARG	ARG	51	X0008	10.25	LYS	LYS	97	D0323	0.00	GLU	GLU
6	A0140	13.94	ARG	ARG	52	A0025	3.81	ASP	ASP	98	D0337	1.53	GLU	GLU
7	A0257	14.00	ARG	ARG	53	A0059	0.01	ASP	ASP	99	E0007	0.83	GLU	GLU
8	A0269	13.92	ARG	ARG	54	A0061	1.48	ASP	ASP	100	E0071	4.64	GLU	GLU
9	A0323	14.00	ARG	ARG	55	A0103	0.44	ASP	ASP	101	E0077	3.61	GLU	GLU
10	A0334	13.99	ARG	ARG	56	A0170	0.88	ASP	ASP	102	H0047	4.67	GLU	GLU
11	C0207	13.66	ARG	ARG	57	A0319	0.00	ASP	ASP	103	I0002	3.20	GLU	GLU
12	C0261	13.32	ARG	ARG	58	A0342	0.28	ASP	ASP	104	O0074	3.58	GLU	GLU
13	C0262	13.36	ARG	ARG	59	C0360	2.28	ASP	ASP	105	T0002	1.31	GLU	GLU
14	C0357	14.00	ARG	ARG	60	B0477	0.68	ASP	ASP	106	T0025	1.68	GLU	GLU
15	C0447	13.97	ARG	ARG	61	B0483	1.27	ASP	ASP	107	B0094	2.66	GLU	GLU
16	C0449	13.96	ARG	ARG	62	D0016	1.06	ASP	ASP	108	A0092	8.01	HIP	HIS
17	B0272	14.00	ARG	ARG	63	D0019	0.06	ASP	ASP	109	A0118	0.00	HIS	HIS
18	B0326	14.00	ARG	ARG	64	D0020	0.03	ASP	ASP	110	A0190	13.59	HIP	HIP
19	B0472	14.00	ARG	ARG	65	D0025	0.33	ASP	ASP	111	A0195	0.00	HIS	HIS
20	D0024	13.99	ARG	ARG	66	D0100	1.92	ASP	ASP	112	A0198	0.00	HIS	HIS
21	D0026	12.69	ARG	ARG	67	D0297	0.00	ASP	ASP	113	A0215	0.00	HIS	HIS
22	D0103	14.00	ARG	ARG	68	D0333	0.00	ASP	ASP	114	A0252	6.33	HIS	HIS
23	D0128	14.00	ARG	ARG	69	E0012	1.15	ASP	ASP	115	A0272	0.00	HIS	HIS
24	D0134	14.00	ARG	ARG	70	E0045	2.67	ASP	ASP	116	A0304	7.99	HIP	HIS
25	D0139	14.00	ARG	ARG	71	E0068	1.87	ASP	ASP	117	A0332	0.00	HIS	HIS
26	D0180	14.00	ARG	ARG	72	I0027	0.28	ASP	ASP	118	A0337	11.74	HIP	HIS
27	D0251	14.00	ARG	ARG	73	X0035	1.96	ASP	ASP	119	C0237	0.00	HIS	HIS
28	D0265	13.30	ARG	ARG	74	B0087	3.72	ASP	ASP	120	C0251	0.00	HIS	HIS
29	D0294	14.00	ARG	ARG	75	A0015	3.29	GLU	GLU	121	C0430	0.00	HIS	HIS
30	D0326	14.00	ARG	ARG	76	A0098	2.94	GLU	GLU	122	C0441	0.00	HIS	HIS
31	D0348	14.00	ARG	ARG	77	A0104	0.75	GLU	GLU	123	C0444	0.00	HIS	HIS
32	E0069	14.00	ARG	ARG	78	A0132	0.00	GLU	GLU	124	B0201	0.00	HIS	HIS

33	F0019	14.00	ARG	ARG	79	A0189	3.74	GLU	GLU	125	B0455	0.00	HIS	HIS
34	I0030	13.70	ARG	ARG	80	A0244	0.69	GLU	GLU	126	B0466	0.00	HIS	HIS
35	L0014	14.00	ARG	ARG	81	A0329	0.51	GLU	GLU	127	B0469	0.00	HIS	HIS
36	O0073	14.00	ARG	ARG	82	A0333	1.09	GLU	GLU	128	D0061	6.01	HIS	HIS
37	O0115	14.00	ARG	ARG	83	C0221	3.62	GLU	GLU	129	D0087	6.44	HIS	HIS
38	T0024	13.21	ARG	ARG	84	C0348	0.00	GLU	GLU	130	D0117	0.00	HIS	HIS
39	T0028	13.86	ARG	ARG	85	C0354	1.55	GLU	GLU	131	D0189	0.00	HIS	HIS
40	C0215	12.96	LYS	LYS	86	C0413	0.00	GLU	GLU	132	D0197	0.00	HIS	HIS
41	B0321	12.84	LYS	LYS	87	C0456	0.15	GLU	GLU	133	D0214	0.00	HIS	HIS
42	B0498	13.98	LYS	LYS	88	B0364	0.00	GLU	GLU	134	D0268	0.00	HIS	HIS
43	D0023	13.34	LYS	LYS	89	B0492	3.29	GLU	GLU	135	D0336	0.07	HIS	HIS
44	D0264	14.00	LYS	LYS	90	D0069	0.23	GLU	GLU	136	E0023	0.00	HIS	HIS
45	D0317	13.54	LYS	LYS	91	D0096	0.00	GLU	GLU	137	F0024	0.00	HIS	HIS
46	E0073	11.83	LYS	LYS	92	D0131	0.54	GLU	GLU	138	B0100	0.00	HIS	HIS

Supplementary Table 5. Transition charges of the heavy atoms in chlorophyll-a (CLA).

Heavy Atom	Transition Charge (10 ⁻³ a.u.)	Heavy Atom	Transition Charge (10 ⁻³ a.u.)
MG	-1.894	CBB	46.82
CHA	90.271	NC	-9.557
CHB	-62.995	C1C	149.735
CHC	-147.722	C2C	-43.008
CHD	97.711	C3C	49.929
NA	4.767	C4C	-109.913
C1A	-110.93	CMC	-3.008
C2A	-43.189	CAC	-8.925
C3A	26.176	CBC	9.819
C4A	118.865	ND	64.76
CMA	2.357	C1D	-93.646
CAA	29.619	C2D	-61.496
CBA	-14.464	C3D	-14.39
CGA	-0.876	C4D	-71.773
O1A	-0.215	CMD	-10.571
O2A	1.107	CAD	-4.562
NB	-83.358	OBD	-30.959
C1B	91.813	CBD	-35.788
C2B	18.399	CGD	26.717
C3B	21.035	O1D	-12.033
C4B	136.789	O2D	-4.184
CMB	14.71	CED	-7.358
CAB	-12.938	C1	-1.644

Supplementary Table 6. Transition charges of the heavy atoms in pheophytin (PHO).

Heavy Atom	Transition Charge (10 ⁻³ a.u.)	Heavy Atom	Transition Charge (10 ⁻³ a.u.)
O2A	1.129	CBD	-37.024
CGA	-1.677	CGD	26.28
O1A	0.555	O1D	-10.407
CBA	-11.228	O2D	-4.426
CAA	24.597	CED	-6.803
C2A	-30.733	CAD	-3.767
C3A	19.225	OBD	-29.2
CMA	1.284	C3D	-18.891
C4A	86.228	C2D	-62.016
CHB	-30.642	CMD	-3.733
C1B	36.691	C1D	-53.312
NB	-54.169	CHD	83.394
C4B	81.492	C4C	-87.215
C3B	24.854	C3C	22.875
C2B	31.144	CAC	-5.093
CMB	11.317	CBC	6.59
CAB	-15.068	C2C	-22.848
CBB	40.579	CMC	-2.518
NA	7.639	NC	-17.917
C1A	-87.81	C1C	129.146
CHA	53.418	CHC	-115.944
C4D	-13.471	C1	-1.483
ND	38.954		

Supplementary Methods

Molecular dynamics simulations

We have constructed the system by embedding the PSII complex (PDBID:3ARC¹) into the membrane. To smooth the contact between the membrane and the protein, we performed a 10000-steps energy minimization with the steepest descent algorithm² by freezing the PSII complex. Furthermore, another 5000 steps energy minimization was performed for the whole system. Next, to further relax the lipids surrounding the PSII complex, position restraint with a force constant of $10 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ was enforced on all the heavy atoms of the PSII complex and the whole system was simulated for 5ns under NVT ensemble (T=300 or 77K), followed by another 10ns simulation under NPT ensemble (T=300 or 77K and P=1bar). Afterwards, the restraint was released and Molecular dynamics (MD) simulations were performed for PSII in the POPC membrane at both 300K and 77K under the NPT ensemble. At each temperature, five independent simulations were performed starting from the last configuration of the position restraint simulation with different initial velocities. We save the snapshots every 20ps. In the NPT simulations, we applied V-rescale thermostat³ and Parrinello-Rahman barostat⁴ with the coupling time constant of 0.1ps and 2.0ps, respectively. The long-range electrostatic interactions beyond the cut-off at 12Å were treated with the Particle-Mesh Ewald (PME) method⁵. The Lennard-Jones interactions were smoothly switched off from 10Å to 11Å. The neighbors list was updated every 10 steps. An integration time step of 2.0ps was used and the LINCS algorithm⁶ was applied to constrain all the bonds.

To examine if our MD simulations are sufficiently long to provide a reasonable structural ensemble, we have extended two of our MD simulations to 100ns at 300K (system size ~580,000 atoms). The RMSDs of both protein and cofactors with

respect to the crystal structure reaches plateau of $\sim 1.7\text{\AA}$ at around 20ns, indicating the system is fully equilibrated (see Supplementary Fig. 1a-b). To calculate the RMSD of cofactors (see Fig. 2a and Supplementary Fig. 1b), we first make the alignment of the MD conformations to the crystal structure using the C_α atoms of protein and then calculated the RMSD of the individual co-factors compared to their conformations in the crystal structure. To check the equilibrium of the system, we have left out the tail region of the cofactors in the RMSD calculations (Supplementary Fig. 1b) since they are high flexible, but have little effect on the site energy calculations^{7,8}.

In order to compare with the experimental results, we also performed MD simulations of the Thr179Glu mutant. The mutation of Thr to Glu is modeled with PYMOL⁹ based on the last frame from the wildtype positive restraint simulation. To allow the structural relaxation of Glu, a 1000-step energy minimization was performed with the steepest descent algorithm² followed by a 100ps NVT simulation ($T=300\text{K}$) by only allowing the Glu residue to move. We next performed a 5000-steps energy minimization with the whole system flexible. Finally we performed a 20ns production MD simulation under NPT ensemble ($T=300\text{K}$ and $P=1\text{bar}$) with the temperature annealing from 50K to 300K at the first nanosecond. All other parameters for the MD simulations are identical with the wildtype simulations.

Site energy calculations

We adopted a hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) method by treating the cofactors quantum mechanically while the protein environment as cloud of partial charges when computing the electronic excitations of the PSII cofactors. We chose the Zerner's Intermediate Neglect of Differential Orbital with parameters for Spectroscopic properties (ZINDO/S)^{10,11} method to compute the

excitation energies of cofactors due to its computational efficiency. In particular, we employed the **ORCA code** (University of Bonn, Germany)¹² for these calculations and considered only the excitation energy from the ground state to the first excited state (Q_y state).

For each cofactor, we performed calculations on a total of 1250 MD conformations (250 snapshots from the last 5-ns of each of the five MD simulations). Furthermore, since the PSII complex contains two monomers, we computed the excitation energies for each monomer independently.

For each MD snapshot, we have included point charges of those atoms within 22Å of each cofactor in our QM/MM calculation. This distance cut-off of around 20 Å for the MM region has been suggested to be an optimal number in previous studies of FMO complex¹³⁻¹⁵ and bacterial antenna system¹⁶. Previous studies have shown that the lipids' major role is in structural assembly and the placement of reduced plastoquinone, but not in determining the electron transfer pathway in the reaction center²⁵⁻²⁹. In addition, different crystal structures contain different composition of lipids depending on the organisms and crystallography condition¹⁰⁻²⁴. **Therefore, we did not include the contributions of lipids to the excitation energies.**

In order to elucidate the contributions of different components to the excitation energy, we have also performed separate calculations by only considering point charges from sub-groups of atoms in the environment such as protein, cofactors, waters, charged residues and neutral residues. As shown in Fig. 5 and Table 1, the charged protein residues play a critical role to stabilize the excitation of CLA606.

The above observation that the presence of charged residues in proximity to CLA606 can greatly alter its excitation energy has been found by previous site-directed

mutagenesis experiments. For example, Thr179 forms interactions with CLA606 through a bridging water molecule (see Fig. 2c). When the neutral Thr179 is mutated to the charged Glu, there is a blue shift of $\sim 1\text{nm}$ in the absorption difference spectrum¹⁷, indicating a higher excitation energy of CLA606. We have performed mutant MD simulations and site energy calculations. Our results show that the site energy of CLA606 is indeed raised by $\sim 0.005\text{eV}$ in the mutant system, though there is a large uncertainty (with the standard deviation of $\sim 0.02\text{eV}$) in our site energy calculations. These results also confirm that alternation of charge environment especially the presence of charged residues could substantially tune the site energy of CLA606. In previous studies, the Thr179His mutant has also been found to greatly alter the site energy of CLA606^{17,18}. This effect is suggested to be largely caused by the dispersion interactions^{17,18}, which is not considered in our QM/MM calculations where the protein environment is treated as the point charge cloud.

In order to pinpoint specific charged residues that may largely affect the excitation energy of CLA606, we have performed additional calculations by only including point charges from each individual residue within 22\AA cut-off in the MM region. From these calculations, we have identified seven residues that could lower the site energy of CLA606 by over 0.0035eV , and all these residues are charged (see Fig. 6a).

We performed further calculations by separately mutating each of these seven charged residues to alanine (Ala) to mimic the mutagenesis experiment. In these calculations, we adopted a residue-based cut-off in the MM region, i.e. those residues with their center of mass (averaged over all the MD conformations) lies within the 22 \AA cut-off were included in the MM region. To model each mutant, we simply replaced the point charges for the residue in the wildtype to those in Ala. We found that all the seven single mutants are able to raise the site energy of CLA606 at a certain extend (see Fig.

6b). Interestingly, if one collects the effect of these seven single mutants, the site energy CLA606 is raised to the similar level of its counterpart CLA607 in the inactive chain (see “CLA606-collective” in Fig. 6b). These results demonstrated that these seven residues play an essential role and work collectively to stabilize the excitation energy of CLA606.

In order to validate the site energies obtained from ZINDO/S, we have employed the Time-Dependent Density Functional Theory (TDDFT) in our QM/MM scheme for the site energy calculations. To select the proper density functional in the TDDFT calculations, we have first compared the computed the Q_y excitation energy of the chlorophyll-a in diethyl ether using different density functionals to the experimental value¹⁹. In particular, we have selected four density functionals that have been previously applied for site energy calculations of cofactors²⁰⁻²³: B3LYP, B3PW91, cam-B3LYP, and LC-wPBE. The initial conformation was taken from the small crystal structure of chlorophyll-a²⁴ followed by an energy optimization using B3LYP/6-31G(d). Our results show that LC-wPBE (1.93eV) has the best agreement with the experimental value (1.87eV). While the site energies obtained using B3LYP (2.10eV) and B3PW91 (2.11eV) are substantially higher than the experimental value. The value from cam-B3LYP (2.05eV) shows better agreement with experiment than B3LYP and B3PW91, but still produces a noticeable higher value than the experiment. Therefore, we have selected LC-wPBE for our further TDDFT calculations. All the calculations are performed using the 6-31G(d) basis set.

As shown in Supplementary Fig. 3a, in TDDFT (LC-wPBE) calculations, the site energy of the active chain cofactor CLA606 is greatly reduced in the presence of the protein environment, while its counterpart in the inactive chain CLA607 displays only a small decrease in its site energy. This observation from TDDFT calculations is

consistent with the results obtained from ZINDO/S calculations (see Supplementary Fig. 3b). Moreover, two sets of calculations show reasonable agreement in the magnitude of site energy difference due to the presence the protein environment for CLA606 and CLA607 (see Supplementary Fig. 3c). As a control, we have also performed the same calculations on a different pair of cofactors CLA604 and CLA605. Again, the TDDFT calculation results agree well with those from ZINDO/S. Specifically, the protein environment has very small effect on the site energies of these two cofactors (see Supplementary Fig. 3). Due to the limitation of computing resources, the site energy of each cofactor in TDDFT calculations is obtained from 25 representative MD conformations.

The hybrid QM/MM method with the TDDFT theory has been successfully applied to compute the site energies of cofactors in photosynthetic systems²⁵. However, since TDDFT is computationally much more expensive than ZINDO/S, the semi-empirical ZINDO/S method may still be a more feasible choice for the investigation the effect of dynamic protein conformational ensemble on the site energies, where thousands of conformations are used to obtain the site energy of each cofactor in the PSII reaction site. ZINDO/S has already been extensively used in previous studies to compute low lying optically allowed excited states of pigment in various photosynthetic systems^{8,13-15,26-30}.

To investigate if our MD simulations (at 20ns) were sufficiently long to provide a reasonable protein ensemble for the site energy calculations, we have extended two of the MD simulations to 100ns and selected 500 MD conformations to re-compute the site energies from the last 5ns of these two simulations. As shown in Supplementary Fig. 4, the results are consistent with our previous calculations using 20ns MD simulations. In particular, the protein environment can largely reduce the site energy

of CLA606 in the active chain, but has minor effect on its counterpart CLA607 in the inactive chain (see Supplementary Fig. 4b). On the other hand, the protein environment has little effect on the site energy of CLA604 (active chain) and CLA605 (inactive chain). From these observations, we conclude that the conformations we selected in our site energy calculations represent a reasonable protein structural ensemble.

Our calculations show that the site energy of CLA606 is the lowest among all the cofactors in the RC and this is consistent with previous studies³¹⁻³⁶ to obtain the site energies by fitting to experimental spectra. However, we also notice that our calculations of the excitation energies for other cofactors, especially for PHO608 and CLA610 show differences from the previous theoretical analysis. We believe that these discrepancies may arise from the comparison of two different protein systems; i.e., the minimal scaffold PSII (denoted also as D1-D2-cyt559) in experiments, *versus* the whole PSII complex in our simulations. Experimental spectrums of PSII are acquired for the minimal scaffold. Thus, the excitation energies derived by fitting to experimental spectrums are applied for this simplified PSII central complex where the arrangement of various protein domains could be different to the full PSII complex. Unfortunately, MD simulations on the minimal scaffold D1-D2-cyt559 are unfeasible now due to the lack of a crystallographic structure. In our work the MD simulations were performed on the whole PSII complex, based on its most recent X-ray crystallography structure, which includes protein subunits (such as the core antenna CP47 and CP43) that are not present in the minimal scaffold of PSII used in spectroscopic experiments and the subsequent theoretical analyses. We believe that the presence of these subunits may significantly perturb the PSII conformations. Consequently, the interactions in the whole PSII complex are likely to be different

from those in the minimal scaffold. Therefore, we believe that it is reasonable to find some differences in the excitation energies calculated by using our method on the whole PSII complex and those values calculated by fitting to the spectrum of D1-D2-cyt559.

Protonation State of Charged Residues

We found that a few charged residues work collectively to reduce the site energies of CLA606 (see Fig. 6). However, the buried charged residues could have large pKa shift and thus take non-standard protonation states. In some cases, the protonation states may even vary under different protein conformations. In order to investigate the protonation states of charged residues, we have performed pKa calculations for all the ARG, LYS, ASP, GLU and HIS residues within 22Å of the eight reaction center cofactors using the MCCE software³⁷. In order to examine the effect of dynamic protein conformational ensemble on the protonation states of these residues, we have selected 50 representative conformations (5 conformations separated by 1ns from the last 5ns of the 5x20ns MD simulations). As shown in Supplementary Table 4, the predicted protonation states from the average pKa values agree well with what we have adopted in our MD simulations except three chain A residues: HIS92, HIS304, and HIS337 (highlighted in Supplementary Table 4).

For HIS337, the MCCE program suggests a positively charged protonation state (HIP), while we used a neutral form (HIE). The crystal structure shows that HIS337 is one of the coordination ligands of the OEC complex, and this coordination is through the N_δ atom of the histidine. Therefore, it is crucial for HIS337 to adopt a neutral protonation state and leave the N_δ atom for its interaction with the OEC complex.

For HIS92 and HIS304, we have performed calculations of site energies with the suggested protonation states (HIP) by the MCCE program. As shown in Supplementary Fig. 5, the alteration of the protonation states for residue HIS92 and HIS304 has negligible effect on the site energy of the nearby cofactors: CLA610 and CLA607 respectively. We also noticed that the locations of these two residues are at least 10Å away from any cofactor. These observations suggest that the protonation states of these two histidine residues may not play a large role on determining the cofactors site energies.

In order to investigate the effect of dynamic protein conformational ensemble on the protonation states, we have also computed the probabilities of the predictions from the MCCE program that match with our MD simulations (see Supplementary Fig. 6). The results indicate that the predicted protonation states for all the ARG and LYS residues do not vary with conformations and agree with our MD simulations. For GLU and ASP, their protonation states display small variations, but in general agree well with those used in our MD simulations (over 80% except GLU189). For GLU189, the MCCE predicts that it has a small probability (30%) to adopt the protonated form, different from that used in our MD simulations. We noticed that this residue also coordinates with the OEC complex, where a protonated form may not be favored since it may increase the steric effect and reduces the ability to form the coordination. For HIS, similar results with GLU and ASP are obtained except the three residues (HIS92, HIS304, and HIS337), which we have discussed in the previous paragraph. The above results show that our MD simulations adopt a reasonable set of residue protonation states for the subsequent site energy calculations.

Coupling strength calculations

We employed the Transition Charges from Electrostatic Potentials (TrEsp) method^{38,39} to determine the coupling strength between the cofactors (see Equation (1) in the main text). The TrEsp approach has been widely applied to other photosynthetic systems^{13-16,27,40}.

In Equation (1), the atomic transition charges $q_I^T(q_J^T)$ were calculated following a similar scheme used before³⁸: (1) The geometry was optimized by a DFT method with B3LYP functional and 6-31G** basis set; (2) The same TDDFT⁴¹⁻⁴⁵ method was used to calculate the transition charges by fitting to the excitation transition density using the CHELPG scheme⁴⁶. The Gaussian03 program⁴⁷ was used for all the DFT calculations and the calculated transition charges were listed in Supplementary Table 5 and 6. Because TDDFT/B3LYP typically overestimates the transition dipole moment, the transition charges were scaled by a factor of 0.769³⁸.

The position of each atom $R_n^J(R_m^I)$ was taken from the MD conformations (see Equation (1) in the main text). The distance-dependent factor f was derived from the distance d between pairs of cofactors in the MD conformations by the following scheme: (1) For CLA, the magnesium was selected to measure its distance d to other cofactors; (2) For PHO, the center of mass of the four nitrogen atoms was used to determine d between the respective PHO and other cofactors.

Our results showed that the coupling strengths between all the pairs of cofactors were smaller than 20meV. The strongest coupling at ~15meV was observed between the “special pair” (P_{D1} - P_{D2}), which agree well with the experimental data of 10meV~20meV⁴⁸⁻⁵². It is clearly shown that the coupling strength of the special pair was not significantly greater than other pairs, which explained the serious overlap in the absorption spectrum of the PSII complex⁴⁸⁻⁵². The coupling strengths from our

calculations were similar to the previous calculations using different methods<sup>35,38,39,52-
⁵⁴ (see Supplementary Fig. 7).</sup>

Supplementary References

- 1 Umena, Y., Kawakami, K., Shen, J. R. & Kamiya, N. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 angstrom. *Nature* **473**, 55-U65 (2011).
- 2 Lewis, J. M., S. Lakshmivarahan, & Dhall., S. *Optimization: steepest descent method Dynamic Data Assimilation*. (Cambridge University Press, 2006).
- 3 Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys* **126**, 014101 (2007).
- 4 Parrinello, M. & Rahman, A. Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *J Appl Phys* **52**, 7182-7190 (1981).
- 5 Essmann, U. *et al.* A Smooth Particle Mesh Ewald Method. *J Chem Phys* **103**, 8577-8593 (1995).
- 6 Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* **18**, 1463-1472 (1997).
- 7 Cory, M. G., Zerner, M. C., Xu, X. C. & Shulten, K. Electronic excitations in aggregates of bacteriochlorophylls. *J Phys Chem B* **102**, 7640-7650 (1998).
- 8 Janosi, L., Kosztin, I. & Damjanovic, A. Theoretical prediction of spectral and optical properties of bacteriochlorophylls in thermally disordered LH2 antenna complexes. *J Chem Phys* **125** (2006).
- 9 Schrodinger, LLC. *The PyMOL Molecular Graphics System, Version 1.3r1* (2010).
- 10 Ridley, J. & Zerner, M. An intermediate neglect of differential overlap technique for spectroscopy: Pyrrole and the azines. *Theor Chim Acta* **32**, 111-134 (1973).
- 11 Thompson, M. A. & Zerner, M. C. A Theoretical-Examination of the Electronic-Structure and Spectroscopy of the Photosynthetic Reaction Center from Rhodospseudomonas-Viridis. *J Am Chem Soc* **113**, 8210-8215 (1991).
- 12 Petrenko, T. & Neese, F. Analysis and prediction of absorption band shapes, fluorescence band shapes, resonance Raman intensities, and excitation profiles using the time-dependent theory of electronic spectroscopy. *J Chem Phys* **127**, 164319 (2007).
- 13 Olbrich, C. *et al.* From Atomistic Modeling to Excitation Transfer and Two-Dimensional Spectra of the FMO Light-Harvesting Complex. *J Phys Chem B* **115**, 8609-8621 (2011).
- 14 Olbrich, C., Strumpfer, J., Schulten, K. & Kleinekathofer, U. Quest for Spatially Correlated Fluctuations in the FMO Light-Harvesting Complex. *J Phys Chem B* **115**, 758-764 (2011).

- 15 Olbrich, C., Strumpfer, J., Schulten, K. & Kleinekathofer, U. Theory and Simulation of the Environmental Effects on FMO Electronic Transitions. *J Phys Chem Lett* **2**, 1771-1776 (2011).
- 16 Olbrich, C. & Kleinekathofer, U. Time-Dependent Atomistic View on the Electronic Relaxation in Light-Harvesting System II. *J Phys Chem B* **114**, 12427-12437 (2010).
- 17 Schlodder, E. *et al.* Site-directed mutations at D1-Thr179 of photosystem II in *Synechocystis* sp PCC 6803 modify the spectroscopic properties of the accessory chlorophyll in the D1-branch of the reaction center. *Biochemistry* **47**, 3143-3154 (2008).
- 18 Schlodder, E. *et al.* Site-directed mutations at D1-His198 and D1-Thr179 of photosystem II in *Synechocystis* sp PCC 6803: deciphering the spectral properties of the PSII reaction centre. *Philos T R Soc B* **363**, 1197-1202 (2008).
- 19 Shipman, L. L., Cotton, T. M., Norris, J. R. & Katz, J. J. Analysis of Visible Absorption-Spectrum of Chlorophyll a Monomer, Dimer, and Oligomers in Solution. *J Am Chem Soc* **98**, 8222-8230 (1976).
- 20 Sundholm, D. A density-functional-theory study of bacteriochlorophyll b. *Phys Chem Chem Phys* **5**, 4265-4271 (2003).
- 21 Vokacova, Z. & Burda, J. V. Computational study on spectral properties of the selected pigments from various photosystems: Structure-transition energy relationship. *J Phys Chem A* **111**, 5864-5878 (2007).
- 22 Balanay, M. P., Lee, S. H., Yu, S. C. & Kim, D. H. Application of Long-range-corrected Density Functional in Metallated Porphyrin Analogues for Dye-sensitized Solar Cells. *B Korean Chem Soc* **32**, 705-708 (2011).
- 23 Zhao, Y. D., Fu, J. J., Li, H. B., Dong, H. & Liao, Y. Photoinduced charge transfer processes of zinc porphyrin derivatives for dye-sensitized solar cells. *Chem Res Chinese U* **29**, 974-981 (2013).
- 24 Kratky, C. & Dunitz, J. D. Methylchlorophyllide A dihydrate. *Acta Crystallographica Section B* **33**, 545-547 (1977).
- 25 Shim, S., Rebentrost, P. & Aspuru-Guzik, A. Atomistic simulations of photosynthetic excitonic energy transfer. *Abstr Pap Am Chem S* **242** (2011).
- 26 Gao, J. K. *et al.* QM/MM Modeling of Environmental Effects on Electronic Transitions of the FMO Complex. *J Phys Chem B* **117**, 3488-3495 (2013).
- 27 Jing, Y., Zheng, R., Li, H.-X. & Shi, Q. Theoretical Study of the Electronic-Vibrational Coupling in the Qy States of the Photosynthetic Reaction Center in Purple Bacteria. *The Journal of Physical Chemistry B* **116**, 1164-1171 (2011).

- 28 Zwier, M. C., Shorb, J. M. & Krueger, B. P. Hybrid molecular dynamics-quantum mechanics simulations of solute spectral properties in the condensed phase: Evaluation of simulation parameters. *J Comput Chem* **28**, 1572-1581 (2007).
- 29 Damjanovic, A., Kosztin, I., Kleinekathofer, U. & Schulten, K. Excitons in a photosynthetic light-harvesting system: A combined molecular dynamics, quantum chemistry, and polaron model study. *Phys Rev E* **65** (2002).
- 30 Linnanto, J. & Korppi-Tommola, J. Quantum chemical simulation of excited states of chlorophylls, bacteriochlorophylls and their complexes. *Phys Chem Chem Phys* **8**, 663-687 (2006).
- 31 Shibata, Y., Nishi, S., Kawakami, K., Shen, J. R. & Renger, T. Photosystem II Does Not Possess a Simple Excitation Energy Funnel: Time-Resolved Fluorescence Spectroscopy Meets Theory. *J Am Chem Soc* **135**, 6903-6914 (2013).
- 32 Lewis, K. L. M. *et al.* Simulations of the Two-Dimensional Electronic Spectroscopy of the Photosystem II Reaction Center. *J Phys Chem A* **117**, 34-41 (2013).
- 33 Gelzinis, A. *et al.* Tight-binding model of the photosystem II reaction center: application to two-dimensional electronic spectroscopy. *New J Phys* **15** (2013).
- 34 Novoderezhkin, V. I., Romero, E., Dekker, J. P. & van Grondelle, R. Multiple Charge-Separation Pathways in Photosystem II: Modeling of Transient Absorption Kinetics. *Chemphyschem* **12**, 681-688 (2011).
- 35 Raszewski, G., Saenger, W. & Renger, T. Theory of optical spectra of photosystem II reaction centers: Location of the triplet state and the identity of the primary electron donor. *Biophys J* **88**, 986-998 (2005).
- 36 Renger, T. & Schlodder, E. Primary Photophysical Processes in Photosystem II: Bridging the Gap between Crystal Structure and Optical Spectra. *Chemphyschem* **11**, 1141-1153 (2010).
- 37 Song, Y. F., Mao, J. J. & Gunner, M. R. MCCE2: Improving Protein pK(a) Calculations with Extensive Side Chain Rotamer Sampling. *J Comput Chem* **30**, 2231-2247, doi:Doi 10.1002/Jcc.21222 (2009).
- 38 Madjet, M. E., Abdurahman, A. & Renger, T. Intermolecular Coulomb couplings from ab initio electrostatic potentials: Application to optical transitions of strongly coupled pigments in photosynthetic antennae and reaction centers. *J Phys Chem B* **110**, 17268-17281 (2006).
- 39 Renger, T. Theory of excitation energy transfer: from structure to function. *Photosynthesis Research* **102**, 471-485 (2009).

- 40 Madjet, M. E., Muh, F. & Renger, T. Deciphering the Influence of Short-Range Electronic Couplings on Optical Properties of Molecular Dimers: Application to "Special Pairs" in Photosynthesis. *J Phys Chem B* **113**, 12603-12614 (2009).
- 41 Bauernschmitt, R. & Ahlrichs, R. Treatment of electronic excitations within the adiabatic approximation of time dependent density functional theory. *Chem Phys Lett* **256**, 454-464 (1996).
- 42 Casida, M. E., Jamorski, C., Casida, K. C. & Salahub, D. R. Molecular excitation energies to high-lying bound states from time-dependent density-functional response theory: Characterization and correction of the time-dependent local density approximation ionization threshold. *J Chem Phys* **108**, 4439-4449 (1998).
- 43 Amos, R. D. & Van Caillie, C. Geometric derivatives of excitation energies using SCF and DFT. *Chem Phys Lett* **308**, 249-255 (1999).
- 44 Amos, R. D. & Van Caillie, C. Geometric derivatives of density functional theory excitation energies using gradient-corrected functionals. *Chem Phys Lett* **317**, 159-164 (2000).
- 45 Furche, F. & Ahlrichs, R. Adiabatic time-dependent density functional methods for excited state properties. *J Chem Phys* **117**, 7433-7447 (2002).
- 46 Breneman, C. M. & Wiberg, K. B. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials - the Need for High Sampling Density in Formamide Conformational-Analysis. *J Comput Chem* **11**, 361-373 (1990).
- 47 G03 (Gaussian, Inc., Wallingford CT, 2004).
- 48 Tetenkin, V. L., Gulyaev, B. A., Seibert, M. & Rubin, A. B. Spectral Properties of Stabilized D1/D2/Cytochrome-B-559 Photosystem-II Reaction Center Complex - Effects of Triton X-100, the Redox State of Pheophytin, and Beta-Carotene. *Febs Letters* **250**, 459-463 (1989).
- 49 Braun, P., Greenberg, B. M. & Scherz, A. D1-D2-Cytochrome B559 Complex from the Aquatic Plant *Spirodela-Oligorrhiza* - Correlation between Complex Integrity, Spectroscopic Properties, Photochemical Activity, and Pigment Composition. *Biochemistry* **29**, 10376-10387 (1990).
- 50 Vankan, P. J. M. *et al.* Time-Resolved Spectroscopy at 10k of the Photosystem-II Reaction Center - Deconvolution of the Red Absorption-Band. *Biochimica Et Biophysica Acta* **1020**, 146-152 (1990).
- 51 Durrant, J. R. *et al.* A Multimer Model for P680, the Primary Electron-Donor of Photosystem-II. *P Natl Acad Sci USA* **92**, 4798-4802 (1995).

- 52 Raszewski, G., Diner, B. A., Schlodder, E. & Renger, T. Spectroscopic properties of reaction center pigments in photosystem II core complexes: Revision of the multimer model. *Biophys J* **95**, 105-119 (2008).
- 53 Renger, T. & Marcus, R. A. Photophysical properties of PS-2 reaction centers and a discrepancy in exciton relaxation times. *J Phys Chem B* **106**, 1809-1819 (2002).
- 54 Chang, J. C. Monopole effects on electronic excitation interactions between large molecules. I. Application to energy transfer in chlorophylls. *The Journal of Chemical Physics* **67**, 3901-3909 (1977).