Electronic Supplementary Information (ESI) for

The conformational phase space of the photoprotective switch in the major Light Harvesting Complex II

Vangelis Daskalakis,* Sotiris Papadatos, and Taxiarchis Stergiannakos

Department of Chemical Engineering, Cyprus University of Technology, 95 Eirinis Street, *E-mail: evangelos.daskalakis@cut.ac.cy

Computational Methods

Model setup -

In the present study, the crystal structure of the major light harvesting complex from spinach (LHCII, pdb:1rwt, polypeptide chains C, E and H)[1] has been used for the initial coordinates to build our models. Apart from the LHCII trimer as described in ref [2], we have additionally considered LHCII models that contain zeaxanthin in the place of violaxanthin, in order to take into account, the xanthophyll cycle that is activated under NPQ conditions.[3] Furthermore, LHCII-PsbS complex models are constructed based on the orientations proposed in ref [4], with the only difference that three instead of only two PsbS monomers interact with the LHCII trimer embedded in the thylakoid membrane mimetic. Models that contain different salts (KCl, $MgCl_2$ and $CaCl_2$)[5] at concentrations of either 120, or 500mM are also considered, in addition to a K^+, or Cl^- surplus to neutralize the protein charges in each sample. In conclusion, we have constructed LHCII models at the neutral or low lumenal pH state (**x2**), with either 120, or 500mM (**x2**) of either KCl, $MgCl_2$, or $CaCl_2$ salts (**x3**). The models at the low lumenal pH state are prepared also with zeaxanthin, in the place of violaxanthin (**x2**), in the presence or absence of three interacting PsbS monomers (**x2**). This sums-up to **30 different LHCII states – models** of between 221-260K atoms that take into account all the factors that could possibly affect the LHCII transition between the light harvesting and quenched state.[3,5–7] We provide **Table S1** for a detailed description of each model. Selected models are slso shown in **Fig. S1** for reference. All various models probed can give insight into the conformations of the trimer at the different states, or at least drive the LHCII to switch between states.

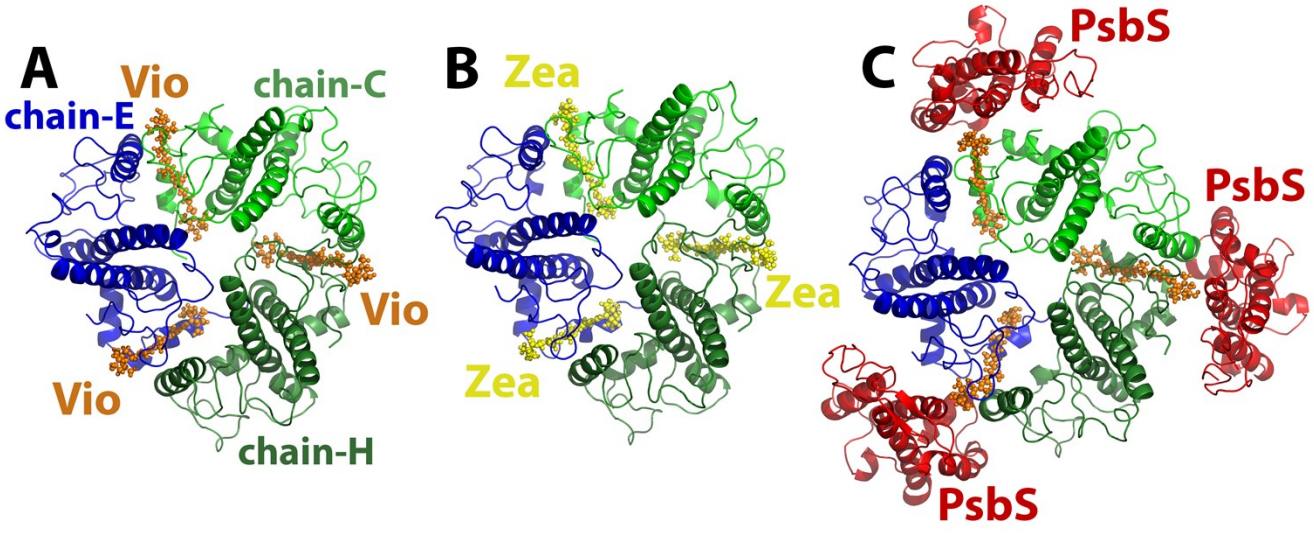

Figure S1 | Selected models probed in this study. **A.** The LHCII trimer (stromal view) with different colors for each polypeptide chain shown as cartoons (green for C, blue for E, and dark green for H). Violaxanthin (Vio) is shown in orange with spheres for the atoms. **B.** The LHCII trimer, depicted as in A, but with Zeaxanthin (Zea) in the place of Vio, shown in yellow spheres. **C.** The LHCII trimer, depicted as in A, with Vio and three PsbS monomers on its periphery, shown as red cartoons.

The LHCII polypeptides are prepared as described earlier and are embedded in a fully hydrated lipid bilayer membrane that mimics the thylakoid membrane.[2,8] In detail, two different initial states of the trimer are considered that differ at the simulated lumenal pH value (either low pH~5.5, or pH > 6).[9,10] Both states are however simulated with constant neutral pH (~7) at the stromal side.[11] To simulate the lumen acidification, associated with NPQ, key lumen exposed residues of the polypeptides (Glu-83,94,107,207 and Asp-111,211,215) are protonated. Titratable residues are treated deprotonated for the neutral lumenal (stromal) pH states. All His residues have been protonated only at the N_δ sites. We have to note that, under NPQ conditions, it is widely accepted that lumen exposed residues exert increased pKa values of around 5-5.5, in the presence of zeaxanthin or the PsbS protein,[11–13] compared to the lower values of 3-4 in the absence of such interactions. Thus, for the low lumenal pH states, we have additionally protonated residues at the lumenal side, due to such increases in their pka values identified in models.[10,14] We also note that the protonations at the protein lumenal sites for the low lumenal pH states, implicitly take into account also possible effects from LHCII aggregation.[15] Both experimental evidence[16–19] and the PDB2PQR (propka 3.0 method)[20] predictions validate our choice. (de-) Protonations are considered according to a-priori predicted pKa and do not refer to constant pH simulations. Constant pH simulations are still prone to instabilities over long simulation times, based on publicly available codes, i.e. in the time scale required to adequately sample the conformational changes within LHCII.[2,7]

Table S1 | Details for the different models probed in this study

#	LHCII	lumen pH	PsbS	Vio/Zea	KCl	$CaCl_2$	$MgCl_2$	Lipids	Waters
1	1	5.5	0	3/0	120mM	0	0	498	44636
2	1	5.5	0	3/0	500mM	0	0	498	44060
3	1	5.5	0	3/0	0	120mM	0	498	44603
4	1	5.5	0	3/0	0	500mM	0	498	44222
5	1	5.5	0	3/0	0	0	120mM	498	44603
6	1	5.5	0	3/0	0	0	500mM	498	44222
7	1	5.5	0	0/3	120mM	0	0	498	44636
8	1	5.5	0	0/3	500mM	0	0	498	44060
9	1	5.5	0	0/3	0	120mM	0	498	44603
10	1	5.5	0	0/3	0	500mM	0	498	44222
11	1	5.5	0	0/3	0	0	120mM	498	44603
12	1	5.5	0	0/3	0	0	500mM	498	44222
13	1	5.5	3	3/0	120mM	0	0	492	51228
14	1	5.5	3	3/0	500mM	0	0	492	50568
15	1	5.5	3	3/0	0	120mM	0	492	51208
16	1	5.5	3	3/0	0	500mM	0	492	50767
17	1	5.5	3	3/0	0	0	120mM	492	51208
18	1	5.5	3	3/0	0	0	500mM	492	50767
19	1	5.5	3	0/3	120mM	0	0	492	51228
20	1	5.5	3	0/3	500mM	0	0	492	50568
21	1	5.5	3	0/3	0	120mM	0	492	51208
22	1	5.5	3	0/3	0	500mM	0	492	50767
23	1	5.5	3	0/3	0	0	120mM	492	51208
24	1	5.5	3	0/3	0	0	500mM	492	50767
25	1	> 6.0	0	3/0	120mM	0	0	498	44636
26	1	> 6.0	0	3/0	500mM	0	0	498	44126

27	1	> 6.0	0	3/0	0	120mM	0	498	44623
28	1	> 6.0	0	3/0	0	500mM	0	498	44242
29	1	> 6.0	0	3/0	0	0	120mM	498	44623
30	1	> 6.0	0	3/0	0	0	500mM	498	44242

Classical Molecular Dynamics -

The all-atom models, as defined previously, are embedded in membrane patches of around 500 lipids (POPC, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine).[9,10,14,21] Around 45000 TIP4P/2005 water molecules[22] are used to hydrate each membrane surface.[23] The OPLS2005-AA[24–26] protein force field is used for the protein. The carotenoids and the native DGDG (digalactosyl diacyl glycerol), LHG (1,2-dipalmitoyl-phosphatidyl-glycerole) lipids are described based on OPLS-AA compatible parameters.[9,14,27] OPLS united-atom lipid parameters for POPC are available and adequately simulate the thylakoid membrane environment,[8,10,28] parametrized at 310K.[29,30] Each model x-y-z dimensions are around 14.0x14.0x11.5 nm^3. Based on published protocols[2] all models are relaxed and equilibrated with gradual removable of constraints on the protein backbone-heavy atoms. In a series of constant volume nVT, and constant pressure nPT ensembles the temperature increases from 100K to 310K,[2] prior to the production runs. For the production classical Molecular Dynamics (MD) simulations, Newton's equations of motion are integrated with a time step of 2.0 fs. The leap-frog integrator in GROMACS 2018 was employed.[31] The production runs have been performed in the constant pressure nPT ensemble, with semi-isotropic couplings in the xy membrane plane and in the z-direction (compressibility at $4.5x10^{-5}$). Van-der-Waals interactions were smoothly switched to zero between 1.0-1.2nm with the Verlet cut-off scheme. Electrostatic interactions were truncated at 1.2nm (short-range) and long-range contributions were computed within the PME approximation.[32,33] All hydrogen – heavy atom bond lengths were constrained employing the LINCS algorithm.[34] The v-rescale thermostat[35] is employed (310K, temperature coupling constant 0.5) and the *Parrinello-Rahman* barostat[36,37] (1 atm, pressure coupling constant 2.0) for two independent trajectories of 0.5μs per model. All parameters can be found in published works.[2,10] The total simulation time sums to **30μs** (30x2x0.5μs) for equilibrium dynamics.

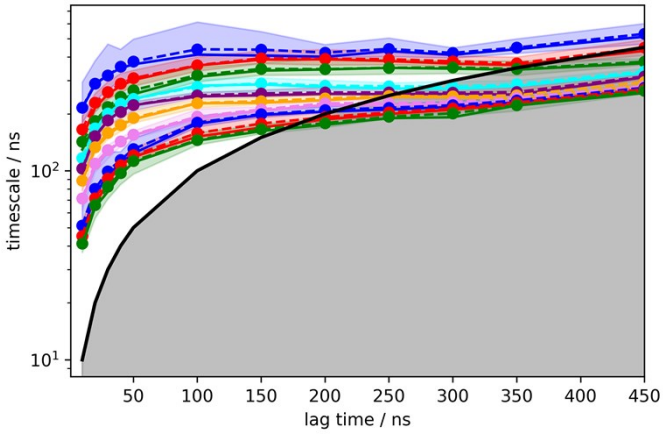

Figure S2 | The convergence behavior of the implied timescales associated with the slowest processes. The solid lines refer to the maximum likelihood result while the dashed lines show the ensemble mean computed with a Bayesian sampling procedure. The black line along with the gray shaded area indicates the timescale horizon below which the MSM cannot resolve processes.

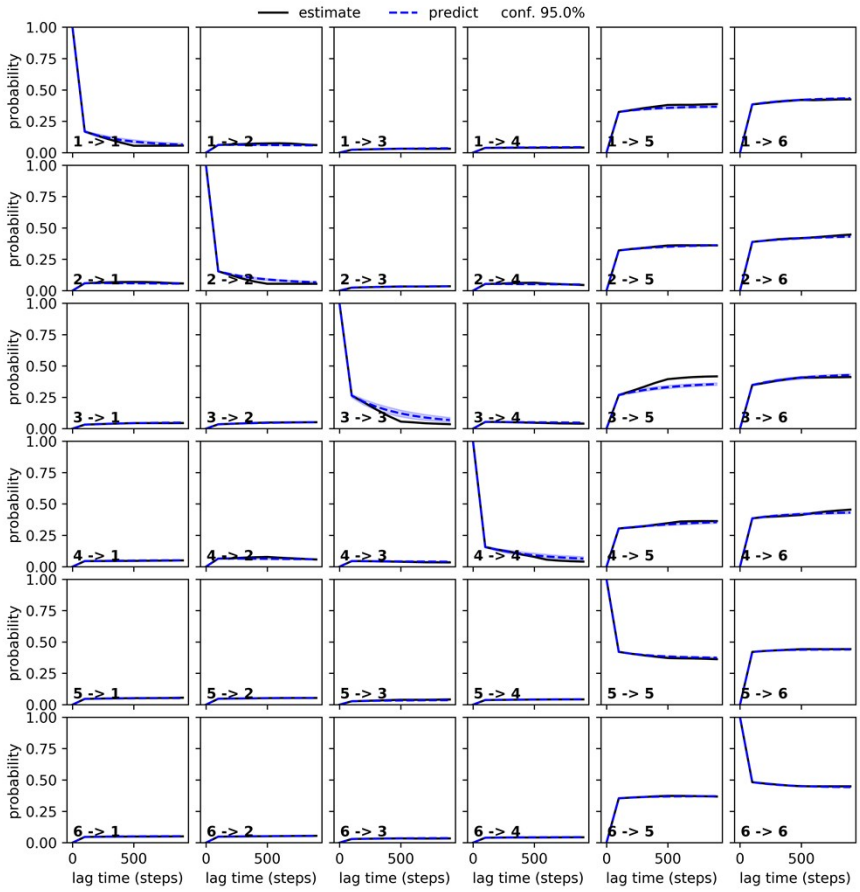

Figure S3 | A. The results of the Chapman-Kolmogorov (CK) computed using an MSM estimated with lag time of 100ns assuming four metastable states. Predictions from the model agree with higher lag time estimates within confidence intervals. The blue shaded areas indicate 95% confidence intervals computed with the aforementioned Bayesian sampling procedure.

Analysis based on Markov state model theory -

It has been shown that monomers within a LHCII trimer can sample different conformations.[2,38] To analyze the 30µs classical MD trajectories of the LHCII trimer, we extract only the LHCII protein scaffold, without protons, pigments, waters, lipids, or ions, and we additionally dismantle the trimer into monomer trajectories. We end up in total with **90µs** of 180 independent trajectories that describe the dynamics of the LHCII monomer protein scaffold. Trajectory frames were taken every 1ns. The frames in all the trajectories were structurally aligned on the same reference initial structure, based on Ca-fitting with VMD 1.9.3,[39] to assure consistency in the analysis. To adequately characterize the structural dynamics of LHCII we combine the all-atom MD simulations with Markov state model (MSM) theory.[40–42] This enables the extraction of long-time-scale individual monomer dynamics from rather short-time-scale MD trajectories of the trimer in different states. The application and accuracy of the powerful MSM theory has been demonstrated in many cases also by experiments that include protein−protein, or protein-drug binding kinetics, as well as protein folding rates and protein dynamics.[43–46]

MSMs of the LHCII monomer were constructed from the trajectories of 90µs total time using the PyEMMA package in Jupyter notebooks.[47] Only the torsional angles of the residues 55-87 (helix-B), 170-201 (helix-A), and 203-214 (helix-D) were selected as the initial input features for model construction. A lag-time of 100ns and 13 tICA eigenvectors were chosen based on the VAMP2-scores,[48] to identify a set of the slowest modes among all the initial input features.[49] These constitute a linearly optimal combination of input features which maximizes their kinetic variance. The slowest degrees of freedom are associated with the torsional angles of the following LHCII residues: Ala-86, Arg-87, Gln-131, Met-135, Glu-171, Ala-172, Phe-194, Val-200, Thr-201, Lys-203, Gly-204, Glu-207, and Leu-209 (*see also main manuscript*). The conformations of the system were projected on these

slowest modes as defined by the tICA method, then the trajectory frames were clustered into 400 cluster-centers (microstates) by k-means clustering, as implemented in PyEMMA.[47] The optimum number of microstates (six) was proposed based on the VAMP2-score.[48] Conformational changes of a system can be simulated as a Markov chain, if the transitions between the different conformations are sampled at long enough time intervals so that each transition is Markovian. This means that a transition from one conformation to another is independent of the previous transitions. The uncertainty bounds were computed using a Bayesian scheme.[50,51] We found that the slowest implied timescales converge quickly and are constant within a 95% confidence interval for lag times above 100ns (**Fig. S2**). The validation procedure is a standard approach in the MSM field., a lag time of 100 ns was selected for Bayesian model construction, and the resulting models were validated by the Chapman-Kolmogorov (CK) test (**Fig. S3**). Subsequently, the resulting MSMs were further coarse grained into a smaller number of six metastable states or microstates, using PCCA++ as implemented in PyEMMA.[47] Both the convergence of the implied timescales (**Fig. S2**), as well as the CK test (**Fig. S3**) confirm the validity and convergence of the MSM. The CK test indicates that predictions from the built MSM (blue dotted lines) agree well with MSMs estimated with longer lag times (black lines). Thus, the model can describe well the long-time-scale behavior of our system within error (blue shaded areas).

The first two tICA vectors (IC1/ IC2) are defined below, as a linear combination of cosine and sine functions of specific Phi-ϕ, or Psi-ψ torsional angles:

IC1 = (-0.2117669227321849)*COS(ALA86PHI)+(0.41775492319058655)*SIN(ALA86PHI)+(-0.559808389273112)*COS(ALA86PSI)+(0.6600856257871487)*SIN(ALA86PSI)+(0.164982040227197)*COS(ARG87PHI)+(-0.5151066102507099)*SIN(ARG87PHI)+(0.2648252866863971)*COS(ARG87PSI)+(0.1147806615069189)*SIN(ARG87PSI)+(-0.22315148525916734)*COS(GLN131PHI)+(-0.09282053052290833)*SIN(GLN131PHI)+(0.28037306486372404)*COS(GLN131PSI)+(0.1676060444049281)*SIN(GLN131PSI)+(-0.4022197155737235)*COS(MET135PHI)+(0.34188752103723796)*SIN(MET135PHI)+(0.41407705908683995)*COS(MET135PSI)+(0.34631941405074795)*SIN(MET135PSI)+(-0.039324855950943725)*COS(GLU171PHI)+(-0.02199693448617469)*SIN(GLU171PHI)+(-0.05709611160009663)*COS(GLU171PSI)+(-0.03689831564314275)*SIN(GLU171PSI)+(0.07242938976235862)*COS(ALA172PHI)+(-0.1500884151457316)*SIN(ALA172PHI)+(-0.0289530278213439)*COS(ALA172PSI)+(-0.02684987505251958)*SIN(ALA172PSI)+(0.12299922568844492)*COS(PHE194PHI)+(0.15886613611446962)*SIN(PHE194PHI)+(0.19275945452977095)*COS(PHE194PSI)+(0.158966653923455)*SIN(PHE194PSI)+(-0.12175076905962252)*COS(VAL200PHI)+(0.14973845208107478)*SIN(VAL200PHI)+(-0.1884368422742053)*COS(VAL200PSI)+(-0.07397544132509451)*SIN(VAL200PSI)+(0.12343743107747304)*COS(THR201PHI)+(0.012277279682527883)*SIN(THR201PHI)+(0.185113415121421)*COS(THR201PSI)+(0.198216195340222)*SIN(THR201PSI)+(0.22159484746627484)*COS(LYS203PHI)+(0.22105176136826485)*SIN(LYS203PHI)+(-0.29625153277577837)*COS(LYS203PSI)+(-0.2724759289390821)*SIN(LYS203PSI)+(0.23200179060373283)*COS(GLY204PHI)+(-0.029544178252826053)*SIN(GLY204PHI)+(-0.12936193807110674)*COS(GLY204PSI)+(-0.09898798823634636)*SIN(GLY204PSI)+(-0.08117864947660554)*COS(GLU207PHI)+(0.07854699696147688)*SIN(GLU207PHI)+(0.13339029827860382)*COS(GLU207PSI)+(0.14640844067087522)*SIN(GLU207PSI)+(-0.16352157687317528)*COS(LEU209PHI)+(0.12290456681009244)*SIN(LEU209PHI)+(-0.07914428035767718)*COS(LEU209PSI)+(-0.03890122183795599)*SIN(LEU209PSI)

IC2 = (0.04557611007995849)*COS(ALA86PHI)+(0.003768733795092695)*SIN(ALA86PHI)+(-0.026975768153629454)*COS(ALA86PSI)+(0.03353158975644035)*SIN(ALA86PSI)+(0.04268398026545617)*COS(ARG87PHI)+(0.05497614555686673)*SIN(ARG87PHI)+(0.023172454946612533)*COS(ARG87PSI)+(0.00562463321158085)*SIN(ARG87PSI)+(0.07600934924882428)*COS(GLN131PHI)+(-0.015296930827965854)*SIN(GLN131PHI)+(-0.043744093672826244)*COS(GLN131PSI)+(-0.04543630550863082)*SIN(GLN131PSI)+(-0.04748475065912107)*COS(MET135PHI)+(0.005865163191593903)*SIN(MET135PHI)+(0.05368144780864727)*COS(MET135PSI)+(0.009502760483985379)*SIN(MET135PSI)+(-

```
0.18722246351319005)*COS(GLU171PHI)+(0.07720415066560098)*SIN(GLU171PHI)+(0.0930278907824
3294)*COS(GLU171PSI)+(0.02676099063183022)*SIN(GLU171PSI)+(0.051412353208551156)*COS(ALA1
72PHI)+(-0.16571192525601933)*SIN(ALA172PHI)+(0.020336802973207228)*COS(ALA172PSI)+(-
0.030466831279901855)*SIN(ALA172PSI)+(-0.03305915032821139)*COS(PHE194PHI)+(-
0.03076189978345216)*SIN(PHE194PHI)+(-0.041955643747273126)*COS(PHE194PSI)+(-
0.0064413819222125645)*SIN(PHE194PSI)+(-
0.1909086241057209)*COS(VAL200PHI)+(0.168379532095841 1)*SIN(VAL200PHI)+(0.610962829993734
5)*COS(VAL200PSI)+(0.6417047985177151)*SIN(VAL200PSI)+(-
0.05168650154896537)*COS(THR201PHI)+(0.4917303283000218)*SIN(THR201PHI)+(-
0.3192514788370159)*COS(THR201PSI)+(-0.3308697132522881)*SIN(THR201PSI)+(-
0.05191430405990459)*COS(LYS203PHI)+(0.07277258342728306)*SIN(LYS203PHI)+(0.0183623841522
9812)*COS(LYS203PSI)+(0.06534971791729183)*SIN(LYS203PSI)+(0.2504715747238633)*COS(GLY204
PHI)+(0.07569615532197012)*SIN(GLY204PHI)+(-0.1790841008466512)*COS(GLY204PSI)+(-
0.24736367974564935)*SIN(GLY204PSI)+(0.14534391708534922)*COS(GLU207PHI)+(-
0.0720655449133264)*SIN(GLU207PHI)+(-0.13357971915246644)*COS(GLU207PSI)+(-
0.1405664686455033)*SIN(GLU207PSI)+(-
0.12245441992008191)*COS(LEU209PHI)+(0.21385034338433614)*SIN(LEU209PHI)+(-
0.05358567741746825)*COS(LEU209PSI)+(-0.04633517640435567)*SIN(LEU209PSI)
```

Enhanced sampling (Replica Exchange, Multiple Walkers Metadynamics) -

To enhance the conformational sampling on the LHCII trimer, we employed the parallel tempering metadynamics in the well-tempered ensemble (PTmetaD-WTE) method.[52–55] Six replicas per sample were run at 310, 320, 330, 341, 352, and 363K, in which only the potential energy (PE) was initially biased (bias factor 120) to achieve large fluctuations in PE and replica overlaps. Replicas were allowed to exchange every 1000 steps for 0.5μs each, which gave an exchange probability between 20-30% in the WTE. The obtained bias was saved and used for the subsequent PTmetaD production runs for another 0.5μs per sample/ replica. Six replicas were again considered at the same temperatures. An exchange was attempted every 1000 steps, that gave an exchange probability between replicas at around 30%, consistent with the large sample sizes. The Collective Variables (CVs) chosen for the PTmetaD runs are described in the main manuscript and are based on the Markov state models and the first two tICA vectors presented above. The PtmetaD runs were performed for the simplest LHCII trimer model structures at neutral and low lumenal pH (without zeaxanthin, or PsbS interactions) and at 120mM KCl ionic strength. A combination of the GROMACS 2018/ PLUMED 2.5[56] engines was employed. A bias factor of 25 at the well-tempered ensemble, along with Gaussians of 1.2 kJ/mol initial height, and sigma values (width) of 0.25 in the CV space, deposited every 2 ps, was employed. The grid space for both CVs is defined between -2 to 2 at a resolution of 0.05. We note that the whole trimer was considered for all enhanced-sampling simulations at the neutral and low lumenal pH states, however four different PTmetaD-WTE setups were employed: three PTmetaD-WTE runs biased the CVs individually for each monomer within the trimer, while the fourth PTmetaD-WTE run biased the CVs in all monomers at the same time. To further exploit the metadynamics capabilities, we performed the Multiple Walkers (MW) variant of the method,[57] with the exact same bias factor, initial height, sigma values and grid space defined before for the PTmetaD-WTE runs. For the MW metadynamics variant, three replicas were run at 310K, and referred each to the biasing of the tICA CVs only for one monomer within the same trimer. Thus, we have exploited the LHCII structure as a trimer, to assign each monomer a different 'walker'. Walkers run as independent simulations sharing the biasing potential. The walkers read with a given frequency the Gaussian kernels deposited by the others and add them to their own. To sum-up, we have performed 8 PTmetaD-WTE runs with 6 replicas each, with a total simulation time of 1μs per replica, including the PE biasing. In addition, two MW metadynamics runs for the LHCII trimer at the neutral and low lumenal pH states, were performed with three replicas per run, for 0.5μs sampling time per replica. These total in **51μs** enhanced sampling time. We have obtained, on the average, around 20% replica exchange rates.

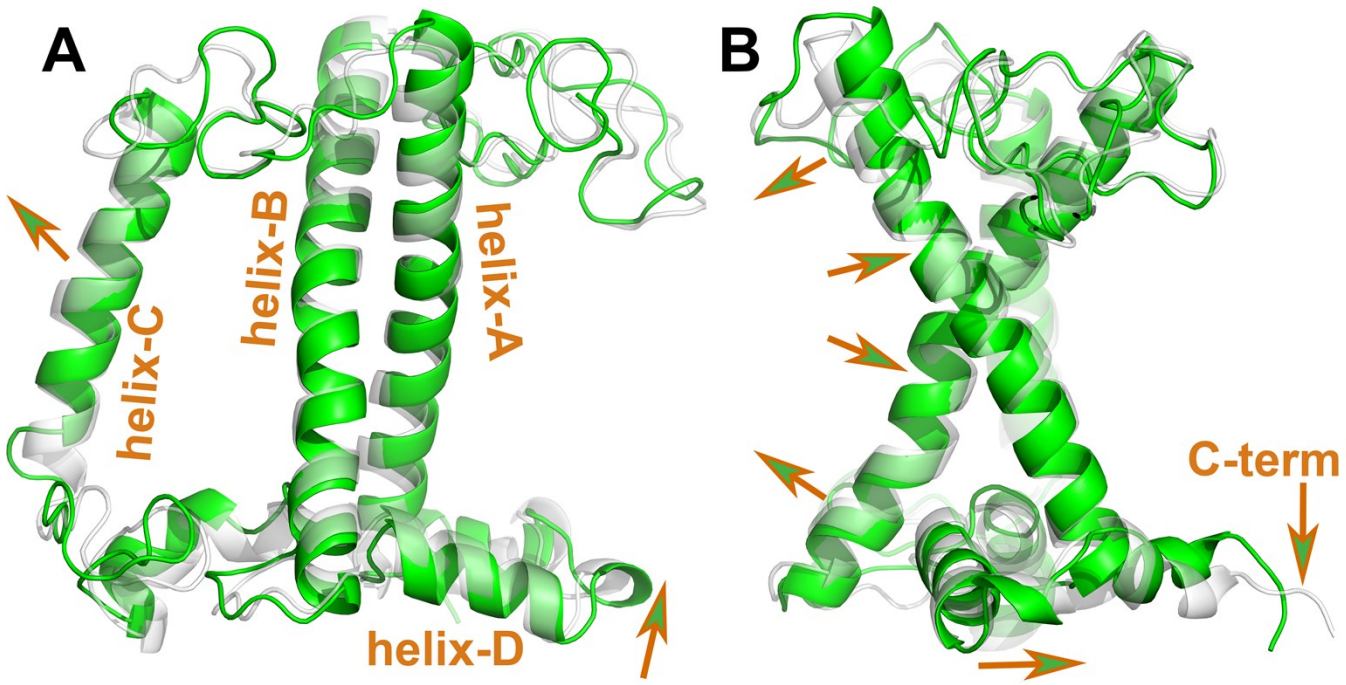

Figure S4 | A-B. Structures of LHCII (chain C) extracted at low ΔpH and Chl-a 612/ Lut-620 excitonic coupling value at ~0 cm^{-1} (solid green cartoon) and at enhanced ΔpH and Chl-a 612/ Lut-620 excitonic coupling value at ~16.2 cm^{-1} (transparent cartoon). Arrows are also drawn for reference, for the transition.

Switching between LHCII trimer states – steered MD

Due to the heterogeneity of the dynamics sampled per monomer, within the same LHCII trimer at the neutral pH state,[38] we have employed the steered MD method[58] to drive all polypeptide conformations within the same trimer towards the different minima (**Fig. 3A-B** *in the main manuscript*). For this, we have chosen to switch to the Amber03 force field[59] to describe the polypeptides, employing the potential setup of refs [38,60]. In detail, the pigments were described by (a) an *ad hoc* force field (ff) developed by Prandi et al. and fine-tuned for carotenoids,[61] (b) a set of parameters reported for Chls-a,[62] along with modifications,[38,60,63] was used for Chls-a/ b. The thylakoid membrane has been parametrized and modelled before in the Amber03 ff.[10,64] It contains a variety of lipids: monogalactosyldiacylglycerol – MGDG (44%), digalactosyldiacylglycerol – DGDG (25%), sulfoquinovosyldiacylglycerol – SQDG (25%), phosphatidylglycerol – PG (6%). Both the established *ad hoc* ff for the carotenoids, and the native environment (thylakoid lipids) for the LHCII trimer, render this setup ideal to probe the associated switch between light harvesting and quenched states. The Amber ff could not be employed *a priori* also for the classical or metadynamics runs, as it has a known weakness by promoting excess aggregation and salt crystallization even at low salt concentrations. For this, we had to turn to OPLS-AA based models of varying salt contents-species, and then switch to the Amber ff for the *ad hoc* description of carotenoids and the native thylakoid environment. The steered MD simulations were initiated at the crystal structure conformation of the trimer, prepared based on the same protocol as described earlier for the OPLS-based systems at neutral and low lumenal pH states at 0.12mM KCl (without Zeaxanthin, or PsbS interactions). The steered MD runs included 50ns of equilibration, 150ns of actual steering of both CV1, and CV2 with a force constant of 1000 kJ/mol/nm^2 towards multiple combinations of CV values. The combinations refer to all minima in Figs. 3A-B of the main manuscript (the combinations of CV1/ CV2 targets are: -1.55/-0.30, -0.75/-0.50, -0.75/0.25, -0.5/-0.5, 0.0/0.0, 0.00/-1.50, 0.25/-0.25, +0.70/-0.30, +0.75/0.0, +1.55/-0.30). A subsequent 110ns equilibration followed. The steered MD runs were performed with the same parameters as reported in the *Classical Molecular Dynamics* section in the methods. The TrESP (transition charges from electrostatic potential)[65] calculations reported in the main manuscript refer to equilibrium trajectories at the end of the steered MD runs, with the Amber03 ff only. All conformations that were derived from the steered MD runs, were first clustered to the

three main minima shown in **Figs. A-B** of the main manuscript, prior to the calculations of the TrESP couplings in each cluster of structures (C1, C2 and C3). In **Fig. S4** we depict the structures of LHCII (chain C) at (i) low ΔpH and Chl-a 612/ Lut-620 excitonic coupling value at ~0 cm^{-1} (solid green cartoon) and (ii) enhanced ΔpH and Chl-a 612/ Lut-620 excitonic coupling value at ~16.2 cm^{-1} (transparent cartoon). Arrows are also drawn for reference, for the transition from (i) to (ii). For the coupling values and their distribution, please refer also to the main manuscript and **Fig. 3C**.

NOTES AND REFERENCES

1 Z. Liu, H. Yan, K. Wang, T. Kuang, J. Zhang, L. Gui, X. An and W. Chang, *Nature*, 2004, **428**, 287–292.
2 V. Daskalakis, S. Maity, C. L. Hart, T. Stergiannakos, C. D. P. Duffy and U. Kleinekathöfer, *J. Phys. Chem. B*, 2019, **123**, 9609–9615.
3 J. Sacharz, V. Giovagnetti, P. Ungerer, G. Mastroianni and A. V Ruban, *Nat Plants*.
4 V. Daskalakis, S. Papadatos and U. Kleinekathöfer, *Biochim. Biophys. Acta - Biomembr.*, 2019, **1861**, 183059.
5 G. Radek Kaňa, *Front. Plant Sci.*
6 A. V Ruban and P. Horton, *Plant Physiol.*, 1999, **119**, 531–542.
7 N. Liguori, S. R. R. Campos, A. M. Baptista and R. Croce, *J. Phys. Chem. Lett.*, 2019, **10**, 1737–1742.
8 N. Liguori, X. Periole, S. J. Marrink and R. Croce, *Sci. Rep.*
9 S. Papadatos, A. C. Charalambous and V. Daskalakis, *Sci. Rep.*, , DOI:10.1038/s41598-017-02892-w.
10 V. Daskalakis, *Phys. Chem. Chem. Phys.*, , DOI:10.1039/c8cp01226a.
11 A. V Ruban, M. P. Johnson and C. D. P. Duffy, *Biochim. Biophys. Acta (BBA)-Bioenergetics*, 2012, **1817**, 167–181.
12 A. V Ruban, *Plant Physiol.*, 2016, **170**, 1903–1916.
13 G. Noctor, D. Rees, A. Young and P. Horton, *Biochim. Biophys. Acta - Bioenerg.*, 1991, **1057**, 320–330.
14 V. Daskalakis and S. Papadatos, *Biophys. J.*, **113**, 2364–2372.
15 A. V Ruban, *FEBS Lett.*, 2018, **592**, 3030–3039.
16 C. Liu, Y. Rao, L. Zhang and C. Yang, *J. Biochem.*, 2014, mvu028.
17 R. G. Walters, A. V Ruban and P. Horton, *Proc. Natl. Acad. Sci.*, 1996, **93**, 14204–14209.
18 X.-P. Li, A. M. Gilmore, S. Caffarri, R. Bassi, T. Golan, D. Kramer and K. K. Niyogi, *J. Biol. Chem.*, 2004, **279**, 22866–22874.
19 E. Bergantino, A. Segalla, A. Brunetta, E. Teardo, F. Rigoni, G. M. Giacometti and I. Szabò, *Proc. Natl. Acad. Sci.*, 2003, **100**, 15265–15270.
20 T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, *Nucleic Acids Res.*, 2004, **32**, W665–W667.
21 N. E. Ioannidis, S. Papadatos and V. Daskalakis, *Biochim. Biophys. Acta - Bioenerg.*, , DOI:10.1016/j.bbabio.2016.07.005.
22 D. C. Elton and M.-V. Fernández-Serra, *J. Chem. Phys.*, 2014, **140**, 124504.
23 O. Fisette, C. Päslack, R. Barnes, J. M. Isas, R. Langen, M. Heyden, S. Han and L. V Schäfer, *J. Am. Chem. Soc.*, 2016, **138**, 11526–11535.
24 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.
25 M. L. P. Price, D. Ostrovsky and W. L. Jorgensen, *J. Comput. Chem.*, 2001, **22**, 1340–1352.
26 K. Karki and D. Roccatano, *J. Chem. Theory Comput.*, 2011, **7**, 1131–1140.
27 L. S. Dodda, I. Cabeza de Vaca, J. Tirado-Rives and W. L. Jorgensen, *Nucleic Acids Res.*, 2017, **45**, W331–W336.
28 J. I. Ogren, A. L. Tong, S. C. Gordon, A. Chenu, Y. Lu, R. Blankenship, J. Cao and G. Schlau-Cohen, *Chem. Sci.*, , DOI:10.1039/C7SC04814A.
29 J. P. Ulmschneider and M. B. Ulmschneider, *J. Chem. Theory Comput.*, 2009, **5**, 1803–1813.
30 A. Laganowsky, E. Reading, T. M. Allison, M. B. Ulmschneider, M. T. Degiacomi, A. J. Baldwin and C.

V Robinson, *Nature*, 2014, **510**, 172–175.
31 H. J. C. Berendsen, D. van der Spoel and R. van Drunen, *Comput. Phys. Commun.*, 1995, **91**, 43–56.
32 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
33 I.-C. Yeh and M. L. Berkowitz, *J. Chem. Phys.*, 1999, **111**, 3155–3162.
34 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.
35 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 14101.
36 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
37 S. Nosé and M. L. Klein, *Mol. Phys.*, 1983, **50**, 1055–1076.
38 L. Cupellini, D. Calvani, D. Jacquemin and B. Mennucci, *Nat. Commun.*, 2020, **11**, 662.
39 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.
40 V. S. Pande, K. Beauchamp and G. R. Bowman, *Methods*, 2010, **52**, 99–105.
41 J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte and F. Noé, *J. Chem. Phys.*, 2011, **134**, 174105.
42 J. D. Chodera and F. Noé, *Curr. Opin. Struct. Biol.*, 2014, **25**, 135–144.
43 N. Plattner and F. Noé, *Nat. Commun.*, 2015, **6**, 7653.
44 N. Plattner, S. Doerr, G. De Fabritiis and F. Noé, *Nat Chem*, , DOI:10.1038/nchem.2785http://www.nature.com/nchem/journal/vaop/ncurrent/abs/nchem.2785.html#supplementary-information.
45 V. A. Voelz, G. R. Bowman, K. Beauchamp and V. S. Pande, *J. Am. Chem. Soc.*, 2010, **132**, 1526–1528.
46 J. D. Durrant, S. E. Kochanek, L. Casalino, P. U. Ieong, A. C. Dommer and R. E. Amaro, *ACS Cent. Sci.*, 2020, **6**, 189–196.
47 M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz and F. Noé, *J. Chem. Theory Comput.*, 2015, **11**, 5525–5542.
48 H. Wu and F. Noé, *J. Nonlinear Sci.*, 2020, **30**, 23–66.
49 M. M. Sultan and V. S. Pande, *J. Chem. Theory Comput.*, 2017, **13**, 2440–2447.
50 F. Noé, *J. Chem. Phys.*, 2008, **128**, 244103.
51 B. Trendelkamp-Schroer, H. Wu, F. Paul and F. Noé, *J. Chem. Phys.*, 2015, **143**, 174101.
52 L. Sutto, S. Marsili and F. L. Gervasio, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2012, **2**, 771–779.
53 A. Barducci, G. Bussi and M. Parrinello, *Phys. Rev. Lett.*, 2008, **100**, 20603.
54 M. Bonomi and M. Parrinello, *Phys. Rev. Lett.*, 2010, **104**, 190601.
55 G. Bussi, F. L. Gervasio, A. Laio and M. Parrinello, *J. Am. Chem. Soc.*, 2006, **128**, 13435–13441.
56 G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput. Phys. Commun.*, 2014, **185**, 604–613.
57 P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti and M. Parrinello, *J. Phys. Chem. B*, 2006, **110**, 3533–3539.
58 C. He, G. Z. Genchev, H. Lu and H. Li, *J. Am. Chem. Soc.*, 2012, **134**, 10428–10435.
59 Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, *J Comput Chem*, 2003, **24**, 1999–2012.
60 V. Balevičius, K. F. Fox, W. P. Bricker, S. Jurinovich, I. G. Prandi, B. Mennucci and C. D. P. Duffy, *Sci. Rep.*, 2017, **7**, 13956.
61 I. G. Prandi, L. Viani, O. Andreussi and B. Mennucci, *J. Comput. Chem.*, 2016, **37**, 981–991.
62 M. Ceccarelli, P. Procacci and M. Marchi, *J. Comput. Chem.*, 2003, **24**, 129–142.
63 L. Zhang, D. A. Silva, Y. Yan and X. Huang, *J Comput Chem*, 2012, **33**, 1969–1980.
64 M. Retegan and D. A. Pantazis, *J. Am. Chem. Soc.*, 2017, **139**, 14340–14343.
65 M. E. Madjet, A. Abdurahman and T. Renger, *J. Phys. Chem. B*, 2006, **110**, 17268–17281.