Supplemental Information for

Structural Basis for Allosteric Regulation in the Major Antenna Trimer of Photosystem II

Vangelis Daskalakis*1, Sayan Maity2, Cameron Lewis Hart3, Taxiarchis Stergiannakos1, Christopher D.P. Duffy3, and Ulrich Kleinekathöfer2

1Department of Environmental Science and Technology, Cyprus University of Technology, 30 Archbishop Kyprianou Str., 3603, Limassol, Cyprus, 2Jacobs University Bremen, Department of Physics & Earth Sciences, Campus Ring 1, 28759 Bremen, Germany, 3School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom.

1. Classical Molecular Dynamics Methods

The samples

The crystal structure of the major light harvesting complex from spinach (LHClII, pdb:1rwt, polypeptide chains C, E and H)1 has been used for the initial coordinates to build our models. Each polypeptide is prepared as described earlier.2-3 In detail, NPQ conditions, i.e., the lumen acidification or the cross-membrane pH gradient, are simulated by protonating key lumen exposed residues (Glu-E, Asp-D) of the protein (E83,94,107,207 and D111,211,215). These protonations are based on experimental4-7 and PDB2PQR (propka 3.0 method)8 predictions to simulate LHClII with constant neutral pH (~7) at the stromal side, but at (a) enlarged ΔpH with low pH~5.5 at the lumenal side, or (b) low ΔpH with neutral pH > 6 at the lumenal side.2-3 All His residues have been protonated only at the Nδ sites. Under NPQ conditions, it has been proposed that the pKa values of lumen exposed residues are increased, e.g. by the interaction with zeaxanthin or the PsbS protein, from values around 3-4, to values around 5-5.5.9-11 Thus, for the enlarged ΔpH, or low lumenal pH state, we have additionally protonated residues at the lumenal side, due to such increases in their pka values in previous models.3,12 These titratable residues are treated deprotonated for the neutral lumenal pH or low ΔpH samples. We note that these states are approximated by protonating key residues according to their pKa and do not refer to constant pH simulations (see methods). 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 LHClII.13 The complexes are embedded in a fully hydrated lipid bilayer membrane that mimics the native thylakoid membrane.14 All crystallographic water molecules are retained within the trimer structures.1 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).2,12,15-16 Around 45000 TIP4P/2005 water molecules17 are used to hydrate each membrane surface.18 Ion (K+, Cl−) concentration at 120mM is added to mimic the physiological thylakoid salt concentration induced by ion fluxes,19 in addition to a K+ or Cl− surplus to neutralize the protein charges in each sample. The OPLS2005-AA20-22 protein force field is used for the protein. Moreover, the carotenoids and the native DGDG (digalactosyl diacyl glycerol), LHG (1,2-
dipalmitoyl-phosphatidyl-glycerole) lipids are described based on OPLS-AA compatible parameters.\textsuperscript{2, 12, 23} OPLS united-atom lipid parameters for POPC are available and adequately simulate the thylakoid membrane environment.\textsuperscript{3, 14, 24} They have been parametrized at 310K to probe the membrane dynamics, along with the TIP4P/2005 water model.\textsuperscript{25-26} The above setup yields systems with x-y-z dimensions of around 13.5x13.5x11.5 nm\textsuperscript{3}.

The equilibration-relaxation for the all-atom systems is employed based on published protocols\textsuperscript{3} with slight variations. This contains a steepest descend energy minimization with a tolerance of 0.5 kJoule/mol for 1000 steps, and a sequence of isothermal (nVT), isothermal-isobaric (nPT) runs with the gradual relaxation of the constraints on protein heavy atoms (from 10\textsuperscript{4} in steps-1-2 to 10\textsuperscript{3} kJ/mol/nm\textsuperscript{2} in step-4) and Ca atoms (from 10\textsuperscript{3} in step-5, to 10\textsuperscript{2} in step-6, 10 in step-7, 1 in step-8 and 0 kJ/mol/nm\textsuperscript{2} in step-9) for around 30ns with a time step of 1.0 fs (steps 1-4) and 2.0 fs (steps 5-9). In detail: (step-1) Constant density and temperature (nVT) Brownian dynamics at 100K for 50ps that employ the Berendsen thermostat with a temperature coupling constant at 1.0 fs.\textsuperscript{27} (steps 2-3) Two short constant density (nVT) and constant pressure (nPT) runs for 100ps each. For the latter, the weak coupling Berendsen thermostat and barostat\textsuperscript{27} at 100K were employed with time coupling constants of 0.1ps (temperature) and semi-isotropic 50.0ps coupling (pressure) with a compressibility of 4.6x10\textsuperscript{-5} (x-y dimension) and 4.5x10\textsuperscript{-5} (z dimension). (step-4) Heating from 100K to 250K in a constant density ensemble (nVT) for 3ns employing a v-rescale thermostat\textsuperscript{28} with a time coupling constant of 0.1ps. (step-5) Heating from 250K to 310K in a constant pressure ensemble (nPT) for 2ns, employing the v-rescale thermostat\textsuperscript{28} and Berendsen barostat\textsuperscript{27}, with time coupling constants of 0.1ps for the temperature and 2.0ps for the pressure. For this we remove also all but the Ca-atom protein position restraints. (step-6) Equilibration at 310K (0.1ps temperature coupling constant) for 5ns (nPT, 1atm, 2.0ps coupling constant for pressure). (steps 7-8) Equilibration at 310K (0.5ps temperature coupling constant) for 5ns (nPT, 1atm, 2.0ps coupling constant for pressure). (step-9) Equilibration at 310K (0.5ps temperature coupling constant) for 10ns (nPT, 1atm, 2.0ps coupling constant for pressure). The barostats – thermostats employed for steps 6-9 were the same as in the production trajectories that follow.

**Classical Molecular Dynamics**

For the production all-atom classical Molecular Dynamics (MD) simulations, Newton’s equations of motion are integrated with a time step of 2.0 fs. All production simulations were run with the leap-frog integrator in GROMACS 5.1.5.\textsuperscript{29} They 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\textsuperscript{-5}). The v-rescale thermostat\textsuperscript{28} is employed (310K, temperature coupling constant 0.5) and the Parrinello-Rahman barostat\textsuperscript{30-31} (1 atm, pressure coupling constant 2.0) for 1.5\micro s trajectories per sample. Details for the parameters can be found in Ref. 3. The first 1.2\micro s are considered further equilibration from each trajectory per sample and are thus disregarded for the analysis. The total simulation time sums to 3.0\micro s for systems that contain ~221k atoms. 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.\textsuperscript{32} All hydrogen – heavy atom bond lengths were constrained employing the LINCS algorithm.\textsuperscript{33}

**2. Enhanced and biased sampling**

The initial LHCII trimer structures for both WT-metaD runs at (a) neutral lumenal pH, or low ΔpH, and (b) low lumenal pH, or enlarged ΔpH, were produced as the central structures of the most populated
clusters along the respective previous classical MD trajectories. A combination of the GROMACS 5.1.5/PLUMED 2.4 engines was employed. Collective Variables (CVs) within the WT-metaD method were defined. As reported in the main manuscript, we have identified an allosteric pathway (P1) that responds to pH changes at the lumenal side of LHCII and involves parts of helices A/ B. Its characteristic P1 dihedral angle (A2-A1–B1-B2), as described and shown in Fig. 1D in the main manuscript, can be defined as the first CV (CV1). This dihedral refers to the interhelical crossing angle between helices A and B of LHCII. We also consider the proposed plasticity of helix-D under NPQ conditions, that can be described by the φ torsional angle of Gly-204 (G204) at the start of helix-D. The G204-φ torsional has been defined as the second CV (CV2). Here the residue numbering is based on that of the major LHCII crystal structure from spinach. A bias factor of 15 at the well-tempered ensemble, along with Gaussians of 1.2 kJ/mol initial height, and sigma values (width) of 0.35 deg in the CV space, deposited every 2 ps, has been employed. The grid space for both CVs is defined between -π to π at a resolution of 0.1. Metadynamics simulations in their well-tempered variant were run for 200ns at each LHCII state, biasing the CV1-CV2 only within the chain C monomer of the trimer. The whole trimer was, however, considered for the simulation. The simulation time was proven adequate as the FES profile does not change past the time of ~60ns with relative minima differences of less than 1 kJ/mol. Furthermore, the enhanced sampling techniques provided an adequately sampled CV space.

The single linkage method was employed to cluster the structures within a 0.1nm cutoff for the Ca atoms of the three monomers within the trimer metadynamics. Central structures of the most populated clusters around selected FES minima are extracted. These are used as starting structures to initiate additional equilibrium dynamics at neutral and low lumenal pH by classical MD trajectories (0.5 μs each). The production runs were performed with the same parameters as reported in the Classical Molecular Dynamics section in the methods. Only the last 0.3μs out of each 0.5μs classical MD trajectory were used for the analysis.

3. The Graph theory framework

In order to probe allosteric pathways that connect the lumen (helix-D, C-terminus) to stroma (N-terminus) sites, and obtain the fine details of the divergence in the LHCII fluctuations under light-harvesting and NPQ conditions, we employed a graph-theoretical approach. This approach evaluates the correlations between the motion of adjacent residues within the protein. Within this theory, protein residues form “nodes” in a graph and “edges” represent contacts or interactions, e.g. hydrogen bonds, salt bridges and hydrophobic tethers, between nodes over an extended network. The correlations (as pathways) usually mean that perturbations to one residue can create long-range allosteric effects by their propagation through the network. Protein networks can enhance the identification of allosteric sites and identify the pathways through which excitations propagate across the protein. For more details we refer the interested reader to reviews or successful applications of the method on protein dynamics. The intrinsic changes within the major LHCII can thus be revealed under the principle of allosteric connections from the lumenal to the stromal sites or vice versa. We first employ RING to convert the protein crystal structure to a graph of nodes and edges and then Gephi for the analysis. The Thr-57/201 residues of the LHCII were chosen as end-points to define the shortest-pathway within the graph theory approach for LHCII. They were chosen as the residues closest to the lumen (helix-D, C-terminus)/stroma (N-terminus). Moreover, they are the furthest apart in the sequence within well-defined secondary structure features to belong to complete pathways within the graph theory.
4. TrESP coupling calculations

The excitonic couplings along the MD trajectories for the Chl-a/ Lut1 pairs were computed based on the electrostatic TrESP approach given by Renger and co-workers\textsuperscript{41-42}.

\[ V_{DA} = \frac{f}{4\pi\varepsilon_0} \sum_{i,j=1}^{D,A} \frac{q_i^T q_j^T}{|r_i - r_j|} \]

where, \( q_i^T \) and \( q_j^T \) are the electrostatic transition charges localized on the respective atoms in the donor and acceptor pigments. The dielectric effect of the protein environment is considered by a screening factor \( f \). In the present work, the transition charges for Chl-a and Lut1 were taken from the recently reported \textit{ab-initio} (RASSCF) calculations\textsuperscript{43} and we assume that they are constant along the MD trajectory. No dielectric effect of the protein environment was not taken into account for any of the couplings since we were mainly interested in relative comparisons.

5. Calculation of LHCII excitation lifetimes

The excitonic couplings within the Chl-a 612/ Lut1 dimer for each trajectory were used to compute the Förster (Golden Rule) rate constants for both forward (Chl-a 612 → Lut1) and backward (Lut1 → Chl-a 612) exciton hopping. As argued in previous work,\textsuperscript{44} the assumption of the incoherent regime is reasonable given the inherently weak (10-20 cm\(^{-1}\)) nature of couplings between the Chl a \( Q_y \) and xanthophyll S\(_1\) transitions. The rate constant for an excitation hopping from donor \( m \) to acceptor \( n \) is given by

\[ k_{mn} = 2|V_{mn}|^2 \text{Re} \int_0^\infty F_m^*(t)A_n(t)dt. \]

Here \( V_{mn} \) is the excitonic coupling and

\[ A_n(t) = e^{-i\omega_n t - \frac{2\lambda_m t}{\hbar}} \]

are the absorption (acceptor) and fluorescence (donor) response functions, respectively. \( \hbar \omega_m/n \) represents the 0-0 absorption maximum of the donor/acceptor and \( \hbar 2\lambda_m/n \) is the Stokes shift (\( \hbar \lambda_m/n \) being the ‘reorganization energy’). The line-broadening function \( g_{m/n}(t) \) is defined as

\[ g_{m/n}(t) = \int_0^\infty d\omega \frac{\pi \omega^2 C_{m/n}''(\omega)}{(1 - \cos(\omega t))\text{coth}\left(\frac{\hbar \omega}{2k_B T}\right) + i(\sin(\omega t) - \omega t)} \]

In this equation \( C_{m/n}''(\omega) \) denotes the spectral density (also referred to as \( J(\omega) \)). Although spectral densities of pigment-protein complexes can be estimated from QM/MM calculations,\textsuperscript{46-47} here we use an estimate extracted from experimental data\textsuperscript{48}

\[ C''(\omega) = \frac{\pi S_0 \omega^5}{s_1 - s_2} \sum_{i=1,2} \frac{s_i}{7!2\omega_i^2\epsilon} - \frac{\omega}{\sqrt{\omega_i}} \]

where \( S_0 = 0.5, s_1 = 0.8, s_2 = 0.5, \omega_1 = 0.56 \text{ cm}^{-1} \) and \( \omega_2 = 1.94 \text{ cm}^{-1} \). For the optically-forbidden Lut \( S_1 \) state the ‘spectrum’ is a weighted density-of-states function obtained from a phenomenological spectral density:

\[ C''(\omega) = \sum_{i=1,2} 2\lambda_i \frac{\omega_0^2 \gamma_i}{(\omega^2 - \omega_i^2)^2} + 2\lambda_0 \frac{\omega_0^2 \gamma_0}{\omega^2 + \gamma_0^2} \]

This spectral density is composed of two under-damped modes representing the C-C and C=C stretching modes and a single over-damped mode which represents the rest of the thermal bath. As previously reported\textsuperscript{44, 49} the under-damped frequencies \( \omega_1 = 1500 \text{ cm}^{-1} \) and \( \omega_2 = 1150 \text{ cm}^{-1} \) are
experimentally derived. By fitting the two-photon absorption spectrum reported by Walla et al. we obtained the parameters $\lambda_1 = \lambda_2 = 900 \text{ cm}^{-1}$ and $\gamma_1 = \gamma_2 = 300 \text{ fs}$ for the two under-damped modes and $\lambda_0 = 450 \text{ cm}^{-1}$ and $\gamma_0 = 53 \text{ fs}$ for the over-damped mode.

The final parameters needed to compute the rate are the energies of the 0-0 transitions for the two pigments, the so-called ‘site energies’. For Chl-a 612 $\hbar \omega_Q = 14900 \text{ cm}^{-1}$ is assumed based on established models of LHCII. For Lut, $\hbar \omega_S$ is not well-characterized due to the transition being optically-forbidden. We choose $\hbar \omega_S = 14000 \text{ cm}^{-1}$ as previously, based on two-photon and excited state absorption studies of Lut in octanol. However, we previously demonstrated that, due to the broad nature of the Lut S1 transition, there is little dependence of the calculated rates on the range $13500 - 14500 \text{ cm}^{-1}$.

As discussed in the main article, we calculated the forward and backward rate constant for each trajectory. Since $k_{mn} \propto |V_{mn}|^2$ the trend of the rates broadly follows those of the couplings, although difference between trajectories are more pronounced due to the quadratic dependence. The average forward and backward inverse rates (‘hopping times’) are listed in Table S1.

<table>
<thead>
<tr>
<th>Neutral pH</th>
<th>$\langle k_{\text{Chl}\rightarrow\text{Lut}}^{-1} \rangle$ (ps)</th>
<th>$\langle k_{\text{Lut}\rightarrow\text{Chl}}^{-1} \rangle$ (ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>235.0</td>
<td>126021.1</td>
</tr>
<tr>
<td>Neutral pH metaD1</td>
<td>419.5</td>
<td>11726.6</td>
</tr>
<tr>
<td>Neutral pH metaD2</td>
<td>289.2</td>
<td>224923.0</td>
</tr>
<tr>
<td>Low pH</td>
<td>218.3</td>
<td>155076.0</td>
</tr>
</tbody>
</table>

**Table S1:** The average forward (Chl-a612→Lut1) and backward (Lut1→Chl-a612) ‘hopping times’ calculated for the four trajectories discussed in the main article.

The forward hopping times are generally slower for neutral pH than for low pH although all are rather slow. This is due firstly to the inherently weak nature of the coupling and the fact that all trajectories sample effectively uncoupled conformations. The backwards hopping times are effectively infinite, meaning that Chl-a 612→Lut1 excitation transfer is essentially irreversible. This is a direct result of the very large Stokes shift, $\hbar 2\lambda_0$, associated with the Lut1 transition. It is worth noting that this does not preclude Lut1→Chl-a 612 energy transfer but shows that it must occur by some (non-trivial) mechanism not captured by Förster theory.

The rate constants allow for simulation of the energy transfer dynamics of the pigment pair, although this is not very meaningful given that they are embedded within a larger network of pigments. As previously, we construct a coarse grained, purely kinetic model of the LHCII trimer. This is reasonable given the fact that equilibration of excitation energy throughout the ‘pool’ of Chl-a in the trimer occurs much faster (~ 10 ps) the Chl-a612-Lut1 energy transfer. The construction of the model is illustrated in Fig. S1. Each LHCII monomer is composed of 3 ‘sites’: Lut1, Chl-a612, the remaining ‘pool’ of 7 Chl-a. We assume that Lut1 is connected only to Chl-a 612, which is justified by previous explicit calculations of the entire LHCII monomer. Further, we assume that the Chl-a pool is iso-energetic and that equilibration of energy within each pool is instantaneous. If we assume that there is no more than one excitation present within the LHCII trimer at any time (the biologically realistic case) then the stochastic excitation dynamics of the system are defined by equations of motion

$$\frac{d}{dt} \mathbf{P}(t) = \mathbf{K} \mathbf{P}(t)$$

Here
\( \mathbf{P}(t) = \begin{pmatrix} P_1(t) \\ P_2(t) \\ \vdots \\ P_n(t) \end{pmatrix} \)

is a column vector of excitation occupation probabilities, i.e., the probability at time \( t \) that excitation is located on a particular site. In the present case we have \( n = 9 \) sites. The time evolution is governed by \( \mathbf{K} \) which is a \( 9 \times 9 \) matrix. For illustrative purposes, we consider the simpler case of a single LHCII monomer

\[
\mathbf{P}(t) = \begin{pmatrix} P_p(t) \\ P_{\text{Chl}}(t) \\ P_{\text{Lut}}(t) \end{pmatrix}
\]

Here \( P_p(t) \) corresponds to the occupation probability of the Chl-a pool, \( P_{\text{Chl}}(t) \) to Chl-a612 and \( P_{\text{Lut}}(t) \) to Lut1.

**Figure S1:** A schematic diagram of the coarse-grained model of energy transfer in LHCII 2. Solid arrows indicate pathways for excitation hopping while dashed arrows indicate dissipation channels. We assume 3 types of site: Lut1 and Chl-a 612 are treated explicitly while the remaining Chl-a in each monomer are lumped into a single ‘pool’ in which excitation energy is assumed to equilibrate instantaneously.

The transfer matrix is:

\[
\mathbf{K} = \begin{pmatrix}
-\frac{k_{\text{Chl}} - P}{N_{\text{Chl}}} & -k_{\text{Chl}} & k_{\text{Chl}} - P & 0 \\
\frac{k_{\text{Chl}} - P}{N_{\text{Chl}}} & -k_{\text{Chl}} - P - k_{\text{Chl}} - k_{\text{Lut}} - k_{\text{Chl}} & k_{\text{Lut}} - \text{Chl} & k_{\text{Lut}} - \text{Chl} - k_{\text{Lut}} \\
0 & k_{\text{Chl}} - Lut & 0 & 0 \\
0 & 0 & k_{\text{Chl}} - Lut & 0
\end{pmatrix}
\]
In this expression $k_{\text{Chl}}^{-1} = 4 \text{ ns}$ denotes the excitation lifetime of free Chl, $k_{\text{Lut}}^{-1} = 14 \text{ ps}$ \cite{49} the lifetime of Lut, as well as $k_{\text{Chl} \rightarrow \text{Lut}}$ and $k_{\text{Lut} \rightarrow \text{Chl}}$ the calculated Chl-a 612→Lut1 and Lut1→Chl-a 612 hopping rates. $k_{\text{Chl} \rightarrow \text{P}}^{-1} = 1 \text{ ps}$ \cite{49} is the rate of transfer between the pool of $N_{\text{Chl}} = 7$ Chl-a and Chl-a 612. For transfer from the large pool to the single Chl-a 612 there is an entropic penalty given by

$$
k_{\text{Pool} \rightarrow \text{Chl} \rightarrow \text{a612}} = \frac{k_{\text{Chl} \rightarrow \text{P}}}{N_{\text{Chl}}}
$$

$$
k_{\text{Chl} \rightarrow \text{a612} \rightarrow \text{Pool}} = k_{\text{Chl} \rightarrow \text{P}}
$$

The model can be trivially extended to the full trimer in which $k_{\text{PP}}^{-1} = 10 \text{ ps}$ is the rate of excitation hopping between monomers. In coarse grained models of energy transfer in PSII super-complexes this rate has been variously assumed to be 5-25 ps. Since this is much faster than the rate of energy transfer to Lut1 the overall excitation lifetime of the complex is essentially insensitive to the choice of $k_{\text{PP}}$ in this range.

For a given snap-shot structure the mean excitation lifetime of the trimer is given by,

$$
\langle \tau \rangle = -\text{tr}\{\mathbf{C}^{-1} \mathbf{P}(0)\}
$$

Here $\lambda$ denotes the diagonal matrix of eigenvalues of $\mathbf{K}$, $\mathbf{C}$ is the transformation matrix and $\mathbf{P}(0)$ is the vector of initial populations. It is assumed that each Chl-a has an equal possibility of absorbing a photon,

$$
P_{\text{P}}(0) = \frac{N_{\text{Chl}} - 1}{3N_{\text{Chl}}}
$$

$$
P_{\text{Chl}}(0) = \frac{1}{3N_{\text{Chl}}}
$$

and

$$
P_{\text{Lut}}(0) = 0
$$

since the Lut S$_1$ transition is optically forbidden.

The model is a coarse-grained approach in which we only explicitly model the Chl-a 612/ Lut1 pair. The remaining ‘pool’ of pigments are treated in purely kinetic sense. Additionally, we neglected several points such as complementary alterations in the couplings within the Chl pool and alterations in the transition energies of Chl-a 612 and Lut1 although this has been shown to have a small effect on the overall lifetime. Lastly, like the couplings on which they are based, the lifetimes are meaningful only in a relative sense (see main manuscript). It is important to note that we use the term lifetime loosely. Here we mean it in the ergodic sense in which for each uncorrelated snap-shop (1ns apart) we computed the excitation lifetime as if it were a stable, static structure. For the fitting to experimental results, we have shifted the LHCII excited state lifetime at neutral lumenal pH to 2.0ns.\cite{49,53}

REFERENCES


3.. Daskalakis, V., Protein-protein interactions within photosystem II under photoprotection: the synergy between CP29 minor antenna, subunit S (PsbS) and zeaxanthin at all-atom resolution. Physical Chemistry Chemical Physics 2018, 20 (17), 11843-11855.
15.. Daskalakis, V., Protein-Protein Interactions within Photosystem II under Photoprotection: The Synergy between CP29 Minor Antenna, Subunit S (PsbS) and Zeaxanthin at all-atom resolution. Physical Chemistry Chemical Physics 2018.


