Energetic driving force for LHCII clustering in plant membranes

Premashis Manna[1]*†, Madeline Hoffmann[1]†, Thomas Davies[2], Katherine H. Richardson[2], Matthew P. Johnson[2], and Gabriela S. Schlau-Cohen[1]*

[1]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[2]Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, United Kingdom
*pmanna@mit.edu, gssc@mit.edu
†these authors contributed equally to this work

Abstract

Plants capture and convert solar energy in a complex network of membrane proteins. Under high light, the luminal pH drops and induces a reorganization of the protein network, particularly clustering of the major light-harvesting complex (LHCII). While the structures of the network have been resolved in exquisite detail, the thermodynamics that control the assembly and reorganization had not been determined, largely because the interaction energies of membrane proteins have been inaccessible. Here, we describe a method to quantify these energies and its application to LHCII. Using single-molecule measurements, LHCII proteoliposomes, and statistical thermodynamic modeling, we quantified the LHCII-LHCII interaction energy as ~-5 *kBT* at neutral pH and at least -7 *kBT* at acidic pH. These values revealed an enthalpic thermodynamic driving force behind LHCII clustering. Collectively, this work captures the interactions that drive the organization of membrane protein networks from the perspective of equilibrium statistical thermodynamics, which has a long and rich tradition in biology.

Teaser

Single-molecule spectroscopy and modeling resolve the thermodynamics of membrane protein interactions in plants.

INTRODUCTION

Biological function is replete with protein-protein interactions that change the enthalpy (H) and entropy (S) of the system. As a result of the Second Law of Thermodynamics, the spontaneous formation of these interactions requires a decrease in the state function called the Gibbs free energy (G), defined as G=H-TS, where T is the temperature. The use of ideas from equilibrium thermodynamics and its microscopic partner, equilibrium statistical mechanics, has a long and rich tradition in biology (*1*, *2*). In particular, the application of statistical thermodynamics to protein-protein interactions has been a powerful tool for understanding processes such as multivalent binding (*3*), complex formation (*4*) and hydration of protein complexes (*5*). However, these studies have been primarily limited to globular proteins because of the challenges associated with studies of membrane proteins. As a result, a thermodynamic understanding of the many biological interactions within membranes is missing. Here, we develop an approach to quantify the thermodynamic parameters for membrane protein interactions and apply this approach to the most abundant membrane protein on Earth, light-harvesting complex II (LHCII) from green plants.

In plants, solar energy is absorbed by a network of LHCII and other photosynthetic proteins and used to drive a cascade of biochemical processes that ultimately convert CO_2 to carbohydrates. In the membrane, these proteins are found at high density (80%) in an arrangement that has been revealed in exquisite detail. For instance, the structure of the Photosystem II (PSII)-LHCII supercomplex was solved at 3.2 Å resolution through single-particle Cryo-EM (*6*). Such high-resolution structures show the position of the proteins, yet the energetics of the protein-protein interactions that control the assembly and organization of such complexes remain elusive. Through these structural studies, it has also been established that the organization and composition of the plant supercomplexes change with growth conditions. In particular, under high light, such as direct sun, LHCII forms clusters as part of a set of photoprotective processes known as non-photochemical quenching (NPQ) (*7*, *8*). The photosynthetic machinery is susceptible to damage by photoinduced reactive oxygen species (ROS) (*8*) and the fast component of NPQ prevents ROS formation by dissipating excess energy as heat (*9*, *10*). NPQ is activated through the formation of a ΔpH across the membrane, which is generated by proton accumulation from water splitting (*11*). Up to 60% of NPQ occurs in LHCII (*12*), where conformational changes of the protein are thought to activate dissipative pathways within the embedded chlorophyll (Chl) and carotenoids (Car). An equilibrium between light-harvesting and dissipative conformations of LHCII likely plays a role in regulating the extent of NPQ in plants and algae (*13–15*). It has been shown that interactions between LHCII and other proteins, such as PsbS or other LHCII, likely modulate this equilibrium (*16*, *17*). In particular, it has been long established that LHCII clusters under many different conditions (*18–23*), yet the thermodynamic origin of this clustering and how it is enhanced in vivo under NPQ conditions all remain unclear. Indeed, the tendency to form clusters at all is surprising owing to the net negative charge of -21e for LHCII trimers. Thus, the thermodynamic driving forces behind the (re)organization of the thylakoid membrane are missing in the literature, including the related parameters, *e.g.*, entropy, enthalpy, and Gibbs free energy.

In this article, we introduce a method to quantify the interaction energies of membrane proteins and the thermodynamic driving forces associated with the membrane organization, which we apply to LHCII. We performed single-molecule spectroscopy of proteoliposome samples and modeled the photophysics and thermodynamics of the protein clusters within the proteoliposomes. By fitting the model to the experimental single-molecule data, the interaction energies were extracted. The LHCII-LHCII interaction energy was found to be attractive in nature, consistent with observations of clustering, and in the range of several k_BT s, where k_B is the Boltzmann constant. The interaction energy at low pH increased in magnitude by ~30% compared to neutral pH, which drives a transition from moderately clustered LHCII configurations to strongly clustered ones. Quantification of the free energy change associated with the pH drop established that the pH-driven clustering of LHCII is primarily enthalpy driven with a magnitude in the tens of k_BT (*i.e.*, few kJ/mol), indicating strength sufficient to induce LHCII clustering yet weak enough for easy reversibility. These results provide a microscopic basis for the dynamic regulation of membrane protein interactions, which may underlie the responsive nature of membrane protein organization and function.

RESULTS

LHCII proteoliposome system to probe protein-protein interactions.

We investigated the protein-protein interactions between LHCII using proteoliposomes with a systematically increasing number of LHCII. The proteoliposomes were prepared by first extruding the thylakoid lipid mixture through a polycarbonate membrane with 25 nm pore radii to generate liposomes of a well-controlled size. The liposomes were mixed with LHCII to form LHCII-containing proteoliposomes and were purified via a sucrose gradient for a final homogeneous sample (Figure S1). The LHCII-to-lipid ratio was varied to produce LHCII proteoliposomes with an average number of LHCII, <N>, from less than one to ten. The ratios were selected based on lipid size, LHCII size, and the measured incorporation efficiency of LHCII (*20*). The calculated <N> was confirmed through the advent of spectroscopic signatures of LHCII-LHCII interaction, as discussed below. Successful formation of the proteoliposomes was determined through dynamic light scattering (DLS), which showed a single peak indicating a monodisperse sample with a hydrodynamic radius of ~30-45 nm (Figure S2). The hydrodynamic radius of the proteoliposomes increased with LHCII content, likely due to the hydrophilic surface of LHCII (Table S1) (*24*). Linear absorption and emission spectra of the purified LHCII proteoliposomes (Figures S3 and S5) showed the intact LHCII trimeric structure was maintained.

LHCII quenching with increasing protein density.

The fluorescence lifetime reports on the overall photophysical pathways, and several studies have established that it decreases in the presence of LHCII-LHCII interactions (*18*, *19*). To characterize the overall dependence of the lifetime on the protein density, we performed ensemble time-resolved fluorescence measurements. The fluorescence decay traces were best fit with a bi-exponential function and the average lifetimes were calculated from these two terms (Figure S8).

To investigate the distribution of fluorescence lifetime values, we performed single-molecule time-resolved fluorescence measurements (Figure S14-15). The lifetimes were extracted from the single-molecule data by fitting the fluorescence decay curves to single exponentials using maximum likelihood estimation (Figure 1a; SI Section 8). Although a bi-exponential function yielded the best fit for the ensemble decay curves, a single-exponential function was used owing to the lower signal-to-noise ratio in single-molecule data. Single-molecule measurements require high excitation fluences, and so light-harvesting proteins photodegrade during the observation time. To minimize the effect of photodegradation, the lifetime was only characterized immediately upon illumination, prior to any changes in the emission properties. The fluorescence lifetimes were measured and analyzed from many individual LHCII-proteoliposomes (~100) for each sample to generate lifetime histograms as shown in Figure 1b.

The average (squares) and median (circles) lifetimes from the ensemble and single-molecule measurements, respectively, for all LHCII-proteoliposome samples at neutral and at acidic pH are shown in Figure 1c, Tables S2 and S3. The lifetime values decreased with protein density and with the pH drop in both ensemble and single-molecule measurements, revealing two types of quenching processes: (1) cluster-dependent quenching; and (2) pH-dependent quenching.

Cluster-dependent quenching. The ensemble measurements yielded an average lifetime of 2.67 ns for <N>= 1, consistent with previous work (*25*). The single-molecule measurements yielded a slightly shorter median lifetime of 2.45 ns for <N>= 1, likely due to a small remaining effect of photodegradation. The lifetimes gradually decreased with protein density down to 1.88 ns for <N> = 10 for the ensemble data and 1.22 ns for <N> = 10 for the single-molecule data, all at pH 7.5. These values give an overall ~30% and ~50% reduction for the ensemble and single-molecule data, respectively (Figure 1d), corresponding a ~3-5% reduction per LHCII. Additionally, a ~20% difference between the ensemble and single-molecule measurements emerged at higher protein densities. Photodegradation typically results in the generation of quenchers, which can also quench neighboring proteins. The effect of photodegradation is thus expected to increase with protein density, consistent with these observations (*25*, *26*).

pH-dependent quenching. The lifetimes were also measured after overnight incubation at pH 5 and decreased from the pH 7.5 values for the samples with <N>=1, 2 and 5 (Figure 1c,e). For LHCII-proteoliposomes with high (<N> >5) protein density, long incubation times (>150 hrs) are required for LHCII to fully equilibrate into clustered configurations in the membrane (SI Sec. 6). However, proteoliposomes become unstable after a few days, particularly at low pH, so the pH-dependence of these samples could not be investigated. The lifetime of LHCII-proteoliposome samples with <N>=1 decreased to an average of 2.57 ns for the ensemble data and a median of 2.02 ns for the single-molecule data, corresponding to a <10% reduction. This minor quenching at low pH could be either due to a pH-dependent conformational change of LHCII or a change in the local electrostatic environment.

Similar to the pH 7.5 data, the lifetimes at pH 5 gradually decreased as the protein density increased down to 1.67 ns for <N>=5 for the ensemble data and 0.96 ns for <N>=5 for the single-molecule data. The magnitude of the pH-induced lifetime reduction increased with protein density as shown in Figure 1e. The pH-induced reduction rose to ~25% and ~30% for the ensemble and single-molecule data, respectively (Figure 1e), for <N>=5, corresponding to a ~5% reduction per LHCII. The difference in magnitude between the ensemble and single-molecule data is likely

due to the presence of photodegradation in single-molecule data, which is more prevalent at low pH. Overall, the larger reduction in lifetime with the pH drop for higher protein densities likely arises from the ability of these samples to adopt more clustered configurations.

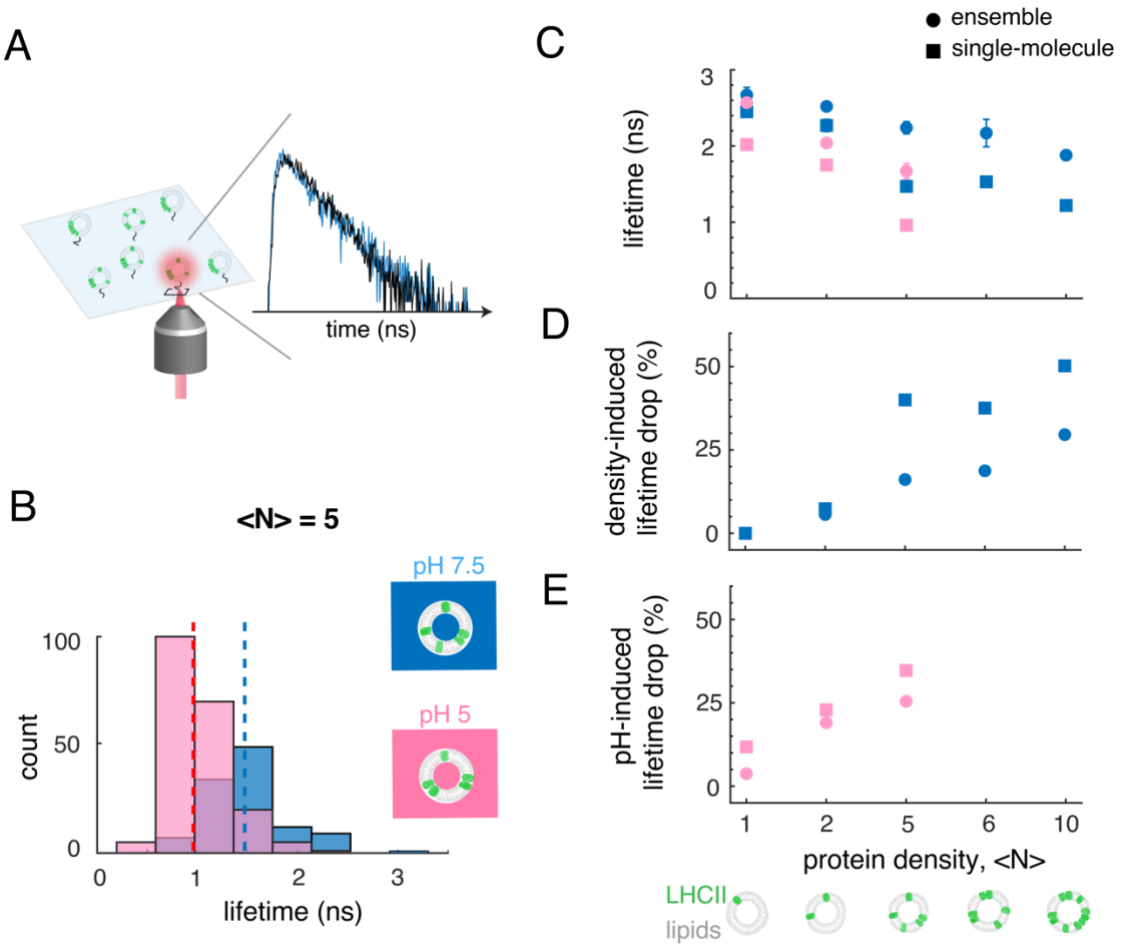


Figure 1. pH- and cluster-mediated quenching of LHCII in the liposome. (a) Schematic of LHCII-proteoliposomes immobilized on a coverslip for single-molecule measurements. A representative single-molecule fluorescence transient shows the fluorescence decay (blue) and its single-exponential fit (black). (b) Representative lifetime distributions obtained from single-molecule measurements of LHCII-proteoliposome samples with an average number of proteins per liposome (<N>) of 5 at pH 7.5 (blue) and 5.0 (pink). The dotted vertical lines are the medians of the distributions. (c) Lifetimes of the LHCII-proteoliposome samples at pH 7.5 (blue) and 5 (pink) at various protein densities. (d) Density- and (e) pH-induced lifetime drop of the proteoliposomes at different protein densities (see SI Sec 5). Circles and squares represent the values from the ensemble and single-molecule measurements, respectively.

Extraction of LHCII-LHCII interaction energies.

The quenching with increasing protein density arises from protein-protein interactions between LHCII. Thus, the fluorescence lifetime provides a reporter for LHCII-LHCII interactions, where shorter lifetimes indicate more (stronger) LHCII-LHCII interactions and longer lifetimes indicate fewer (weaker) LHCII-LHCII interactions. Based on this dependence, we developed an analysis method to quantitatively extract LHCII-LHCII interaction energies from the single-molecule lifetime distributions of the LHCII-proteoliposomes as illustrated in Figure 2 (SI Sec. 15).

In the first step of our method (Figure 2a), we developed a stochastic model adapted from Gruber et al. (*27*) to predict the fluorescence lifetimes of LHCII clusters of different sizes. This model takes known photophysical properties of LHCII and incorporates the influence of singlet-triplet annihilation and the dependency of such annihilation process on cluster size (n), which arises from the ability of a triplet excitation to quench singlets in the surrounding LHCII within the cluster. Figure 2b displays the lifetime as a function of n. The model predicts a lifetime of 2.5 ns for $n = 1$, consistent with the experimental value of 2.7 ns for the <N>=1 proteoliposomes, which decreases to <0.5 ns for $n = 20$. Thus, singlet-triplet quenching can serve as a reporter of n in proteoliposomes, although it is unlikely to be the quenching pathway responsible for NPQ (*8*, *28*). Incorporation of a separate quenching pathway had minimal impact on this model (Figure S27).

In the second step of our method (Figure 2c), we characterized different configurations of the same number of proteins (N) containing different cluster sizes (n). For a given number of proteins, the possible number of configurations (m) was computed. For instance, with N=5, seven different configurations are possible ranging from completely clustered to completely unclustered as illustrated in Figure 2c. The probability of formation (w) of each configuration was also computed based on their geometry, which depends on the number of proteins, the size of the proteins, and the size of the proteoliposome (SI Sec 12). As expected, w decreased for configurations with more clustering. The lifetime of each configuration was also calculated using the lifetime of its constituent clusters as described above.

In the third step, we quantified the equilibrium populations of each configuration, which depend both on the enthalpy and the entropy of the system. The enthalpy of the LHCII proteoliposome system can be mathematically described as a function of pairwise LHCII-LHCII interaction energy (J) while the entropy is related to the probability of formation (w). Figure 2d displays the population of different configurations for N=5 as a function of J.

In the fourth step, we generated simulated lifetime distributions from the number of proteins (N), the relative populations of the different configurations with a given N, and the lifetimes of each configuration (Figure S30 and S31). Figure 2e displays the simulated lifetime distributions (gray) overlaid with the experimental lifetime distribution (blue) for <N> = 5 at J of -8, -5 and -2 k_BT. The simulated lifetime distribution is in good agreement with the experiment at J = -5 k_BT, indicating the value of J is in that vicinity.

In the final step, the simulated lifetime distributions for all <N> were globally fit to the experimental ones at each pH using maximum likelihood estimation (MLE) to extract J (SI Sec 15). Figure 2f shows the log-likelihood estimates as a function of J (fits for individual samples are shown in Figure S32). The best fits for the LHCII-LHCII interaction energy were obtained at -5.4

and -7.2 k_BT for the samples at pH 7.5 and pH 5, respectively. The negative values signify that LHCII-LHCII pairwise interaction energies are attractive in nature, and the attraction is strengthened by ~2 k_BT upon the pH drop.

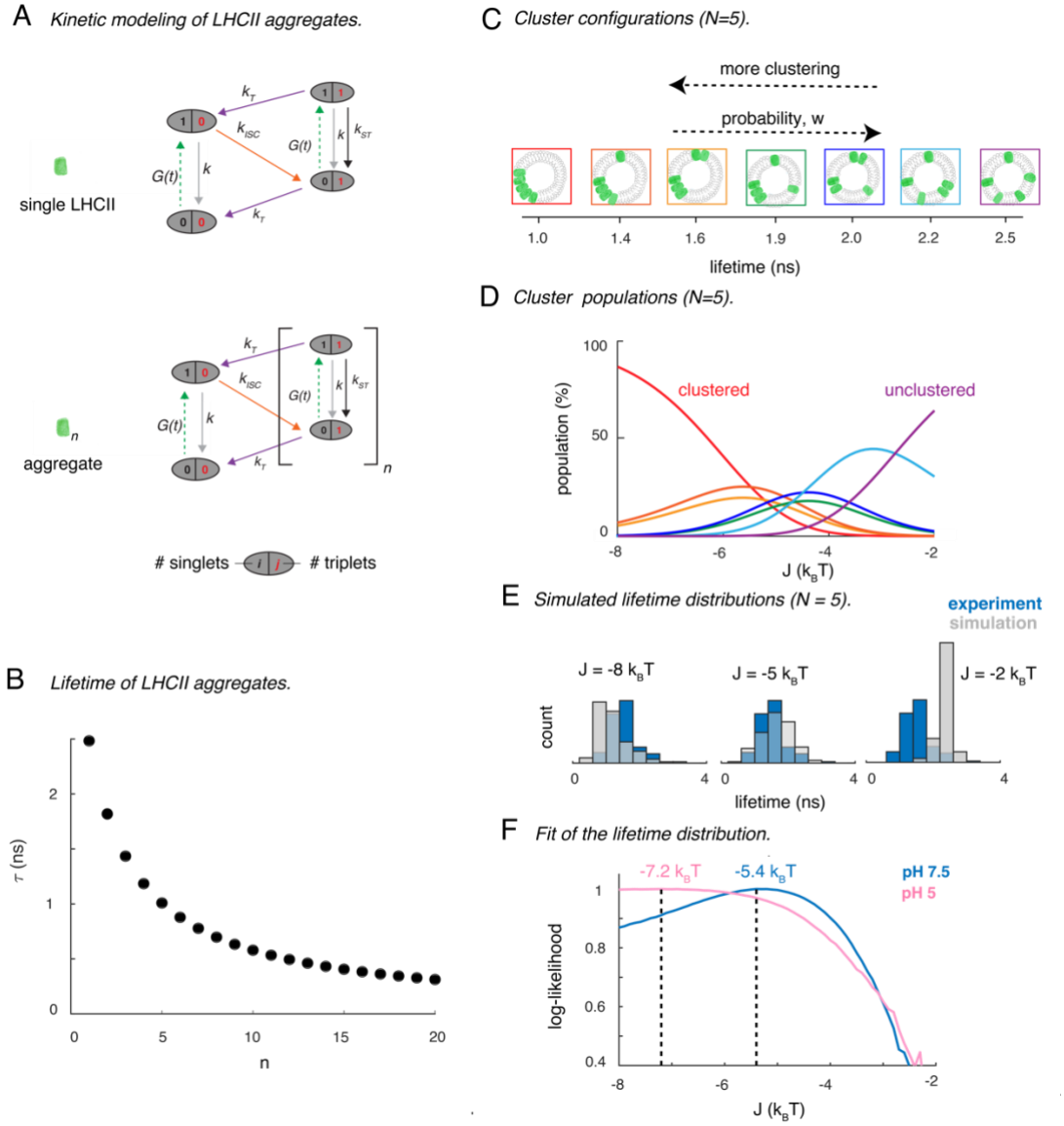


Figure 2. Extraction of LHCII-LHCII interaction energies from single-molecule lifetime data. (a) Stochastic kinetic model of the excited state relaxation pathways of isolated and clustered LHCII. The ovals represent states with different numbers of singlets (black) and triplets (red). $G(t)$ is the excitation rate; k and k_T are rate constants for linear de-excitation from singlet and triplet states, respectively; k_{ISC} is the inter-system crossing rate-constant; and k_{ST} is the rate-constant for singlet-triplet annihilation. (b) Fluorescence lifetime computed from the model in (a) as a function of the number of LHCII complexes in a cluster, n. (c) All possible cluster configurations of LHCII-proteoliposomes upon the incorporation of a total number of proteins, N, of five and the lifetime of each configuration estimated from the model. (d) The populations of these seven configurations (color-matched to the boxes of the respective configurations in (c)) are displayed as a function of LHCII-LHCII interaction energy, J. (e) The relative

populations of these configurations are utilized to obtain lifetime distributions (gray histograms) at J = -8, -5, and -2 k_BT. The simulated distributions are overlaid with the experimental lifetime distribution (blue histograms). (f) Results of global fits of the simulated distributions to the experimental ones to extract J at pH 7.5 (blue) and 5 (pink).

Free-energy driving force for LHCII clustering in liposomes.

Using the identified interaction energies, we quantified the free energy driving force associated with the reorganization of LHCII into clusters upon a pH drop. We derived the following expression for the free energy driving force ($\Delta\Delta G$) for the transition of LHCII-proteoliposomes from neutral to low pH:

$$\Delta\Delta G = \sum_{i=1}^{m} \Delta\Delta H_i(J) - k_B T * \ln\left(\sum_{i=1}^{m} \frac{w_{l,i}}{w_{n,i}}\right) \quad (1)$$

where $\Delta\Delta H_i(J)$ is the change in enthalpy associated with i-th configurations and $w_{l,i}$ and $w_{n,i}$ are the probabilities of such configurations at pH 5.0 and 7.5, respectively. The first term of the right-hand side of the above equation represents the enthalpy change and the second term represents the change in entropy for this transition. For <N>=5, enthalpy and entropy changes were found to be -12.7 k_BT and -4.5 k_BT, respectively. These values show that this transition is mildly exothermic in nature and associated with a reduction in entropy presumably due to a more restricted organization with clusters. From the changes in enthalpy and entropy, the Gibbs free energy change is quantified as -8.2 k_BT for <N>=5. For other protein densities, the free energy changes ranged from -2 to -19 k_BT (Figure 3a, Table S9). While the magnitudes of the changes in entropy, enthalpy and Gibbs free energy all increased with protein density, the change in entropy only increased slightly whereas the change in enthalpy increased appreciably, and so the free energy became even more dominated by the enthalpic contribution at high protein density.

Free-energy driving force for enhanced LHCII clustering in the intact chloroplast.

To quantify the free-energy change that drives the enhancement of clustering under high light in vivo, we analyzed the freeze- fracture electron micrographs of the intact spinach chloroplasts reported by Johnson et al. (*7*) (SI Sec. 18). The dark-adapted (vio, dark) and light-adapted (zea, light) membranes were assigned to light-harvesting and dissipative states, respectively, based on the reported NPQ values (*7*). Particle picker transforms of the electron micrographs were used to identify the LHCII organization in the membrane (Figure 3b, SI Sec. 18). The transforms were analyzed for each ~88 nm × 88 nm region of the micrographs, which is similar in surface area to the proteoliposomes. Based on the extracted membrane organizations and the LHCII interaction energies extracted from the LHCII-proteoliposome experiments, we calculated $\Delta\Delta H$ associated with the transition from the light-harvesting to a dissipative state. The average value of $\Delta\Delta H$ in the chloroplast scaled similarly to the values from the LHCII-proteoliposomes. For instance, in proteoliposome with <N> = 10, we calculated $\Delta\Delta H$ ~- 28 k_BT associated with the transition from high to low pH. Now, with a five-fold increase in the density (<N> = 50) in the chloroplast, the

change in enthalpy is found to be ∼-115 k_BT. The large error associated with the ΔΔH value in the chloroplast reflects the stochasticity in protein organization in the different regions of the membrane.

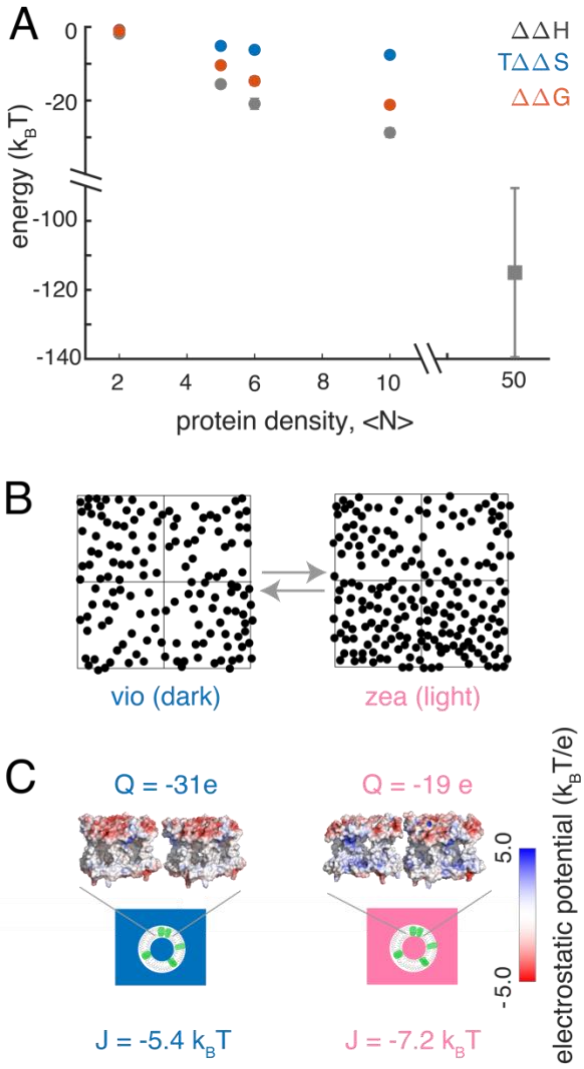


Figure 3. Free-energy change of pH-induced clustering of LHCII. (a) The changes in enthalpy (ΔΔH), entropy (ΔΔS) and Gibbs free-energy (ΔΔG) in the LHCII proteoliposomes (circles) and chloroplasts (square) for the pH-induced (i.e., high light-induced) clustering. The error bars are the standard errors. (b) Masks produced by thresholding the freeze-fracture electron micrographs of the intact spinach chloroplasts of the dark-adapted (vio, dark) and light-treated (zea, light) samples reported in (*7*). The square boxes are the ∼88 nm × 88 nm regions over which the thermodynamics parameters involving a transition from dark to light organizations were calculated. (c) Schematic of lateral interactions between two LHCII apo-proteins at pH of 7.5 (left) and 5 (right). LHCII proteins are shown as potential energy surfaces generated by the Adaptive Poisson-Boltzmann solver in Pymol with the net surface charges of the proteins (Q) also displayed.

DISCUSSION

Molecular-level origin of LHCII-LHCII interactions.

We quantified the LHCII-LHCII interaction energy as -5.4 k_BT at neutral pH and at least -7 k_BT at acidic pH (2f, SI Sec 15, Fig S31). The negative sign implicates these interaction energies are indeed attractive. Schneider et al. estimated LHCII-LHCII stacking interactions across the stromal gap as -4 k_BT based on the electrostatic attraction between two correspondingly charged parallel plates in an aqueous solution of counter ions (*29*). Therefore, the lateral LHCII-LHCII interaction obtained in the current work is ~30% stronger than the predicted stacking interaction, consistent with the formation of lateral protein networks within the thylakoid membrane (*7*, *30*, *31*).

The interactions between the proteins can be viewed as a balance between electrostatic repulsion and van der Waals attraction forces (*32–34*). We calculated the net surface charge of an LHCII trimer at pH 7.5 to be -31e, which arises from the carboxyl groups associated with the glutamic acid and aspartic acid residue side chains (*35*, *36*). The van der Waals attraction arises due to the interactions between uncharged molecules with permanent and/or instantaneously induced electrostatic dipole moments. As the measured LHCII-LHCII interaction energy is attractive in nature, the attractive van der Waals force must overpower the repulsive electrostatic force. Consistently, a recent computational study showed PSII-LHCII interactions are attractive due to the same balance of van der Waals and electrostatic forces (*37*). Because the majority of charges in LHCII are localized at the terminal regions, the transmembrane part of the protein is largely hydrophobic in nature. This charge distribution minimizes the electrostatic repulsion between LHCII trimers, thereby leading to enhanced van der Waals attraction. The net charge and overall surface charge density are highly dependent on the membrane lipid composition, local ionic strength, and charged residues, and so the specific amplitudes of the two opposing forces are likely adjusted by these factors in vivo. For example, phosphorylation of LHCII introduces additional negative charges, which likely increase the electrostatic repulsion among LHCII and weaken the LHCII-LHCII lateral interactions (*32*, *38*, *39*). Alternatively, with an abundance of cations in the medium, the effective net charge of LHCII is decreased and the magnitude of the repulsive interactions is also decreased, strengthening the LHCII-LHCII interactions. Indeed, the effect of screening on LHCII interactions has been observed by Kirchhoff and coworkers using positively charged ions (K^+, Mg^{2+}) (*32*).

What could be the molecular-level origin of enhanced interaction at low pH? van der Waals forces depend on the size of the proteins, inter-protein distances and the dielectric constant of the medium, all of which are nearly constant between low and neutral pH. We propose that the net electrostatic repulsion between the proteins decreases at low pH due to the screening effect of H^+, the protonation of negatively charged lipids in the membrane, or the protonation of certain negatively charged amino acid residues of LHCII exposed to the solvent. Our estimation of the net surface charge of LHCII proteins (Q) in low and neutral pH supports this hypothesis (Figure 3c, Sec SI 15). The surface charge of LHCII is mostly located at its edges facing the stromal and lumenal sides of the membrane while the central transmembrane region is largely neutral. The strength of Q was found to decrease from -31e at pH 7 to -19e at pH 5, corresponding to a 38%

reduction of negative charge at low pH consistent with the measured ~30% increase in pairwise interaction energy.

Enhanced LHCII clustering at low pH.

Intense work in the last few decades has found that LHCII is a major site for NPQ in higher plants and that ΔpH is necessary for its trigger (*7*, *12*, *40–42*). Numerous previous studies have also revealed the tendency of LHCII proteins to cluster (*18*, *19*, *21*, *23*). Our statistical thermodynamical modeling explains the clustering of LHCII and its enhancement at low pH, providing a basis for the organization of LHCII and its dependence on NPQ conditions. LHCII-LHCII interactions are required for cluster formation, as the population of clustered configurations in the proteoliposomes is near zero when the interaction energy is not present, *i.e.*, at J=0 (Figure 2d). At neutral pH where moderate interactions are present (J = -5.4 k_BT), a variety of configurations co-exist primarily containing a mixture of clustered and unclustered LHCII. The interaction energies produce an enthalpic driving force (ΔH) that overcomes the entropic driving force (ΔS) associated with isolated LHCII to induce moderate clustering. At low pH where stronger interactions are present (J = -7.2 k_BT), the population of clustered configurations considerably increases due to the exponential dependence on enthalpic energy ($\Delta H(J)$, see Methods, Eqn. 5). For instance, in the LHCII-proteoliposomes with <N>=5, the percentage of the completely clustered configuration, where all five LHCII form a single cluster, is 25% at the neutral pH and increases to 74% at low pH (Figure S34). Ultimately, the difference in interaction energies leads to an enthalpic driving force (ΔΔH) that induces a transition from moderately clustered configurations at neutral pH to heavily clustered configurations at low pH.

LHCII clustering from a thermodynamic perspective.

We also quantified the changes in enthalpy (ΔΔH), entropy (ΔΔS), and free energy (ΔΔG) for LHCII clusters associated with the transition from a light-harvesting (neutral pH) to dissipative (low pH) state (Figure 3a). Our extracted parameters established a free energy driving force for enhanced clustering, which is consistent with in vivo observations. The change in enthalpy is favorable ($\Delta\Delta H < 0$) owing to the attractive LHCII-LHCII interactions. In contrast, the change in entropy is unfavorable ($\Delta\Delta S < 0$), as expected from the reduced probabilities of highly clustered configurations. The larger magnitude of the enthalpic term as compared to the entropic term leads to a $\Delta\Delta G < 0$, enabling spontaneous or thermodynamically feasible enhancement of LHCII clustering at low pH (Table S9). Thus, these results show that the enhanced clustering of LHCII in the liposome or natural thylakoid membrane at low pH is primarily driven by enthalpy.

The magnitude of the enthalpic term increases dramatically with protein density owing to the additional possible interactions, which scale combinatorically with the number of LHCII. In contrast, the magnitude of the entropic term only changes minimally owing to the relatively dilute LHCII density (Figure 3a). That is, the similar number of configurations available for all the proteoliposome samples in this low-density limit leads to similar entropic values. The increase in the magnitude of ΔΔH, and thus the magnitude of ΔΔG, makes the clustering process more exergonic in nature as protein density increases. However, at the even higher protein densities found in the chloroplast (equivalent to <N>=50), the magnitude of the free energy change may

be different compared to its liposome counterpart. The thylakoid environment is extremely crowded with different proteins, and so the entropic term for LHCII clustering may be smaller owing to the smaller effective area. The clustering in vivo is likely more complicated owing to the presence of other proteins. The membrane reorganization in vivo may, therefore, be affected by a contribution from ΔS, similar to the contribution of this parameter to state transitions (*43*). While the values may be modified, the free energy driving forces quantified here are on the order of a few k_BT s, and thus represent values large enough that they likely induce the membrane reorganization under NPQ conditions, yet are low enough so that their effect can be reversible through thermal motion and modulated by other external factors.

The lateral interactions of membrane proteins play an important role in the regulation of photosynthesis and numerous other fundamental biological processes. Large-scale molecular dynamics simulations have been applied to investigate such interactions (*37*, *44*). However, unlike globular proteins, experimental characterization of membrane protein interactions had not been possible. Here, by employing single-molecule spectroscopy and statistical thermodynamic principles, we have developed a method that directly quantifies LHCII-LHCII interactions. This method can be applied to probe the interactions among other photosynthetic proteins within the PSII-LHCII supercomplexes (*30*, *31*) and can be readily extended to other membrane protein networks with the use of exogenous fluorophores.

In conclusion, we extracted the lateral interaction energies at neutral and acidic pH for LHCII, the most abundant membrane protein on Earth. The LHCII-LHCII interaction energy at neutral pH is approximately -5 k_BT, which creates a strong but reversible binding to form clusters of various sizes. The interaction energy is ∼30% stronger at pH 5 compared to at pH 7.5, which drives a transition from moderate clustering to strong clustering. The free energy change associated with the pH drop is dominated by enthalpy and on the order of 10s k_BT s, which is strong enough to form clusters but low enough to be reversible and regulated by other cellular factors. These values capture the underlying mechanisms behind the observed yet previously unexplained reorganization of the plant membrane, which is thought to play an active and important role in the regulation of photosynthesis. Such dynamic and responsive reorganization of membrane networks is a fundamental component of biological function. Overall, this work presents a framework to analyze membrane protein interactions using equilibrium statistical thermodynamics that can be extended to investigate membrane organization for a range of biological systems.

Materials and Methods

Preparation of LHCII-proteoliposomes

LHCII proteins used in this work were purified from the dark-adapted spinach as described in details elsewhere (*25*). The detailed protocol for the preparation of LHCII proteoliposome at neutral pH (pH, 7.5) and their characterizations are reported in the SI (SI Sec. 3) of this article and also in (*25*). For the low pH measurements, proteoliposome prepared at neutral pH was incubated for ∼10-12 hrs. at (20 mM MES, 40 mM NaCl, pH 5). The size of the incubated liposomes was monitored at the different phases of incubation using dynamic light scattering (Figure S10).

Ensemble Measurements

Absorption and emission spectra of the LHCII samples were collected using an Epoch Microplate Spectrophotometer (BioTek) and a Cary Eclipse Fluorescence Spectrophotometer, respectively (Figure S3-S5). The samples were syringe filtered (GE Healthcare Life Sciences, pore size 0.22 μm) to discard any aggregates before the measurements.

The fluorescence decays were measured with a TCSPC module (Time Tagger 20, Swabian Instruments). For excitation, a tunable fiber laser (FemtoFiber pro, Toptica Photonics, 80 MHz repetition rate, 130 fs pulse duration, 610 nm, 4 nm full-width half maximum (fwhm)), passed through a pinhole and directed into a home-built confocal microscope. The excitation was focused by an oil-immersion objective (UPLSAPO100XO, Olympus, NA 1.4) onto the sample placed on a coverslip. The emission of the sample was collected through the same objective and separated from the excitation using a dichroic (ZT647rdc, Chroma) and bandpass filter (ET700/75m, Chroma and ET690/120x, Chroma)). The fluorescence decay was fit by iterative re-convolution with a bi- exponential function using the measured instrument response function (IRF) of the system with a home-built MATLAB code. The IRF was measured by the scattered signal to be ~400 ps (fwhm). The average lifetime values were obtained through an intensity-weighted average of the fitted bi-exponential lifetime constants (Table S2, Figure S8).

Single-molecule Measurements

The filtered samples were diluted to ~15 pM LHCII in 20 mM HEPES, 40 mM NaCl, pH 7.5 (for neutral buffer) or 20 mM MES, 40 mM NaCl, pH 5 (for acidic buffer). For the measurement at neutral pH (pH 7.5), we used an enzymatic scavenging mixture containing 2.5 mM protocatechuic acid (PCA) and 25 nM protocatechuate-3,4-dioxygenase (PCD). A mixture of 20 nM Pyranose Oxidase from Microorganism (Creative Enzymes, Shirley, NY), 1.2 μg/ml of catalase (Sigma-Aldrich) and 100 mM glucose was used as scavenging solution for low pH measurements. For the low pH measurements, argon gas was also used to avoid intense photobleaching. The proteoliposomes were immobilized on a homemade or commercial (Bio 01 low-density glass coverslip, 22 × 22 mm, Microsurfaces Inc.) biotinylated coverslip via neutravidin-biotin interactions. The method of single-molecule data collection and analysis has been described in the SI Sec. 7-8.

Modeling

Stochastic modeling of LHCII clusters

We developed a model of the excited-states of LHCII clusters to predict the fluorescence lifetimes. Upon photoexcitation of their singlet states, chlorophyll molecules relax via several competing pathways (*45*): (1) radiative decay from the Chl singlet states ($1/k$ ~5 ns); (2) inter-system crossing to Chl triplet states ($1/k_{ISC}$ ~10 ns), which subsequently transfer to Cars. Under the high excitation fluence and repetition rate of single-molecule measurements, there can be a substantial accumulation of such Car triplets. The Car triplets de-excite any singlet excited states via singlet-triplet annihilation (*46*), which has been extensively described in LHCII (*27*, *47*). In

this study, the average number of proteins in the proteoliposome is ten or below, and these discrete numbers of trimers are best described with a stochastic model, similar to Gruber et al. (*27*) (SI Sec. 13). The contribution of singlet-singlet annihilation is insignificant (<2.7%) under the excitation fluences used in this work and therefore was excluded from the model.

Owing to the influence of S-T annihilation, the excited-state lifetime of a single LHCII depends on the triplet concentration. Assuming a random distribution of triplet quenchers over LHCII aggregates and the S-T annihilation process being independent of their spatial coordinates, the dependence of lifetime on the triplet concentration ([T]) can be calculated (*46*, *47*):

$$\tau = \frac{1}{k + k_{ISC} + k_{ST} * [T]} \tag{2}$$

where k, k_{ISC} and k_{ST} are the rate constants of the radiative decay, inter-system crossing and singlet-triplet annihilation, respectively (*27*). The equation also assumes that the triplet population has reached a quasi-stationary condition as the decay of triplet states (microseconds) is much slower than the singlet state decay (nanoseconds). The rate constants used in this model were obtained from (*27*) and shown in the SI Sec 13. At higher illumination intensities, the triplet state population ([T]) increases, resulting in an enhanced quenching of the singlets (Eqn.2). In the first step of the analysis, we described the excited-state lifetime of LHCII and LHCII clusters (Figure 2a). To validate our model, we performed intensity-dependent single-molecule measurements of unclustered LHCII (Figure S28). The lifetime decreased with illumination intensity, as expected. The model predictions matched reasonably well with the median values of the experimental lifetime distributions (Table S7).

The amount of S-T annihilation is also dependent on the cluster size (*48*). As the size of the cluster increases, the singlets can be quenched by triplets from other LHCII, increasing the probability of the S-T annihilation. In the small clusters investigated here, we assume a triplet present in a single LHCII can quench the excitations in the entire cluster. In this limit, the total number of effective triplets or magnitude of triplet quenching can be modeled to increase linearly with the cluster size. Therefore, for a cluster consisting of n number of LHCII molecules, the lifetime of its singlet excited state is given as:

$$\tau = \frac{1}{k+k_{isc}+k_{ST}*(n*[T])} \tag{3}$$

where, [T] is the triplet concentration in an isolated LHCII. Next, we used this model to calculate the excited-state lifetime as a function of the size of the LHCII cluster (n) as displayed in Figure 2b. Stochastic modeling of LHCII photophysics, characterization of the different clusters and their equilibrium populations, generation of simulated lifetime distributions at different protein-protein interaction energies (J), and extraction of J based on maximum likelihood estimation have been described in detail in the Supplementary Information section of this article.

Cluster configurations and populations

At high protein densities, LHCII clusters form on liposomes that exhibit quenching of the fluorescence emission (*19*, *48*). The cluster formation mimics the in vivo formation of LHCII

arrays under high light conditions (*7*, *49*). In the second step of the analysis (Figure 2c), we characterized the clusters by employing a classical statistical mechanics description of an interacting system (*50*). The configurations for a given number of proteins, *N*, can be obtained by solving the equation below (SI Sec 10) (*50*):

$$\sum_{l=1}^{N} l * m_l = N \quad (4)$$

where the system forms m_l number of clusters of l proteins. The m-matrix constructed from the solutions to the equation above contains the possible configurations of the N proteins, *i.e.*, the sizes of the clusters. For example, when $N = 5$, seven different configurations of the five LHCII are possible as shown in Figure 2c. In the most clustered configuration (left), all the LHCII form one cluster of five proteins whereas in the least clustered configuration (right), the proteins are all separated. As the number of proteins increases, the number of configurations also increases (Figure S16b). For instance, when $N = 10$, 42 different configurations are possible.

However, the probability (w) of forming these different types of configurations varies depending on the total number of proteins involved (N), the geometry of clusters, i.e. the connectivity of the proteins in the configurations, and also the available area to form the clusters. For instance, the probability of forming a cluster of two proteins in a liposome can be estimated as follows: assuming one LHCII protein is already on the liposome, if we incorporate another protein, what is the chance that the second protein will be within 1 nm distance (which is our definition of the cluster) to the first one to form a cluster of two proteins vs. both remains unclustered. It turns out to be that assuming zero interactions between the proteins, the probability of forming two clusters on a liposome with a radius of 25 nm is only 0.006. The probability of forming different types of configurations on a spherical surface (w) is discussed in detail in the Supplementary Information section of this article (SI Sec 12).

The lifetime of each configuration is computed by first calculating the lifetime for each cluster within the configuration using the model values displayed in Figure 2c and second combining the values to describe the full configuration. Figure 2b shows that the configurations with more clustering are characterized by a shorter lifetime and lower probability of formation (w).

At equilibrium, the complex interplay between entropy and enthalpy determines the population of a configuration (*51*). In a clustered system like the LHCII-proteoliposomes, the enthalpy (ΔH) can be mathematically represented as a function of the pairwise interaction energy between proteins. We introduce this parameter, the pairwise LHCII-LHCII interaction energy, as J. For a given number of LHCII in the liposome, the equilibrium population of the i^{th} configuration (p_i) depends on the probability of forming that configuration (w_i) and its total enthalpy ($\Delta H_i(J)$) in the following way:

$$p_i = w_i * e^{-\frac{\Delta H_i(J)}{k_B T}} \quad (5)$$

where, k_B and T are the Boltzmann constant and temperature, respectively.

The enthalpy ($\Delta H(J)$) of a configuration depends on how the LHCII proteins are clustered in that configuration. For instance, in a configuration of five LHCII where two and three proteins are clustered (Figure 2c), $\Delta H(J)$is given as:

$$\Delta H(J) = (2-1) * J + (3-1) * J = 3J \quad (6)$$

Figure 2d displays the population of the seven configurations for N=5 as a function of *J*. When *J* is close to zero, mostly unclustered configurations are favored as there is no energetic driving force to form clusters. In contrast, when *J* is large (close to -10 k_BT), clustered configurations are favored. At an intermediate *J*, populations with both clustered and unclustered configurations coexist.

Simulated lifetime distribution and extraction of LHCII-LHCII interaction energy

For a liposome with a certain number of LHCII embedded in it, we simulate the possible cluster configurations, their excited state lifetimes, and relative populations from our model. This enables us to generate the lifetime distribution of that sample (SI Sec 15). First, the number of proteins incorporated into a given liposome, *N*, is selected with a probability given by a Poissonian distribution with the average value, <N>, for the sample. All possible configurations for *N* are modeled. For instance, if an *N* of five is selected, we model all seven configurations as shown in Figure 2c. The relative population of these configurations depends on the pairwise LHCII-LHCII interaction energy (*J*) as discussed above (Figure 2d). Next, at a certain *J*, a possible configuration is selected and weighted by its relative population. The lifetime of that configuration is computed from the stochastic model as discussed above (also in SI Sec 13). This whole process is repeated multiple times to generate a lifetime distribution for that sample at a particular *J* (Figure 2e).

Free-energy change calculation

At equilibrium, the Gibbs free energy ($\Delta\Delta G$) change for a transition from neutral to low pH is described in the following way:

$$\Delta\Delta G = -k_B T ln\left(K_{eq}\right) = -k_B T ln\left(\sum_{i=1}^{m} \frac{p_{l,i}}{p_{n,i}}\right) \quad (7)$$

where, K_{eq} is the equilibrium constant, $p_{n,i}$ and $p_{l,i}$ are the population of the *i*-th configurations at the neutral (pH 7.5) and low pH (pH 5), respectively. *m* is the number of possible configurations for a given number of proteins embedded in the liposome as discussed above. Now, the population of the i^{th} configuration, p_i is given as shown in Eqn. 5. Inserting Eqn. 5 in Eqn. 7 and slight re-arranging we obtain Eqn. 1 which is used to compute the free-energy change of the transition from neutral to low pH (see also SI Sec. 17).

REFERENCES

1. K. A. Dill, S. Bromberg, D. Stigter, *Molecular driving forces: statistical thermodynamics in biology, chemistry, physics, and nanoscience* (Garland Science, 2010).
2. H. G. Garcia, J. Kondev, N. Orme, J. A. Theriot, R. Phillips, "Thermodynamics of biological processes" in *Methods in enzymology* (Elsevier, 2011), vol. 492, pp. 27–59.

3. L. M. Stevers, P. J. de Vink, C. Ottmann, J. Huskens, L. Brunsveld, A Thermodynamic Model for Multivalency in 14-3-3 Protein–Protein Interactions. *J Am Chem Soc*. **140**, 14498–14510 (2018).
4. P. Sartori, S. Leibler, Lessons from equilibrium statistical physics regarding the assembly of protein complexes. *Proc. Nat. Acad. Sci.* **117**, 114–120 (2020).
5. S.-H. Chong, S. Ham, Dynamics of Hydration Water Plays a Key Role in Determining the Binding Thermodynamics of Protein Complexes. *Sci Rep*. **7**, 8744 (2017).
6. X. Wei, X. Su, P. Cao, X. Liu, W. Chang, M. Li, X. Zhang, Z. Liu, Structure of spinach photosystem II–LHCII supercomplex at 3.2 Å resolution. *Nature*. **534**, 69 (2016).
7. M. P. Johnson, T. K. Goral, C. D. P. Duffy, A. P. R. Brain, C. W. Mullineaux, A. V Ruban, Photoprotective energy dissipation involves the reorganization of photosystem II light-harvesting complexes in the grana membranes of spinach chloroplasts. *Plant Cell*. **23**, 1468–1479 (2011).
8. R. Bassi, L. Dall'Osto, Dissipation of light energy absorbed in excess: the molecular mechanisms. *Annu Rev Plant Biol*. **72**, 47–76 (2021).
9. A. V Ruban, Nonphotochemical Chlorophyll Fluorescence Quenching: Mechanism and Effectiveness in Protecting Plants from Photodamage. *Plant Physiol*. **170**, 1903–1916 (2016).
10. V. Giovagnetti, A. V Ruban, The evolution of the photoprotective antenna proteins in oxygenic photosynthetic eukaryotes. *Biochem Soc Trans*. **46**, 1263–1277 (2018).
11. P. Muller, X.-P. Li, K. K. Niyogi, Non-photochemical quenching. A response to excess light energy. *Plant Physiol*. **125**, 1558–1566 (2001).
12. L. Nicol, W. J. Nawrocki, R. Croce, Disentangling the sites of non-photochemical quenching in vascular plants. *Nat Plants*. **5**, 1177–1183 (2019).
13. P. Manna, G. S. Schlau-Cohen, Photoprotective conformational dynamics of photosynthetic light-harvesting proteins. *Biochim Biophys Acta Bioenerg*. **1863**, 148543 (2022).
14. B. van Oort, A. van Hoek, A. V Ruban, H. van Amerongen, Equilibrium between quenched and nonquenched conformations of the major plant light-harvesting complex studied with high-pressure time-resolved fluorescence. *J Phys Chem B*. **111**, 7631–7637 (2007).
15. S. Santabarbara, P. Horton, A. V Ruban, Comparison of the thermodynamic landscapes of unfolding and formation of the energy dissipative state in the isolated light harvesting complex II. *Biophys J*. **97**, 1188–1197 (2009).
16. M. Son, R. Moya, A. Pinnola, R. Bassi, G. S. Schlau-Cohen, Protein–Protein Interactions Induce pH-Dependent and Zeaxanthin-Independent Photoprotection in the Plant Light-Harvesting Complex, LHCII. *J Am Chem Soc*. **143**, 17577–17586 (2021).
17. L. Wilk, M. Grunwald, P.-N. Liao, P. J. Walla, W. Kühlbrandt, Direct interaction of the major light-harvesting complex II and PsbS in nonphotochemical quenching. *Proc. Natl. Acad. Sci.* **110**, 5452–5456 (2013).
18. A. Natali, J. M. Gruber, L. Dietzel, M. C. A. Stuart, R. van Grondelle, R. Croce, Light-harvesting complexes (LHCs) cluster spontaneously in membrane environment leading to shortening of their excited state lifetimes. *J. Biol. Chem.* **291**, 16730–16739 (2016).
19. M. Tutkus, J. Chmeliov, G. Trinkunas, P. Akhtar, P. H. Lambrev, L. Valkunas, Aggregation-related quenching of LHCII fluorescence in liposomes revealed by single-molecule spectroscopy. *J Photochem Photobiol B*. **218**, 112174 (2021).

20. M. Tutkus, P. Akhtar, J. Chmeliov, F. Görföl, G. Trinkunas, P. H. Lambrev, L. Valkunas, Fluorescence Microscopy of Single Liposomes with Incorporated Pigment–Proteins. *Langmuir*. **34**, 14410–14418 (2018).
21. P. Akhtar, F. Görföl, G. Garab, P. H. Lambrev, Dependence of chlorophyll fluorescence quenching on the lipid-to-protein ratio in reconstituted light-harvesting complex II membranes containing lipid labels. *Chem Phys*. **522**, 242–248 (2019).
22. M. K. Shukla, A. Watanabe, S. Wilson, V. Giovagnetti, E. I. Moustafa, J. Minagawa, A. V Ruban, A novel method produces native light-harvesting complex II aggregates from the photosynthetic membrane revealing their role in nonphotochemical quenching. *J. Biol. Chem.* **295**, 17816–17826 (2020).
23. S. Wilson, D.-H. Li, A. V Ruban, The Structural and Spectral Features of Light-Harvesting Complex II Proteoliposomes Mimic Those of Native Thylakoid Membranes. *J Phys Chem Lett*. **13**, 5683–5691 (2022).
24. Z. Liu, H. Yan, K. Wang, T. Kuang, J. Zhang, L. Gui, X. An, W. Chang, Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution. *Nature*. **428**, 287–292 (2004).
25. P. Manna, T. Davies, M. Hoffmann, M. P. Johnson, G. S. Schlau-Cohen, Membrane-dependent heterogeneity of LHCII characterized using single-molecule spectroscopy. *Biophys J*. **120**, 3091–3102 (2021).
26. Q. Wang, W. E. Moerner, Dissecting pigment architecture of individual photosynthetic antenna complexes in solution. *Proc. Nat. Acad. Sci.* . **112**, 13880–13885 (2015).
27. J. M. Gruber, J. Chmeliov, T. P. J. Krüger, L. Valkunas, R. Van Grondelle, Singlet–triplet annihilation in single LHCII complexes. *Phys. Chem. Chem. Phys.* **17**, 19844–19853 (2015).
28. O. C. Fiebig, D. Harris, D. Wang, M. P. Hoffmann, G. S. Schlau-Cohen, Ultrafast Dynamics of Photosynthetic Light Harvesting: Strategies for Acclimation Across Organisms. *Annu. Rev. Phys. Chem.* **74**, 493–520 (2023).
29. A. R. Schneider, P. L. Geissler, Coexistence of fluid and crystalline phases of proteins in photosynthetic membranes. *Biophys J*. **105**, 1161–1170 (2013).
30. X. Su, J. Ma, X. Wei, P. Cao, D. Zhu, W. Chang, Z. Liu, X. Zhang, M. Li, Structure and assembly mechanism of plant C2S2M2-type PSII-LHCII supercomplex. *Science (1979)*. **357**, 815–820 (2017).
31. L. Shen, Z. Huang, S. Chang, W. Wang, J. Wang, T. Kuang, G. Han, J.-R. Shen, X. Zhang, Structure of a $C_2S_2M_2N_2$-type PSII–LHCII supercomplex from the green alga *Chlamydomonas reinhardtii*. *Proc. Nat. Acad. Sci.* **116**, 21246–21255 (2019).
32. S. Puthiyaveetil, B. Van Oort, H. Kirchhoff, Surface charge dynamics in photosynthetic membranes and the structural consequences. *Nat Plants*. **3**, 1–9 (2017).
33. J. Barber, Influence of surface charges on thylakoid structure and function. *Annu Rev Plant Physiol*. **33**, 261–295 (1982).
34. W. S. Chow, E.-H. Kim, P. Horton, J. M. Anderson, Granal stacking of thylakoid membranes in higher plant chloroplasts: the physicochemical forces at work and the functional consequences that ensue. *Photochemical & Photobiological Sciences*. **4**, 1081–1090 (2005).
35. M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, J. H. Jensen, PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput*. **7**, 525–537 (2011).

36. E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D. W. Gohara, T. Dolinsky, R. Konecny, D. R. Koes, J. E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G.-W. Wei, M. J. Holst, J. A. McCammon, N. A. Baker, Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 (2018).
37. R. Mao, H. Zhang, L. Bie, L.-N. Liu, J. Gao, Million-atom molecular dynamics simulations reveal the interfacial interactions and assembly of plant PSII-LHCII supercomplex. *RSC Adv*. **13**, 6699–6712 (2023).
38. S. Bellafiore, F. Barneche, G. Peltier, J.-D. Rochaix, State transitions and light adaptation require chloroplast thylakoid protein kinase STN7. *Nature*. **433**, 892–895 (2005).
39. W. H. J. Wood, M. P. Johnson, Modeling the role of LHCII-LHCII, PSII-LHCII, and PSI-LHCII interactions in state transitions. *Biophys J*. **119**, 287–299 (2020).
40. F. Saccon, V. Giovagnetti, M. K. Shukla, A. V Ruban, Rapid regulation of photosynthetic light harvesting in the absence of minor antenna and reaction centre complexes. *J Exp Bot*. **71**, 3626–3637 (2020).
41. A. V Ruban, S. Wilson, The mechanism of non-photochemical quenching in plants: localization and driving forces. *Plant Cell Physiol*. **62**, 1063–1072 (2021).
42. L. Nicol, R. Croce, The PsbS protein and low pH are necessary and sufficient to induce quenching in the light-harvesting complex of plants LHCII. *Sci Rep*. **11**, 7415 (2021).
43. H. Jia, J. R. Liggins, W. S. Chow, Entropy and biological systems: Experimentally-investigated entropy-driven stacking of plant photosynthetic membranes. *Sci Rep*. **4**, 1–7 (2014).
44. N. Liguori, R. Croce, S. J. Marrink, S. Thallmair, Molecular dynamics simulations in photosynthesis. *Photosynth Res*. **144**, 273–295 (2020).
45. E. E. Ostroumov, Y. R. Khan, G. D. Scholes, others, "Photophysics of Photosynthetic Pigment-Protein Complexes" in *Non-Photochemical Quenching and Energy Dissipation in Plants, Algae and Cyanobacteria* (Springer, 2014), pp. 97–128.
46. L. Valkunas, G. Trinkunas, V. Liuolia, R. Van Grondelle, Nonlinear annihilation of excitations in photosynthetic systems. *Biophys J*. **69**, 1117–1129 (1995).
47. Y. Zaushitsyn, K. G. Jespersen, L. Valkunas, V. Sundström, A. Yartsev, Ultrafast dynamics of singlet-singlet and singlet-triplet exciton annihilation in poly (3-2'-methoxy-5'octylphenyl) thiophene films. *Phys Rev B*. **75**, 195201 (2007).
48. V. Barzda, C. J. de Grauw, J. Vroom, F. J. Kleima, R. van Grondelle, H. van Amerongen, H. C. Gerritsen, Fluorescence lifetime heterogeneity in aggregates of LHCII revealed by time-resolved microscopy. *Biophys J*. **81**, 538–546 (2001).
49. P. H. Lambrev, Z. Várkonyi, S. Krumova, L. Kovács, Y. Miloslavina, A. R. Holzwarth, G. Garab, Importance of trimer–trimer interactions for the native state of the plant light-harvesting complex II. *Biochim Biophys Acta Bioenerg.* **1767**, 847–853 (2007).
50. S. K. Sinha, "Classical statistical mechanics of interacting system" in *Introduction to statistical mechanics* (Alpha Science Int'l Ltd., 2005), pp. 204–239.
51. A. R. Schneider, P. L. Geissler, Coarse-grained computer simulation of dynamics in thylakoid membranes: methods and opportunities. *Front Plant Sci*. **4**, 555 (2014).
52. E. Crisafi, A. Pandit, Disentangling protein and lipid interactions that control a molecular switch in photosynthetic light harvesting. *Biochim Biophys Acta Biomembr.* **1859**, 40–47 (2017).

53. V. Daskalakis, S. Papadatos, U. Kleinekathoefer, Fine tuning of the photosystem II major antenna mobility within the thylakoid membrane of higher plants. *Biochim Biophys Acta Biomembr.* **1861**, 183059 (2019).
54. P. G. Saffman, M. Delbrück, Brownian motion in biological membranes. *Proc. Nat. Acad. Sci.* **72**, 3111–3113 (1975).
55. H. Kirchhoff, Diffusion of molecules and macromolecules in thylakoid membranes. *Biochim Biophys Acta Bioenerg.* **1837**, 495–502 (2014).
56. T. Kondo, A. Pinnola, W. J. Chen, L. Dall'Osto, R. Bassi, G. S. Schlau-Cohen, Single-molecule spectroscopy of LHCSR1 protein dynamics identifies two distinct states responsible for multi-timescale photosynthetic photoprotection. *Nat Chem*. **9**, 772 (2017).
57. L. P. Watkins, H. Yang, Detection of intensity change points in time-resolved single-molecule measurements. *J Phys Chem B*. **109**, 617–628 (2005).
58. M. Maus, M. Cotlet, J. Hofkens, T. Gensch, F. C. De Schryver, J. Schaffer, C. A. M. Seidel, An experimental comparison of the maximum likelihood estimation and nonlinear least-squares fluorescence lifetime analysis of single molecules. *Anal Chem*. **73**, 2078–2086 (2001).
59. F. Pedregosa, G. Varoquaux, A. Gramfort, B. Michel V. and Thirion, O. Grisel, M. Blondel, R. Prettenhofer P. and Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* . **12**, 2825–2830 (2011).

Acknowledgments

Funding: This work is supported by the Human Frontiers Science Program (Award #RGY0076 to MPJ and GSSC). PM, MH, and GSSC also acknowledge support from the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences under Award # DE-SC0018097 to GSSC. GSSC thanks the CIFAR Bio-inspired Solar Energy program for support.

Author contributions:
Funding Acquisition: GSSC, MPJ
Conceptualization: PM, GSSC
Investigation: PM, MH, TD, KHR
Analysis: PM, MH
Supervision: GSSC, MPJ
Writing—original draft: PM, GSSC
Writing—review & editing: PM, GSSC, MH, MPJ, TD, KHR

Competing interests: The authors declare no competing interests.

Data and materials availability: All data needed to evaluate the conclusions in this article are available in the main text and/or the supplementary materials. Data available at 10.5281/zenodo.10050211.

Supplementary Materials

This PDF file includes

Supplementary Text
Figs. S1 to S36
Tables S1 to S9

Science Advances
AAAS

Supplementary Materials for

Energetic driving force for LHCII clustering in plant membranes

Premashis Manna *et al.*

*Corresponding author. Email: pmanna@mit.edu, gssc@mit.edu

This PDF file includes:

Supplementary Text S1 to S18
Figs. S1 to S37
Tables S1 to S9

Supplementary Text

S1. Purification of LHCII

LHCII trimers used in this study were purified from spinach. The details of the LHCII purification and the characterizations are provided in (*25*). Briefly, fresh market spinach leaves were dark-adapted at 4°C overnight then blended in ice-cold grinding medium (330 mM Sorbitol, 10 mM $Na_4P_2O_7$, 5 mM $MgCl_2$, 2 mM D(+) iso-ascorbate, pH 6.5). The homogenate was filtered once through two layers of muslin and one layer of cotton wool. The filtrate was centrifuged at 4000 x g for 20 minutes and the pellet resuspended in wash medium (330 mM Sorbitol, 10 mM MES, pH 6.5) followed by centrifugation at 4000 x g for 20 minutes. The pellet was resuspended in resuspension medium (330 mM Sorbitol, 40 mM MES, 5 mM $MgCl_2$, pH 6.5). Break medium (5 mM $MgCl_2$, 40 mM MES, pH 7.6) was added to triple the original volume to lyse any unbroken chloroplasts. The osmotic potential was restored after 45 seconds with the addition of an equal volume of osmoticum medium (660 mM Sorbitol, 40 mM MES, 5 mM $MgCl_2$, pH 6.5). The sample was then centrifuged at 4000 x g for 20 minutes and the thylakoid membranes were then resuspended in a small volume of storage medium (300 mM Sucrose, 10 mM HEPES, 5 mM $MgCl_2$, pH 7.5). Thylakoids were solubilized at a concentration of 1 mg/ml chlorophyll in 0.55% n-Dodecyl-β-D-Maltopyranoside (βDDM) in the dark on ice for 10 minutes, with occasional mixing. Partially solubilized thylakoids were centrifuged at 1000 x g for 10 minutes to pellet any insoluble material. The supernatant was removed and centrifuged 38,000 x g for 40 minutes to pellet Photosystem II membrane fragments (known as BBYs). BBYs with a final chlorophyll concentration of 0.5 mg/ml were solubilized in 10mM HEPES (pH 7.5), 0.5% n-Hexadecyl-β-D-Maltopyranoside (HDM), 0.1 % β-DDM for 30 minutes at room temperature, with occasional mixing. The solubilized BBYs were centrifuged at 22,000 x g for 10 minutes and the supernatant was retained. LHCII was purified from solubilized BBYs via sucrose density gradient centrifugation. Solubilized BBYs were loaded onto a continuous 650 mM freeze/thaw sucrose gradient in a buffer of 20 mM HEPES, pH 7.5, 0.06% (w/v) glyco-diosgenin (GDN) in SW32 Ti rotor tubes (Beckman Coulter, Inc.). Fractionation of protein complexes was performed by centrifugation at 175,000 x g in an SW32 Ti swinging bucket rotor (Beckman Coulter, Inc.) for 28-32 hours at 4°C. After centrifugation, trimeric LHCII was harvested using a peristaltic pump. Finally, LHCII samples were flash-frozen in liquid nitrogen and stored at -70°C.

S2. Preparation of LHCII-proteoliposomes

For the preparation of liposomes, a thylakoid lipid mixture containing 50% monogalactosyldiacylglycerol (MGDG), 30% digalactosyldiacylglycerol (DGDG), 10% sodium salt of L-α-phosphatidylglycerol (Soy PG) and 10% sulfoquinovosyldiacylglycerol (SQDG) was dissolved in 1 ml of 7:3 chloroform: MeOH. Biotinylated lipids (sodium salt of 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine-N-(cap biotinyl), 18:1 Biotinyl Cap PE, Avanti Polar Lipids, Inc.) were added to the mixture for a final ratio of 1:250 (biotinylated lipid: thylakoid lipid). The lipids were completely dried down using N_2 gas and transferred to a vacuum

desiccator for 2-3 hrs. The lipid films prepared this way are stored at -70°C and used subsequently as needed. The dried lipid mix was resuspended in a working buffer (20 mM HEPES, 40 mM NaCl, pH 7.5) by vortexing vigorously. The lipid mixture was subjected to 8 freeze/thaw cycles to break down large multilamellar vesicles. The vesicles were then extruded through a polycarbonate membrane with 50 nm (diameter) pore ~21-times using an extrusion kit (Avanti Polar Lipids, Inc.) to generate unilamellar vesicles with uniform sizes. 0.03% β-DDM was gently mixed with the sample to destabilize the unilamellar liposomes and the mixture was left for 30 minutes with occasional gentle mixing. The desired amount of LHCII was added to the sample containing empty liposomes to obtain a lipid:protein molar ratio of 7000, 5000, 2500, 1000, 880, and 510 with the average number of proteins per liposome (<N>) <1, 1, 2, 5, 6 and 10, respectively. This number is based on a ~ 20-25% incorporation of LHCII in proteoliposome as shown here (*20*). The reaction mixture was left for 1 hr. in the dark with occasional gentle mixing. Bio-beads (Biobeads SM-2 adsorbent media, Bio-Rad Laboratories, Inc.) were added in a step-wise manner into the solution for 2-3 hrs. to extract detergent molecules from the solution.

A 5-step sucrose gradient (15%, 20%, 25%, 30%, and 35%) was prepared in SW41 ultracentrifuge tubes (Beckman Coulter, Inc.) using a peristaltic pump. LHCII-proteoliposomes were loaded onto the sucrose gradients which were then centrifuged at 150,000 x g for 14 hrs. to remove liposome-free LHCII aggregates. After centrifugation, the green band corresponding to LHCII-proteoliposomes was collected from the gradient using a peristaltic pump and subjected to further ensemble characterization and single-molecule measurements within one week.

S3. Characterization of LHCII proteoliposomes

Dynamic light scattering

For DLS measurements, the samples were filtered to remove large aggregates. DynaPro Nanostar (Wyatt Technologies) was used for DLS measurements with illumination at 658 nm. A commercial software (Dynamics 7.1.9.3) was used for the data analysis in DLS to report the intensity distribution of the scattered particles. The diameters of the proteoliposome samples are displayed in Table S1.

Absorption and Emission

Absorption and emission spectra of the LHCII samples were collected using an Epoch Microplate Spectrophotometer (BioTek) and a Cary Eclipse Fluorescence Spectrophotometer, respectively. The samples were syringe filtered (GE Healthcare Life Sciences, pore size 0.22 μm) to discard any aggregates before the measurements.

Excited state lifetime

The fluorescence decays were measured with a TCSPC module (Time Tagger 20, Swabian Instruments). For excitation, a tunable fiber laser (FemtoFiber pro, Toptica Photonics, 80 MHz repetition rate, 130 fs pulse duration, 610 nm, 4 nm full-width half maximum (fwhm)), passed through a pinhole and directed into a home-built confocal microscope. The excitation was focused by an oil-immersion objective (UPLSAPO100XO, Olympus, NA 1.4) onto the sample placed on a coverslip. The emission of the sample was collected through the same objective and separated

from the excitation using a dichroic (ZT647rdc, Chroma) and bandpass filters (ET700/75m, Chroma and ET690/120x, Chroma). The fluorescence decay was fit by iterative re-convolution with a bi-exponential function using the measured instrument response function (IRF) of the system with a home-built MATLAB code. The IRF was measured by the scattered signal to be ~ 400 ps (fwhm). The average lifetime values were obtained through an intensity-weighted average of the fitted bi-exponential lifetime constants.

S4. Trypsin Digest

Trypsin is an enzyme that cleaves arginine and lysine residues in a protein sample. In LHCII, trypsin selectively cleaves the N-terminus of the protein (*52*). If LHCII inserts directionally into a liposome, then the LHCII should either be all uncleaved or totally cleaved after trypsin digestion. If it inserts randomly, we should see a mixture of cleavage products. Previous studies of LHCII in liposomes have reported mixed results (*42*, *52*). We performed trypsin digest experiments on the LHCII proteoliposomes used in this work to confirm orientation.

In brief, LHCII proteoliposome samples with <N> = 5 were mixed with trypsin in a 5:1 mass ratio and heated on a Thermomixer at 37◦C for up to 2 hours. SDS-PAGE was used to visualize the reaction products. Standard protocols were used with the modification of not adding 2-mercaptoethanol prior to the heating step. 12% Mini-PROTEAN® TGX gels from Bio-Rad were found to have the best separation of the cleavage products.

S5. Lifetime Measurements at Low pH

Protein incorporation efficiency in liposomes depends on the pH of the solvent and therefore liposomes prepared with the same protein-to-lipid ratio but at different pH differ in protein density. In order to disentangle the pH-induced and density-induced effects in the quenching of LHCII, it is paramount that the protein density in the liposome remains the same. Therefore, to overcome the uncertainty of protein density at neutral and low pH, we incubated the proteoliposome samples at pH 5 buffer rather than preparing the liposome at this low pH. We monitored the change in excited state lifetimes of the proteoliposomes over time to identify an equilibrium point (Figure S11). We also verified that the proteoliposome samples remain intact by performing dynamic light scattering (DLS) at different phases of the incubation process (Figure S10). It was found that the lifetime of LHCII-proteoliposome samples reached equilibrium after ~10 hrs. of incubation and remained stable in the course of single-molecule data collection (S11 b,c).

As discussed in the main text, the enhanced quenching at low pH is driven by increased clustering of LHCII proteins which require diffusion of the proteins on the liposome surface. However, the mobility of LHCII protein becomes up to ~10 fold slower at acidic pH as shown by (*53*). Also, as the size of the clusters gets bigger, the lateral diffusion of the aggregates becomes even slower restricting their mobility. This is explained by Saffman-Delbruck equation, $D \propto ln(1/R)$, with D and R as the diffusion constant and the radius of the protein assembly (*54*). Once these factors are taken into account, the longer time (> 10 hrs.) required by the system to reach an equilibrium is justified. A more quantitative description of the kinetics of cluster formation in liposomes is given in the following section.

S6. Kinetics of Cluster Formation in Liposomes

Let's assume there are five proteins on the liposome surface with a radius *R*. What is the rate of conversion from configuration A to configuration B as shown in Figure S12? The lateral diffusion coefficient of LHCII on the lipid surface is *D*. Considering a Brownian motion of LHCII on a liposome which is true in a protein-poor system like this, at a certain time period of Δt, the mean area the proteins travel is $4D\Delta t$. Therefore, the fraction of the surface covered by the protein (*f*) at this time interval is,

$$f = \frac{4D\Delta t}{A} = \frac{4D\Delta t}{4\pi R^2} = \frac{D\Delta t}{\pi R^2} \quad (S1)$$

where *A* is the surface area of the liposome.

Now, the probability that within this random walk, the single LHCII trimer comes close to the other cluster of four proteins and quenching happens is:

$$\frac{D\Delta t}{\pi R^2} * p \quad (S2)$$

where *p* is the encounter probability. Here, we assume that the diffusion coefficient of the bigger cluster is slow and therefore it remains relatively immobile compared to the single LHCII.

During this quenching event, the concentration of unquenched configuration (*C*), decreased by an amount of Δ*C*. Therefore,

$$\frac{\Delta C}{C} = -\frac{D\Delta t}{\pi R^2} * p \quad (S3)$$

At the limit of $\Delta t \to 0$, the above equation becomes,

$$\frac{dC}{C} = -\frac{D * p}{\pi R^2} dt \quad (S4)$$

Integration of the above equation gives,

$$C = a * exp\left(-\frac{t}{\tau_c}\right) + b. \quad (S5)$$

where, $\tau_c = \frac{\pi R^2}{D*p}$ is the time constant of cluster formation. a and b are constants.

Figure S13, shows the drop in lifetime for <N> sample as a function of incubation time obtained experimentally. A single-exponential fitting of the trace gives a τ_c of 7.1 hrs. Again, the encounter probability for two proteins on the liposome size of 25 nm is given as p = 0.006 (ref. analytical solution of cluster formation, Sec 12). The above results give us a diffusion time constant of LHCII under low pH:

$$D = \frac{\pi R^2}{\tau_c * p} = 1.3 \times 10^{-17} m^2/s \quad (S6)$$

The above results assume that the drop in lifetime for <N> = 5 sample is largely from the clustering of configuration A to B (Figure S12). This is approximately true as considering the LHCII-LHCII interaction energy at neutral and low pH as -5.4 and -7.2 k_BT, the population of configuration B changes from 23% to 74%. For more rigorous estimation, the contributions from all other configurations and their quenching events need to be accounted for. It is notable that the diffusion constant obtained here ~ is 100-fold slower than that reported from FRAP measurements (*55*). This could be due to the slow mobility of LHCII at low pH due to hydrophobic mismatch as shown here by MD simulation (*53*).

Using the above analysis, it is straightforward to show that with bigger liposomes, the time constant for cluster formation is very slow. For instance, with a liposome of radius 50 nm (R=50 nm) and <N> = 5,

$$\tau_c = \frac{\pi R^2}{D * p} = 129hrs \qquad (S7)$$

Here, the encounter probability, p, is used as 0.0013 (ref. analytical solution, Sec 12).

For proteoliposomes with higher protein density, the analysis of cluster kinetics is complex because it involves multiple configurations. For instance, in proteoliposomes with <N>=10, the 42 different configurations get clustered at different extents upon low pH incubation. However, only accounting for the conversion of the configuration of 9 clustered and one free LHCII to a configuration where all 10 LHCII are clustered, the time constant of cluster formation is estimated as ~ 50 hrs. Therefore, to observe the steady state drop in the lifetime of the <N>=10 samples upon low pH incubation, its needs to be incubated at least 3 × 50 = 150 hrs. considering exponential kinetics. However, LHCII-proteoliposomes are not stable for that many hours particularly at low pH as revealed by a reduction in their size observed in dynamic light scattering. Therefore, the low pH measurements of proteoliposomes with protein densities higher than five could not be measured reliably in this work.

S7. Single-molecule Measurements

The protocol for immobilization of LHCII-proteoliposome samples on biotinylated coverslip has been described in the main text.

For the single-molecule measurement of immobilized samples, the excitation was generated by a tunable fiber laser (FemtoFiber pro, Toptica Photonics, 80 MHz repetition rate, 130 fs pulse duration, 610 nm, 4 nm full-width half maximum (fwhm)), passed through a pinhole, and directed into a home-built confocal microscope. The excitation was focused by an oil-immersion objective (UPLSAPO100XO, Olympus, NA 1.4) onto the samples immobilized on a coverslip. The coverslip was mounted on a piezostage controlled by a home-written Labview-based software. The stage was used to raster-scan on a 5µm x 5µm area to detect single particles (Figure S14).

The concentration of the sample was diluted to ~ 4-6 LHCII-proteoliposomes per 25 µm^2 scanned area. Subsequently, the laser was parked on the detected spots. The emission of the

sample was collected through the same objective and separated from the excitation using a dichroic (ZT647rdc, Chroma) and bandpass filters (ET700/75m, Chroma and ET690/120x, Chroma). The laser spot size was 280 nm - 380 nm (fwhm) (Figure S8 b). The average power and intensity of excitation used for the single-molecule measurements were 350 nW and 5250 nJ/cm^2 per pulse, respectively. Emission was detected by a silicon-based single photon counting avalanche photodiode (SPCMAQRH, Excelitas Technologies). A time-correlated single-photon counting (TCSPC) module (Time Tagger 20, Swabian Instruments) was used to record the macrotime and microtime for each detected photon (*56*). The instrument response function (IRF) measured from the scattered signal was ~ 400 ps (fwhm). Fluorescent traces from >100 complexes were recorded for each set of measurements.

S8. Single-molecule Data Analysis

Lifetime fits by maximum likelihood estimation (MLE). The TCSPC module recorded the absolute arrival time of the photons to the detector (macrotime) and the time relative to a trigger pulse (microtime). The macrotime was binned at 100 ms resolution to generate fluorescence intensity traces. Regions of constant intensity (states) were identified by a change-point algorithm (*57*). The threshold for the change-point algorithm was adjusted manually based on the inspection of fluorescence traces of the individual complexes. The microtimes of the photons from each state were binned at 280 ps resolution and the resultant histogram was fit with a single-exponential function convolved with the IRF to obtain excited state lifetimes. States with fewer than 500 photons were removed from the analysis as reliable fitting of a lifetime could not be obtained. States with a lifetime below 0.5 ns and above 5 ns were filtered due to being near the IRF or unphysiological. Due to the low signal to-noise (S/N) ratio encountered in single-molecule measurements, the decay curves were fit by an algorithm based on maximum likelihood estimation (MLE). MLE captures the inherent Poissonian noise of the single-molecule photon detection and provides robust fitting of the data compared to the conventional non-linear least square method in the low S/N regime (*58*). The details of the fitting are provided in (*25*).

S9. Density- and pH-induced lifetime drop in LHCII proteoliposomes

In recent years, the aggregation of LHCII has been explored in model membranes including liposomes and nanodiscs (*16*, *18*, *19*, *21*). It was shown that with the increase of protein density in LHCII proteoliposomes, the excited state lifetime of the LHCII complexes shortened from >3 ns (in detergent) to less than 1 ns. This establishes that LHCII proteins have a very strong tendency to aggregate even in the absence of a pH gradient. Here, we also observed similar protein density dependent quenching of LHCII in the liposome as described in the literature (*18*, *19*, *21*). We define the density-induced lifetime drop ($\Delta\tau_{density}$) in the following way:

$$\Delta\tau_{density} = \frac{\tau_N - \tau_1}{\tau_1} \times 100 \qquad (S8)$$

where, τ_N and τ_1 are the lifetimes of proteoliposome with average protein densities of N and 1, respectively. We observe a ~ 30% density-induced drop in ensemble lifetime once the average proteins in liposome are increased from less than one to 10 (Figure 1d main text, Table S2). A

similar trend is observed in the lifetimes obtained from single-molecule measurements (Table S3).

It has been shown in the intact chloroplast (*7*) and in the model membranes (*16*, *42*) that lowering the pH enhances fluorescence quenching. Nicol et al. (*42*) reported that in the LHCII-packed liposomes (lipid:protein = 100:1 or 131:1), lowering the pH from 7.5 to 5.5 shortens the lifetime by 20-30%. Here, we define the pH-induced lifetime drop ($\Delta\tau_{pH}$) as the following way:

$$\Delta\tau_{pH} = \frac{\tau_{N,7.5} - \tau_{N,5.0}}{\tau_{N,7.5}} \times 100 \qquad (S9)$$

where, $\tau_{N,7.5}$ and $\tau_{N,5.0}$ are the lifetimes of the LHCII proteoliposome with average protein densities of N at pH 7.5 and 5.0, respectively.

S10. Clustering of Interacting Proteins

When a certain number of proteins with pairwise attractive potential are allowed to interact with each other by restricting them to a finite surface area they tend to form clusters. The driving force of this cluster formation is the reduction of free energy. Based on the mode of clustering, i.e., how the proteins in the cluster are connected, we can classify different types of configurations from these interacting proteins. The number of possible type of configurations for N number of proteins can be obtained by solving the equation below (*50*):

$$\sum_{n=1}^{N} n * m_n = N \qquad (S10)$$

where, the system forms a cluster with m_1 number of unit cluster, m_2 number of clusters of 2 proteins, ... m_n number of cluster of l proteins and so on.

For instance, with $N = 5$, there are seven solutions that satisfy the above equation (Eqn. S10). In other words, seven different configurations are possible as shown in Figure S16a. These seven solutions of the equation S10 are displayed in Table S4. The first solution ([0 0 0 0 1]) represents the configuration where all the proteins form a cluster while the last solution ([5 0 0 0 0]) represents where all the proteins remain unclustered. The rest of the five configurations consist of various numbers of clustered and free proteins.

As the number of proteins increases, the possible number of configurations also increases. For example, with 10 proteins, 42 configurations are possible. Figure S16b displays the number of configurations as a function of the number of proteins. The number of possible configurations is obtained by solving Eqn. S10.

In this work, we prepared LHCII-proteoliposomes with various lipid-to-protein ratios that generate liposomes with an average number (<N>) of proteins up to 10. For a given number of LHCII, the equilibrium population of a certain configuration (p) arising from the clustering depends on both the probability of forming the configuration (w) and the enthalpy (ΔH) weighted by the Boltzmann factor. This can be expressed as follows:

$$p = w * e^{-\Delta H/k_B T} \quad (S11)$$

where, k_B and T are Boltzmann constant and temperature, respectively. In the following section, we discuss the way we compute the equilibrium population (p) of these configurations. The populations are then exploited to extract the pairwise interaction energies (J) of LHCII proteins.

S11. Determination of the equilibrium population of cluster configurations from simulation

To determine the probability of forming different configurations (w), we use Monte-Carlo simulations to randomly distribute the proteins on the liposome surface. A homemade Matlab (Mathworks, Inc.) code was used for this. The liposome is modeled as a spherical surface with a radius, R, of 25 nm. Proteins (LHCII) are randomly assigned x, y, and z coordinates between 0 and 1. These coordinates, C, are then projected onto the surface of this sphere using the equation:

$$C' = \frac{R.C}{\sqrt{C.C}} \quad (S12)$$

The distance, d, between proteins is then calculated by:

$$\theta = cos^{-1}\left(\frac{C_1'.C_2'}{R^2}\right) \quad (S13)$$

$$d = R\theta \quad (S14)$$

where θ is the angle between the two proteins.

If any two proteins are separated by a distance of less than 6.5 nm (size of LHCII trimers), the configuration is automatically thrown out and a new random configuration is chosen. This is to avoid any non-physical overlap of two proteins on the liposome surface. Any proteins separated by a distance of 6.5-7.5 nm are assumed to be in the same cluster and able to interact. The number of clusters and the number of proteins in each cluster determines the configuration. Each configuration has an assigned enthalpy calculated by:

$$H = \sum_{i=1}^{N} J * (n_i - 1) * m_i \quad (S15)$$

Where J is the pairwise interaction energy, n_i is the number of proteins in the cluster of size i, m_i is the number of clusters of size i, and N is total the number of proteins in the liposome.

Table S5 displays the probability of different configurations with the number of LHCII in liposomes from 2 to 6. The simulations are typically run for $10^7 - 10^9$ number of iterations until the population of various configurations converges. All of the possible configurations are recovered with these iterations for N up to 5. However, for the systems with molecules beyond 6, all configurations are not observed even after running the iterations up to 10^9 times. As it is displayed in Table S5, for $N = 6$, the configuration where all the LHCII molecules clustered (configuration a) is very unlikely just considering the w factor. However, to properly account for the equilibrium population of the configurations, probabilities of all the configurations are

necessary. This is because some configurations might have a low probability of occurrence just considering the entropy factor but could be possible when the enthalpy factor is also accounted for. Therefore, we solved the probability of the cluster configurations via an analytical approach as described below.

S12. Analytical solution to the probability of cluster formation on liposome

Let us consider a liposome with a radius of R. The total surface area of the liposome is $4\pi R^2$. Let us also assume that there is a protein molecule with radius r embedded on this liposome surface. Now, what is the probability that if we insert another protein, those two will form a cluster? For this, we need to define the cluster. In this work, we define that if the second molecule is within a distance of d away, we consider those two proteins to form a cluster. Now, as the second protein cannot occupy the space of the first protein, the probability of forming this cluster is proportional to $\pi(r+d)^2 - \pi r^2$ (the yellow shaded region in Figure S17). Therefore, the probability of forming a cluster of two proteins (p_2) is given by,

$$p_2 = \frac{\pi(r+d)^2 - \pi r^2}{4\pi R^2 - \pi r^2} \times 2 \qquad (S16)$$

The denominator in the above equation reflects the total surface area available for the second molecule to be on the liposome surface. The factor of 2 accounts for the exchangeability of these two identical proteins. With R = 25 nm, r = 3.25 nm, and d = 1 nm; the above equation predicts the probability of forming a cluster of two LHCII on the liposome surface is 0.006. This exactly matches the result obtained from the Monte Carlo simulation (Table S5).

The probability of higher-order cluster formation can be calculated in a similar way. However, for higher-order clusters, different modes of arrangement need to be considered. For example, the probability of cluster formation of three proteins (p_3) is proportional to:

$$p_3 = p_2 * p_3' \qquad (S17)$$

Where, p_2 is the probability of cluster formation of two proteins and p_3' is the probability that the third proteins find this cluster of two proteins on the liposome surface. As similar to discussed above, p_3' is proportional to the yellow shaded regions in Figure 2. Now, when the third molecule is added to the liposome, the cluster of two molecules can be in two extreme arrangements: 1) they can touch each other (Figure S18, left) or 2) they can be in d distance away (as per the definition of a cluster, Figure S18, right). Let us call them compact and non-compact arrangements, respectively. In reality, most of the time, these two proteins in a cluster can be found in between compact and non-compact arrangements.

Figure S19 a shows the area available for the compact arrangement of clusters having N=2-5 proteins. The probability that when another protein added to the liposome would form a cluster is proportional to the yellow-shaded region. This available interaction area for a compact arrangement is the minimum available area. This minimum available area for a number of clusters is computed using a commercial software (https://www.sketchandcalc.com/). The maximum available area for the non-compact arrangements for a cluster of N particles is $N * \pi((r+d)^2 - r^2)$. Figure S19 b displays the maximum (non-compact), minimum (compact), and average areas (mean of maximum and minimum area) available for interactions

when another molecule/protein is added to the cluster. This average area (A_{avg}) is used to compute the probability of cluster formation. For instance, the probability of cluster formation of three proteins (p_3) is given as,

$$p_3 = \frac{p_2 * A_{(avg,2)} * 2}{4\pi R^2 - 2 * \pi r^2} \quad (S18)$$

This can be generalized to compute the probability of cluster formation of N protein (p_N) as:

$$p_N = \frac{p_{(N-1)} * A_{(avg,N-1)} * 2}{4\pi R^2 - (N-1) * \pi r^2} \quad (S19)$$

Where, $p_{(N-1)}$ and $A_{(avg,N-1)}$ are the probability of cluster formation and the average interaction area available in a cluster of $(N-1)$ number of proteins.

Above, we derived the expression of probability to form configurations where all the molecules are clustered. However, if a certain number of proteins are incorporated into liposomes, all the proteins may not form clusters. Some proteins may remain unclustered or free. For instance, if five molecules are incorporated into the liposome, they form a configuration where three of them form a cluster and the other two remain free. What is the probability of forming such configurations? Let us assume a cluster with m_1 number of unit clusters, m_2 number of clusters of 2 molecules, ...m_n number of n molecules, and so on. The probability of forming such a configuration is given by:

$$\left(\prod_{n=2}^{N} p_l^{m_n}\right) * \frac{N!}{\prod_{n=1}^{N} m_n\,!\,(n!)^{m_n}} \quad (S20)$$

The first part of this equation comes from the joint probability of forming clusters of different sizes, whereas, the second part accounts for the exchangeability of the proteins. Consider the example of the above-mentioned configuration of five proteins with an aggregate of three proteins and two unclustered proteins (unit clusters). The probability of forming such a configuration is given by:

$$(p_3)^1 * \frac{5!}{[2!\,(1!)^2] * [1!\,(3!)^1]} = p_3 * 10 \quad (S21)$$

Figure S20 and Table S6 compares the probability of the configurations obtained from the analytical approach (Eqn. S20) and simulations using both random sampling and MMC algorithms. Clearly, there is a good agreement between these two methods. It looks like that although the population obtained from the MMC algorithm agrees reasonably well with theory at J values in the range of -4 k_BT to 0 k_BT, it deviates from theory at J = -10 to -4 k_BT. The reason behind this may be even after multiple rounds of iterations (~100 million), heavily clustered states (all clustered) do not appear in the MMC simulations and therefore the contribution of these populations comes out as zero (Eqn. S11). However, as these configurations are energetically favorable, they show up in the population obtained from the analytical theory.

S13. Photokinetic modeling of cluster-mediated lifetime quenching of LHCII in lipid environment

Once the chlorophyll molecules of LHCII get excited from the ground state (S_0) to the singlet excited state (S_1), there can be several pathways that lead to its relaxation to the ground state. A fraction of the population relaxes back to the ground state via radiative transition. Another fraction gets trapped into the Chl triplet state which is readily transferred to the Car triplet state via triplet transfer. At higher illumination intensities when the number of excitations per pulse is greater than unity, singlet-singlet annihilation (S-S annihilation) can occur. On the other hand, at a higher repetition rate of the laser, there can be a significant build-up of the triplet state populations due to their longer microsecond lifetimes. This triplet state can quench the singlet excited state population via singlet-triplet annihilation (S-T annihilation) providing another nonradiative pathway for excited state decay of the S1 state. This essentially shortens the excited state lifetime of chlorophyll molecules further (Figure S22).

The number of excitations per pulse (n) of LHCII can be computed as follows:

$$n = \frac{I * \sigma * \lambda}{hc} * \frac{1}{w} \qquad (S22)$$

$$I = \frac{P}{\pi * a * b} \qquad (S23)$$

$$\sigma = 2.303 * \frac{\epsilon_{610}}{NA} \qquad (S24)$$

where, I, σ, and w are laser intensity, absorption cross-section of a LHCII trimer, and repetition rate of the laser (80 MHz) respectively. P is laser power (350 nW). a (280 nm) and b (380 nm) are the laser spot sizes (FWHM). λ is the excitation wavelength (610 nm). ϵ_{610} ($4.36 \times 10^5\ M^{-1}cm^{-1}$) and NA are the extinction coefficient of a LHCII trimer at 610 nm and Avogadro's constant, respectively. h and c are Planck's constant and speed of light, respectively. Based on the above calculations, the excitation per pulse for a single LHCII is 0.0269, which is well below 1 (Figure S23). Therefore, we can ignore the effect of S-S annihilation as a possible quenching pathway of the S_1 population.

Now, assuming S-T annihilation as a quenching pathway in our sample, the decay of the excited population ($[S_1]$) can be described by the kinetic equations below:

$$\frac{d[S_1]}{dt} = G(t) - (k + k_{ISC})[S_1] - k_{ST}[S_1][T] \qquad (S25)$$

$$\frac{d[T]}{dt} = k_{ISC}[S_1] - k_T[T] \qquad (S26)$$

Here, k_{ISC} (1/8.54 ns^{-1}) and k_T (1/7 μs^{-1}) are rate constants for inter-system crossing and triplet state decay, respectively. k (1/5.81 ns^{-1}) is the linear de-excitation rate from S_1 state. k_{ST} (1/36ps) is the rate constant for S-T annihilation. $G(t)$ is the excitation rate which depends on the

absorption cross-section of the sample, illumination intensity, and the repetition rate of the laser. As T_{Chl} readily transfers the population to the triplet state of carotenoids (T), this triplet transfer process is effectively eliminated from the above treatment. All the parameters are taken from Gruber et al. (*27*).

As k and k_T differ by three orders of magnitude, the change in triplet concentration in the steady state regime between two subsequent laser pulses is almost negligible compared to the accumulated triplet concentration. Therefore, the kinetics of $[S_1]$ are governed by the equation below:

$$\frac{d[S_1]}{dt} = G(t) - (k + k_{isc} + k_{ST} * [T]_0)[S_1] \qquad (S27)$$

Where, $[T]_0$ is the population of the triplet state in quasi-stationary condition. Hence, the excited state lifetime of the sample is given by:

$$\tau = \frac{1}{(k + k_{isc} + k_{ST} * [T]_0)} \qquad (S28)$$

In this quasi-stationary regime (i.e., at $\tau = \frac{1}{(k+k_{isc}+k_{ST}*[T]_0)}$), the population of the triplet can be calculated as (*47*):

$$[T]_0 = \frac{k_{isc} * [S_1]_0}{k_T} = \frac{k_{isc}}{k + k_{isc} + k_S T[T]_0} * \left(\frac{n_0}{k_T}\right) * w \qquad (S29)$$

Where, $n_0 = \int G(t)dt$ is the fraction of the population into the S_1 state after single-pulse. From the above equation, $[T]_0$ is given as:

$$[T]_0 = \frac{-k + \sqrt{k^2 + \frac{4k_{ST} * k_{isc} * n_0 w}{k_T}}}{2k_{ST}} \qquad (S30)$$

Combining Eqn. S28 and S30, we get,

$$\frac{1}{\tau} = \frac{k}{2} + k_{isc} + \sqrt{\frac{k^2}{4} + \frac{k_S T * k_{isc} * w}{k_T} * n_0} \qquad (S31)$$

Eqn. S31 can be used to estimate the excited state lifetime of single LHCII complexes in the presence of S-T annihilation.

Lifetime quenching in LHCII clusters Once multiple LHCII complexes are incorporated in the membrane environment, we observe a shortening of lifetime. This quenching of lifetime depends on protein concentration in the membrane. We model this cluster-dependent quenching in LHCII invoking S-T annihilation.

A small cluster of LHCII proteins can be considered a single supermolecule comprising of various accessible energy levels. As the size of these clusters increases, the singlet excited states

can come into contact with triplet states from other LHCII molecules. This enhances the probability of S-T annihilation. Therefore, for a cluster consisting of *n* number of LHCII proteins, the decay of its singlet excited state can be given as:

$$[S_1] = n_0 * e^{-(k+k_{isc}+n*k_{ST}*[T]_0)t} \quad (S32)$$

In other words, the excited state lifetime of its S_1 state is:

$$\tau = \frac{1}{k + k_{isc} + n * k_{ST} * [T]_0} \quad (S33)$$

Figure S24 displays the excited state lifetime of LHCII clusters with cluster sizes from 1-10. As the size of the cluster gets bigger, its lifetime gets shorter due to enhanced S-T annihilation in the bigger clusters.

Now, the computation of excited state lifetime using Eqn. S33 gives a value of 0.14 ns for an isolated LHCII complex. However, this is much shorter than the experimentally observed value of the lifetime of LHCII in detergent (~ 3.3 ns) or singly-occupied LHCII-proteoliposome (2.8 ns). On the other hand, for a cluster of 10, this model predicts a lifetime of <0.01 ns. This contradicts the experimentally observed lifetime of LHCII-proteoliposome samples with <N> of 10 which is ~ 2.0 ns. This discrepancy is probably due to the fact that in small photosynthetic antenna complexes like LHCII or a small cluster of LHCII contain a limited number of triplet states. Therefore, a simple kinetic model as described above overestimates the contribution of S-T annihilation. To account for the limited number of excitons and their discrete nature, a stochastic model needs to be used as developed in (*27*).

Stochastic model of ST annihilation Figure S25 demonstrates the stochastic model used to compute the probability of different states (P_{ij}). $G(t)$ is the rate of excitation. A rectangular pulse profile was used to model $G(t)$ with pulse-width and inter-pulse delay of 130 fs and 12.5 ns respectively, similar to the experiments. The probabilities of each state (P_{ij}) can be computed by solving the set of coupled differential equations below:

$$\frac{dP_{00}}{dt} = -G(t)P_{00} + kP_{10} + k_T P_{01}$$

$$\frac{dP_{10}}{dt} = G(t)P_{00} - (k + k_{isc})P_{10} + k_T P_{11}$$

$$\frac{dP_{01}}{dt} = k_{isc}P_{10} + (k + k_{ST})P_{11} - (G(t) + k_T)P_{01} + 2k_T P_{02}$$

$$\frac{dP_{11}}{dt} = G(t)P_{01} - (k_T + k + k_{ST} + k_{isc})P_{11} + 2k_T P_{12}$$

$$\frac{dP_{02}}{dt} = -(G(t) + 2k_T)P_{02} + k_{isc}P_{11} + (k + 2k_{ST})P_{12}$$

$$\frac{dP_{12}}{dt} = G(t)P_{02} - (k + 2k_T + 2k_{ST})P_{12}$$

These equations were solved numerically using a home-written MATLAB (Mathworks, Inc.) code with the initial condition, $[P_{00},P_{01},P_{10},P_{11},P_{02},P_{12}] = [1,0,0,0,0,0]$. The step size of the simulation was 100 fs (< 130 fs, the pulse peak width). The simulation was run up to w/k_T pulses as the triplet states experience that many numbers of excitations by the pulsed laser.

Solutions of the probability states at $t = 1/k_T$ can be used to determine the population of triplet state ($[T]_0$):

$$[\mathrm{T_0}] = (P_{01} + P_{11}) + 2 * (P_{02} + P_{12}) = 1.8 * 10^{-3} \quad \text{(S34)}$$

However, as it is shown in Figure S26a, only P_{01} contribute significantly to the population of the triplet state. The other states have negligible triplet state contributions. This also suggests that a further extended model is not necessary. Figure S26b displays the excited state lifetime of LHCII as a function of their cluster size (j). The kinetic parameters used for the simulation are provided in the caption. Except for the k_{ST}, all the other parameters are obtained from Gruber et al. (*27*). Here, k_{ST} is adjusted to 12 ps (36 ps reported in (*27*)) to match the median lifetime observed in isolated LHCII complex (~ 2.5 ns) under in single-molecule studies (*25*).

S14. Intensity dependence of excited state lifetime in the isolated LHCII complexes

The Eqn. S33 shows the excited state lifetime of the LHCII complexes depends on the number of proteins in the cluster (j), as well as the stationary triplet state population ($[T]_0$). Now, $[T]_0$ can be increased by increasing the excitation intensity of the sample which in turn should shorten the excited state lifetime. To investigate the validity of this prediction from our model, we measured the single-molecule lifetime distribution of single LHCII complexes immobilized in a PVA matrix at various laser power. The lifetime distributions, the median of these distributions and the prediction from our model are shown in Figure S28. We see a good agreement between the experiment and simulations that validate the model and model parameters.

For the measurements in a PVA matrix, LHCII was diluted to ~ 15 pM in a solution containing 5% PVA (MW 78000, 99% hydrolyzed, Polysciences, Inc.), PCA/PCD scavenging mixture and buffer (10mM HEPES, 20 mM NaCl, 0.06% (w/v) GDN, pH 7.5). An argon chamber was used for the single-molecule measurements of LHCII in the PVA matrix to reduce photodegradation (*25*).

S15. Extraction of LHCII-LHCII interaction energy from lifetime data

Figure S29 demonstrates the steps followed to extract the LHCII-LHCII interaction energy (J) from single-molecule lifetime data. Although the clustering reduces the fluorescence intensity of LHCII proteoliposomes as well as shortens their lifetimes, we employ only lifetime data to extract J. This is because the excited state lifetime is a more robust quantity than the fluorescence intensity as it does not get influenced by day-to-day variation of the alignments of our optical set-up and lifetime is independent of illumination intensity. Only the first levels of single-

molecule data are used to extract J as in the later levels the proteins undergo photodegradation due to a longer exposure and the quenching model described in the previous section does not include such effects.

Although the average number of LHCII (<N>) embedded in liposomes can be controlled by lipid to protein ratio, this number is very heterogeneous (*20*, *25*). We assume that the number of LHCII incorporated in LHCII follows a Poisson distribution. Later, we investigated the effect of Poisson and non-Poisson distribution of LHCII in liposomes on the pairwise interaction energy (Figure S33d). Both analyses show the same LHCII-LHCII interaction energy peaks. Figure S28a shows the Poisson distribution for <N>=5.

In the first step of the simulation, for a sample with a specific <N>, the number of LHCII in a liposome (N) is selected from its corresponding Poisson distribution. Once the N is selected, the Matlab program utilizes the population of various configurations at different J (Figure S29b) for that N to compute the average excited state lifetime for these configurations. Based on the size of clusters in a particular configuration, the quenching model predicts their lifetimes (Figure 2a, main text). For instance, in a liposome with five LHCII incorporated the configurations with a cluster of three and a cluster of 2 LHCII, the lifetime (τ) is given as:

$$\tau = \frac{1}{5} * (2 * \tau_{N2} + 3 * \tau_{N3}) \qquad (S35)$$

where, τ_{N2} and τ_{N3} are the lifetime of clusters consisting of two and three LHCII proteins, respectively. The lifetime of different clusters as a function of sizes is given in Figure 2a (main text) computed from the quenching model (Sec. 13). The steps a-c in Figure S29 are run multiple times to construct a lifetime distribution for a particular LHCII interaction energy (Figure S29d). This figure shows the simulated lifetime distribution for the LHCII-proteoliposome sample with <N> = 5 and $J = -5\ k_BT$.

Finally, the above procedures are repeated for multiple J, ranging from 0 to -10 k_BT, and compared with the experimental lifetime distributions. Figure S30 shows the experimental single-molecule lifetime distribution for LHCII-proteoliposomes with <N> of 5, 6, and 10 along with the simulated distributions for J of -2, -5 and -8 k_BT. J can be systematically varied to generate lifetime distributions and fit to the experimental data to extract LHCII-LHCII interaction energies. The fits involve converting the lifetime distribution into a probability density function by kernel density estimate (KDE) and subsequent maximum likelihood estimation (MLE) as described below.

Cross-validated Kernel Density Estimate. Kernel density estimation (KDE) is a nonparametric method to estimate an underlying probability density function from a data set. We use KDE to estimate the probability density of fluorescent lifetimes as a function of energy. We first simulate 100,000 points at energies ranging from 0 to -10 kBT in 0.1 increments using the method outlined above and then perform KDE on each data set. For an independent, identically distributed set of variables ($x_1, x_2, \ldots, x_n$) the KDE is given by:

$$p(x) = \frac{1}{nh}\sum_{i=1}^{i=n} \mathrm{K}\left(\frac{x - x_i}{h}\right) \quad (S36)$$

where h is the bandwidth and $\mathrm{K}\left(\frac{x-x_i}{h}\right)$ is the kernel function.

A Gaussian kernel was used in this work. Bandwidth selection is crucial to achieving an appropriate distribution that is neither under nor over-smoothed. We used a method called k-fold cross-validation with k = 5 to pick the bandwidth. In k-fold cross-validation, the data is split into k subsets. One subset is chosen as the test data set while the k-1 subsets are combined into the training set. The KDE is performed using the training data set and then evaluated against the test data set. The process is repeated k times, each time with a different subset used as the test data set. This process was implemented using python's sklearn toolkit (*59*). The code is available upon request.

Maximum Likelihood Estimation. Maximum likelihood estimation (MLE) is used to estimate which interaction energy best fits the experimental data. The likelihood function is given by:

$$L(\theta) = f(X_1, X_2, .., X_n|\theta) \quad (S37)$$

In words, the likelihood is the probability of observing $x_1,x_2,...,x_n$ given some parameter θ. The probability of observing a given set of data is:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \prod_{i=1}^{i=n} p_i \quad (S38)$$

So we can then write,

$$L(\theta) = \prod_{i=1}^{i=n} p_i \quad (S39)$$

In practice, the log-likelihood function is computationally more tractable and is often used.

$$ln\ L(\theta) = ln\left(\prod_{i=1}^{i=n} p_i\right) \quad (S40)$$

$$ln\ L(\theta) = \sum_{i=1}^{i=n} l\,n(p_i) \quad (S41)$$

The probability density function as determined by KDE is integrated to determine the probability of each experimentally observed lifetime to calculate the likelihood. For the global fit, the likelihood functions of each <N> are added together and the resultant function is optimized to obtain a global value of *J*. Table S8 displays the LHCII-LHCII interaction energies obtained from the global fits and samples with different <N>.

Parameter dependence. Polycarbonate membranes with a pore size of 50 nm were used to prepare the liposome. However, the dynamic light scattering (DLS) results of the LHCII-proteoliposome show some variation in their observed sizes (Table S1 and Figure S2). Again, the cluster in the liposome is defined if two proteins are within a distance of *r* (the yellow shaded region in Figure S33a). In our model, we have assumed this *r* to be 1 nm. Both the liposome size (*R*) and the interaction distance effect the relative population of different configurations (Eqn. S20) and therefore the extracted LHCII-LHCII interaction energy (*J*). To investigate the dependence of *J* on *R* and *r* used our model, we systematically varied these parameters. Figure S32b and c show the normalized log-likelihood at different *R* and *r*. We found that these parameters do not affect the extracted *J* significantly. With *R* in the range of 25-37.5 nm and *r* in the range of 0.5-1.5 nm, the extracted *J* still lies in the vicinity of -5 k_BT. Figure S33d displays the dependence of *J* on the S-T annihilation rate constants. With k_{ST} between 12 and 20 ps, *J* lies in the range of -5 to -6 k_BT. However, at slower k_{ST} (~ 36 ps), we do not observe a clear peak in the log-likelihood analysis. This indicates, the value of *J* is weakly dependent on the lifetime of LHCII clusters, which in turn is related to k_{ST} (Eqn. S33). We also investigated the effect of *J* on assuming that proteins in liposomes follow a Poisson and no distribution (non-Poisson) at all. In both cases, we found a similar peak value in the log-likelihood plot. However, the plot for non-Poisson distribution was a bit noisy and the peak was less pronounced.

S16. Surface Charge Calculations

The net surface charge of LHCII (*Q*) was calculated using PROPKA software (*35*), which takes account of the pKa of the protonatable residues and then applies perturbation factors from the protein including coulomb interactions, desolvation, and electrostatics. The charged surface is then created using the Adaptive Poisson-Boltzmann Solver, which solves the equations of continuum electrostatics (*36*).

S17. Free energy change in pH-driven clustering in LHCII proteoliposomes

Based on our single-molecule lifetime data and analysis of LHCII clusters in proteoliposome, we observe moderately clustered configurations at neutral pH to heavily clustered ones at low pH. We attribute this to a change in relative configurations at these two conditions. For instance, as Figure S33 shows, the configuration where all the proteins are clustered constitutes 23% of the population. However, at low pH the relative population of this configuration changes to ~ 93%. We aim to quantify the change in enthalpy ($\Delta\Delta H$), entropy ($\Delta\Delta S$), and free energy ($\Delta\Delta G$) of such transition.

Figure S35 shows a schematic representation of how such thermodynamic parameters are calculated. For the calculation of $\Delta\Delta H$, the pairwise interaction energies at neutral and low pH are required (Section 15). In this particular case, the total interaction energies are $3J_1$ and $4J_2$.

Therefore, ΔΔH is given as $4J_2 - 3J_1$. On the other hand, ΔΔS can be estimated from the probabilities (w) of those configurations (see Methods, main text).

Now, for the proper calculation of $\Delta\Delta G$, the contribution from all other configurations should also be considered. This leads to,

$$\Delta\Delta G = \sum_{i=1}^{m} \Delta\Delta H_i(J) - k_B T * \ln\left(\sum_{i=1}^{m} \frac{w_{l,i}}{w_{n,i}}\right) \quad (S42)$$

where, $\Delta\Delta H_i(J)$ is the change in enthalpy associated with i-th configurations; $w_{l,i}$ and $w_{n,i}$ are the probabilities of such configurations at pH low (pH, 5.0) and neutral pH (pH, 7.5), respectively. Please note that the first term of the right-hand side of the above equation represents the enthalpy change, and the second term represents the change in entropy for the transition.

Table S9 demonstrates the ΔΔG, ΔΔH and ΔΔS for pH-driven aggregation in samples with <N> of 2,5 and 10. As expected, the aggregation is not entropically favorable. However, the favorable changes in enthalpy lead to the negative ΔΔGs driving the clustering.

S18. Free energy change in light-driven clustering in the chloroplast

In the previous section, we have discussed various thermodynamic parameters (ΔΔH, ΔΔS and ΔΔG) involving a pH-driven transition from unclustered to a clustered configuration in LHCII proteoliposome. The main difference between a model LHCII proteoliposome system and an in-vivo thylakoid membrane is the protein density. The maximum protein density in the LHCII proteoliposome system studied here is 10 per liposome or 1.3 LHCII per 1000 nm^2 of membrane area. On the other hand, the density of LHCII in the thylakoid membrane can be as high as 6.3 per 1000 nm^2 of membrane area or ~5 times higher. This can potentially alter the thermodynamic parameters associated with the quenching in-vivo. To investigate this, we analyzed the freeze-fracture electron micrographs of dark-adapted vio and light-treated zea samples obtained from intact spinach chloroplast (*7*). Here, we assume that the vio-dark and zea-light samples represent the systems where LHCII remains in unquenched and quenched configurations, respectively. This is supported by their NPQ values (Table 1 in (*7*)).

Figure S35a shows the freeze-fracture electron micrographs of vio-dark and zea-light samples. The images are cropped versions of the figure reported in (*7*). We used ImagePro software (Media Cybernetics) to analyze these images. First, we identified a few LHCII particles and backgrounds manually from the images. Then these were fed into the 'smart segmentation' feature of the software which thresholds each pixel based on the given intensity values of the particles and the backgrounds. This enables us to segment the image as shown in Figure S36b. The identified areas of the images covered by LHCII particles (green areas in Figure S36) are collected and compared between vio-dark and zea-light samples. As shown in Figure S36c, the median LHCII-covered area of zea-light sample was ~25% higher compared to vio-dark sample indicating a more clustered configuration of LHCII in the former.

In the next step, we used the segmented images to create particle transforms as shown in Figure S37 a-b. For this, we filled the identified areas with circles that represent LHCII particles. The area of each circle was taken to be 50 nm^2 as was observed in (*7*). We divided the transforms into regions of $\sim$ 88 × 88 nm^2 square areas which are equivalent to the surface area of proteoliposomes. This enables us to directly compare the ΔΔH values associated with light-treated clustering of LHCII in chloroplast to pH-induced clustering of LHCII in liposomes. However, we could not extract the ΔΔS (and therefore ΔΔG) values from these images using the analytical method described in Sec. 12. The reason for this was the analytical method computes the probability of different configurations only up to $N = 20$ but the average number of LHCII per 88 nm × 88 nm square regions in the chloroplast were ∼50. Anyway, the change in enthalpy (ΔΔH) can be measured reliably as this does not involve the probabilities of different configurations involved for a particular *N*. We used a similar methodology to calculate the average change in enthalpy (ΔΔ*H*) of vio-dark and zea-light samples as was used in the liposome. To give an example, the region highlighted in yellow in Figure S36a contains 45 LHCII proteins. These 45 proteins are in a configuration where they form 19 clusters of 1 (1 × 19), 6 clusters of two (2 × 6), two clusters of three (3× 2), and one cluster of eight (8 × 1). Therefore, the enthalpy in that region is calculated as:

$$\Delta\Delta H(J) = (1-1)19J + (2-1)6J + (3-1)2J + (8-1)J = 17J \qquad (S43)$$

Similarly, we computed the ΔH for all other square regions. Finally, these mean ΔH values are used to calculate light-induced enthalpy change in the chloroplast in the following way:

$$\Delta\Delta H = <\Delta H>_{zea} - <\Delta H>_{vio} \qquad (S44)$$

where $<\Delta H>_{vio}$ is the enthalpy averaged over all the 88 nm × 88 nm regions in vio-dark sample and $<\Delta H>_{zea}$ are the enthalpy averaged over all the 88 nm × 88 nm regions in zea-light sample. Table S9 shows the ΔΔH for a transition of vio-dark to zea-light configuration of the chloroplast.

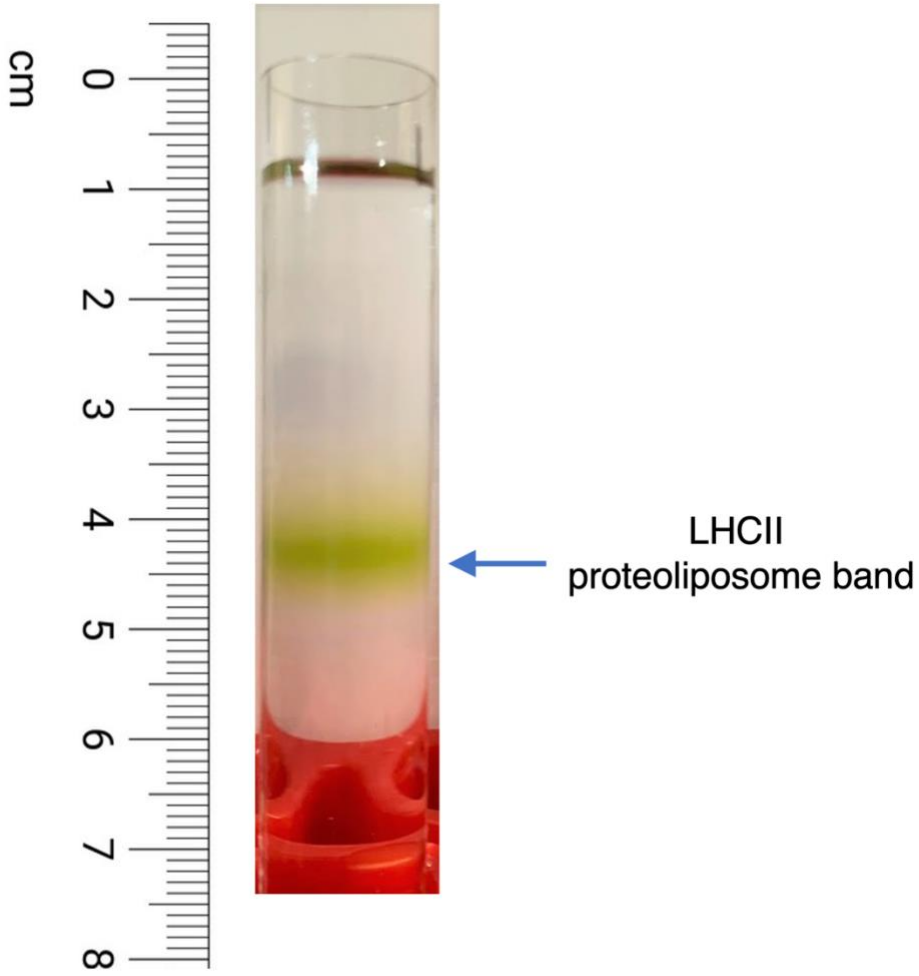


Figure S1: A representative sucrose gradient showing a LHCII proteoliposome band for <N>=5.

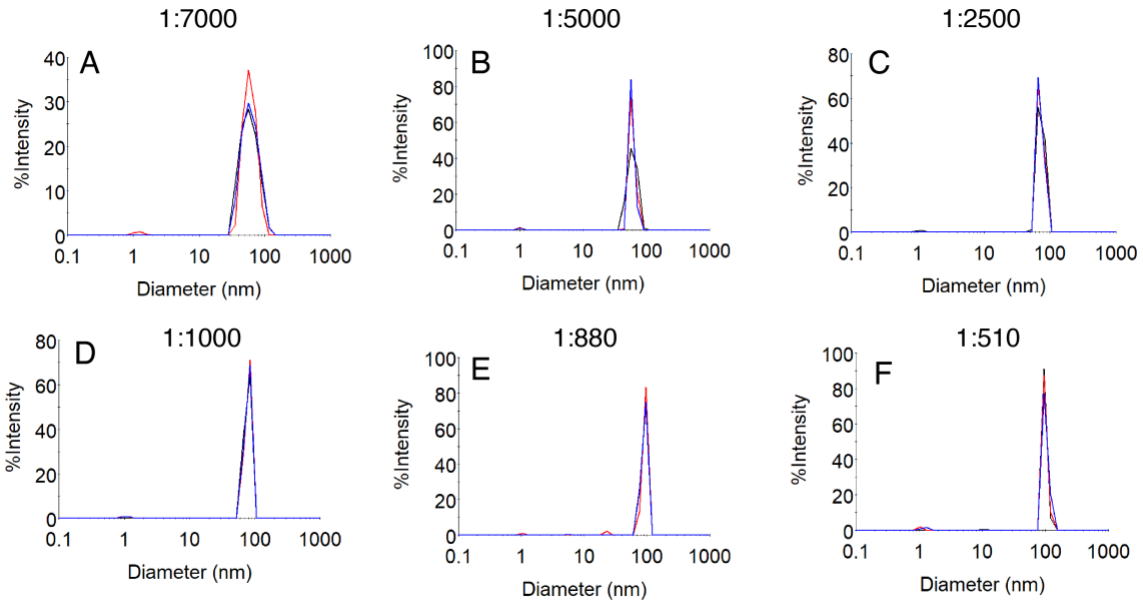


Figure S2: Size distribution of LHCII-proteoliposomes with various protein-to-lipid ratios obtained from dynamic light scattering (DLS) measurements. Blue, red and black lines represent the technical replicates of the measurements.

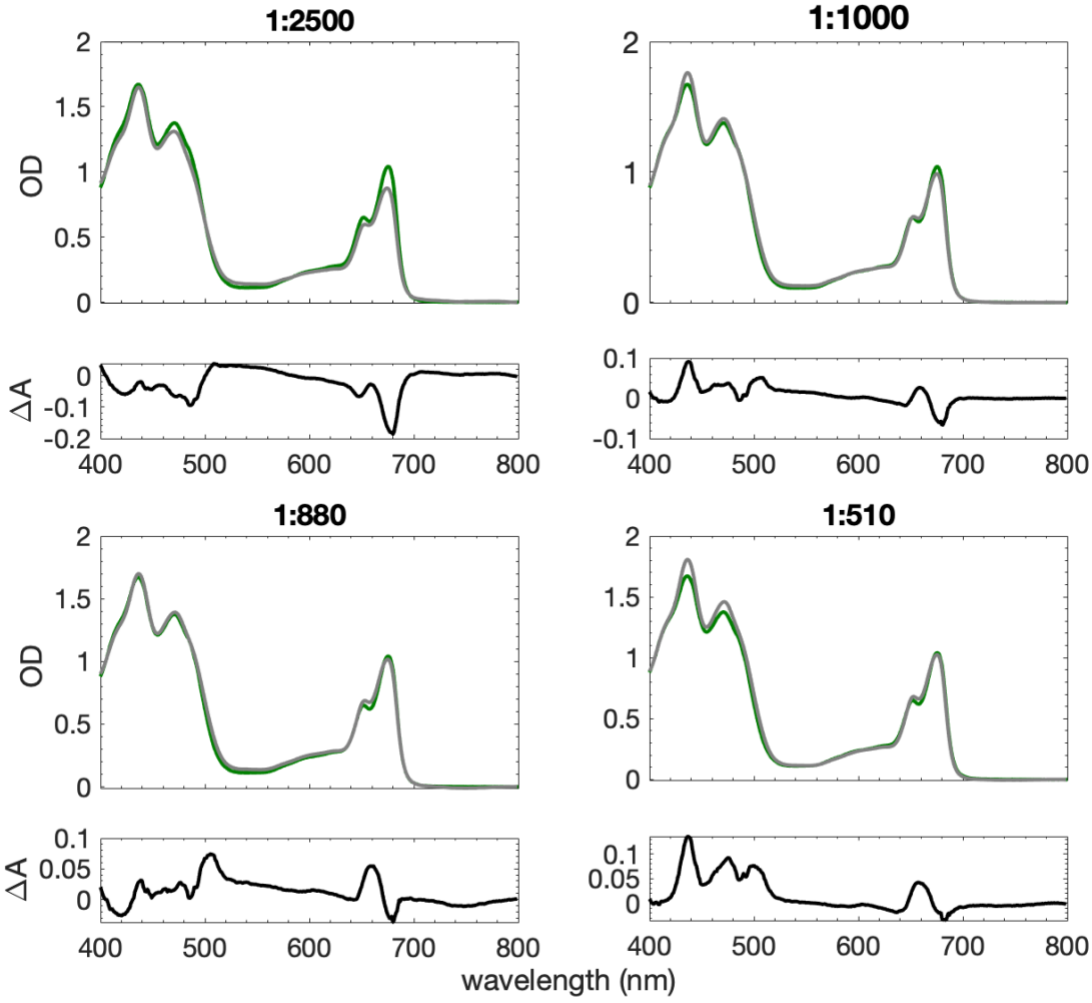


Figure S3: Absorption spectra of LHCII in detergent (10mM HEPES, 20 mM NaCl, 0.06% (w/v) GDN, pH 7.5) and incorporated in liposomes with protein-to-lipid ratio of 2500, 1000, 880 and 510. The difference spectra ($\Delta A = OD_{liposome} - OD_{detergent}$) are also displayed below each of the corresponding absorption spectra. Absorption spectra are normalized at 405 nm.

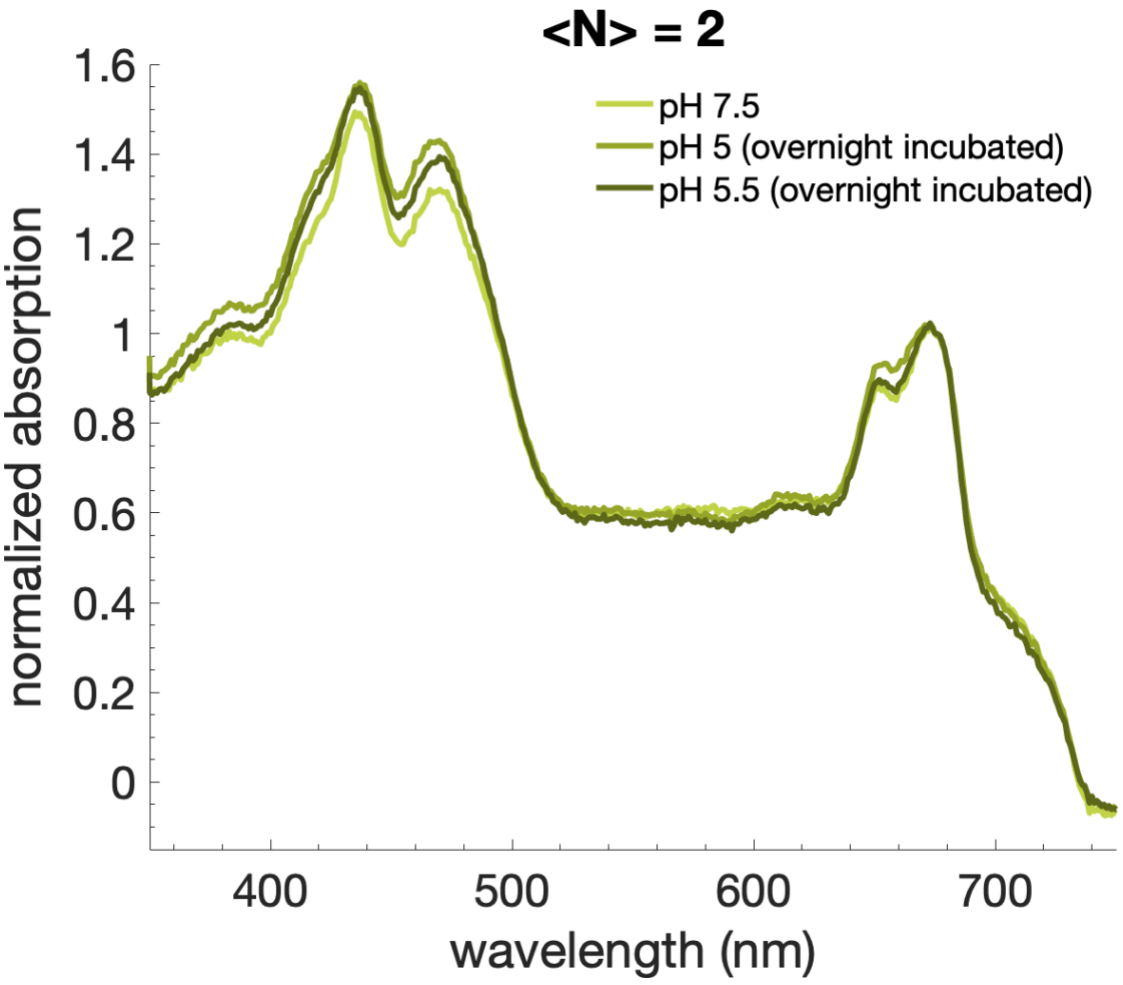


Figure S4: Absorption spectra of LHCII proteoliposome with <N>=2 incubated overnight at low pH.

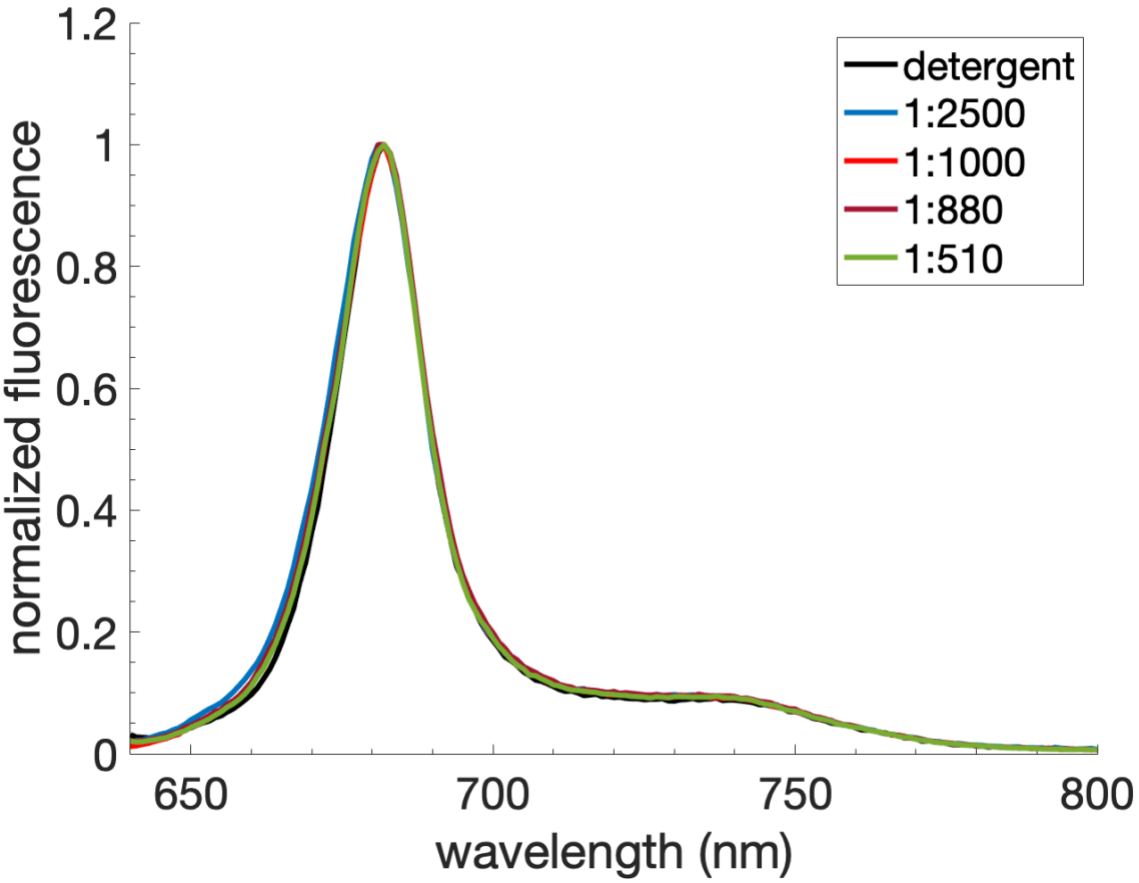


Figure S5: Normalized emission spectra of LHCII in detergent (10mM HEPES, 20 mM NaCl, 0.06% (w/v) GDN, pH 7.5) and incorporated in liposomes with protein-to-lipid ratios of 2500, 1000, 880 and 510.

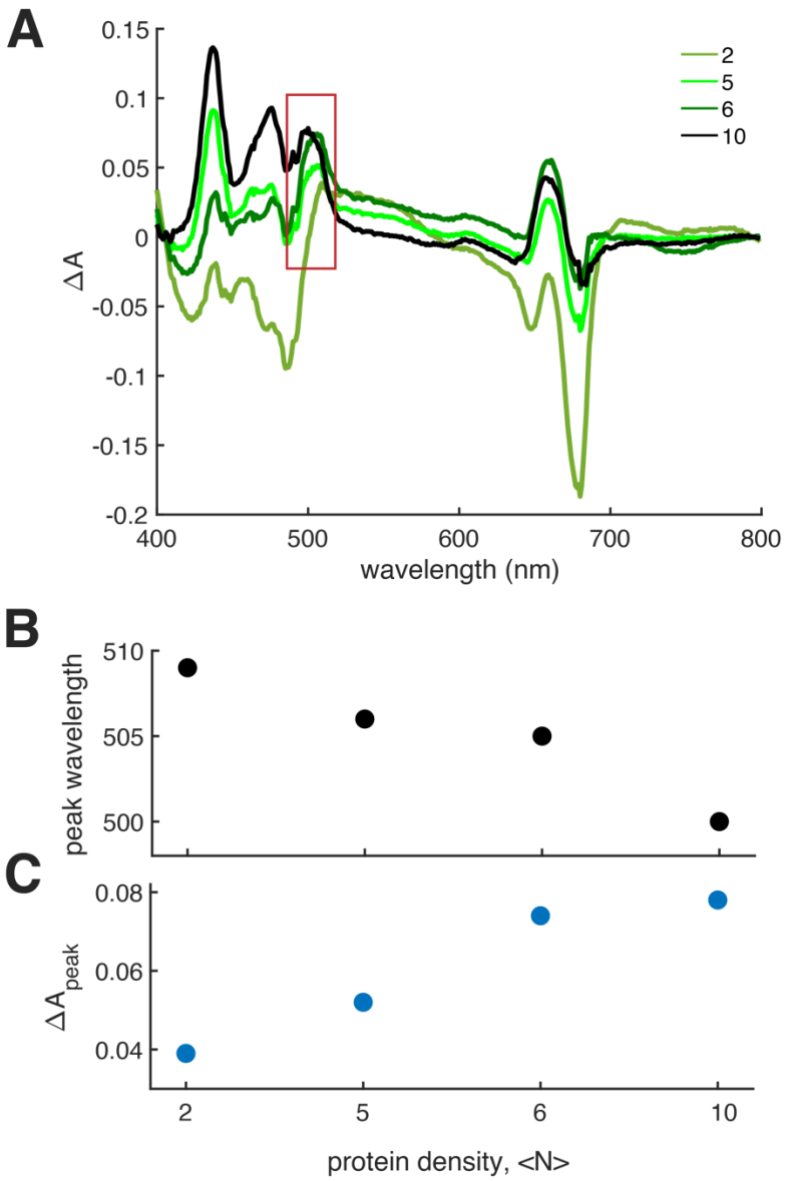


Figure S6: (a) Difference spectra ($\Delta A = OD_{liposome} - OD_{detergent}$) of the LHCII proteoliposome with varying protein density. The peak wavelengths around the 500 nm region of the difference spectra (within the box shown in (a)) and ΔA at that peak wavelengths are displayed in (b) and (c), respectively.

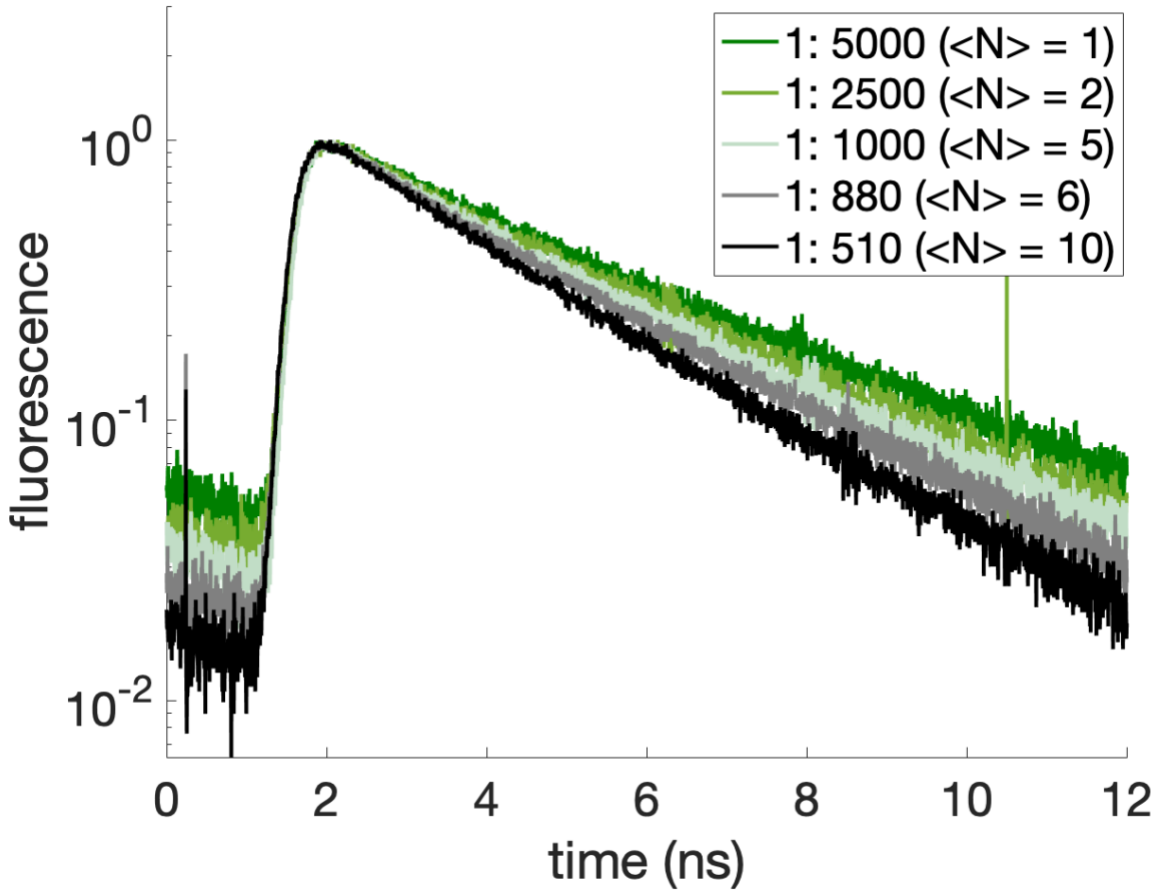


Figure S7: Ensemble lifetime of LHCII proteoliposome samples. Normalized fluorescence lifetime decay of LHCII embedded in liposomes with various protein-to-lipid ratios. The protein-to-lipid ratios used in this study are 1:5000, 1:2500, 1:1000, 1:880 and 1:510. The average number of LHCII (<N>) in liposomes in these samples are 1, 2, 5, 6 and 10 respectively. The fluorescence lifetime decay gets shortened upon increasing protein density in liposomes.

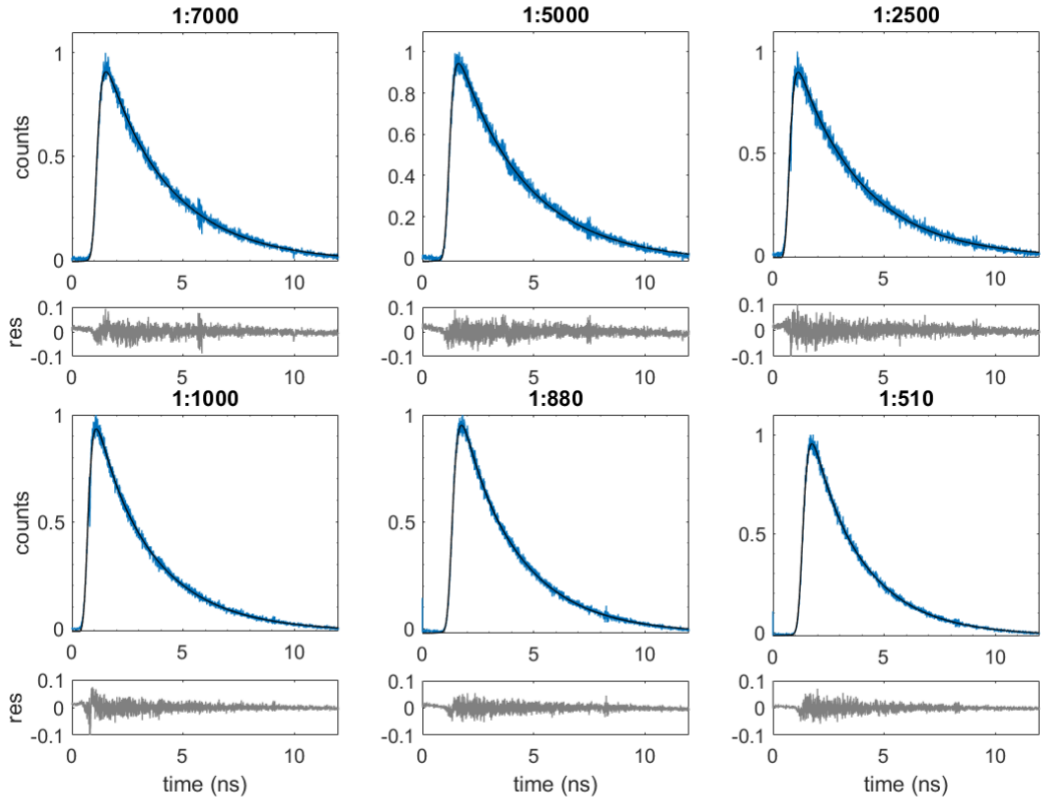


Figure S8: The ensemble lifetime decay of LHCII-proteoliposomes at different protein-to-lipid ratios. The figures display the fluorescence decay (blue lines) of the samples and their corresponding fits (black lines) along with the residuals. The decays are fitted with bi-exponential functions employing the non-linear least square-based method. The fit results are displayed in Table S2.

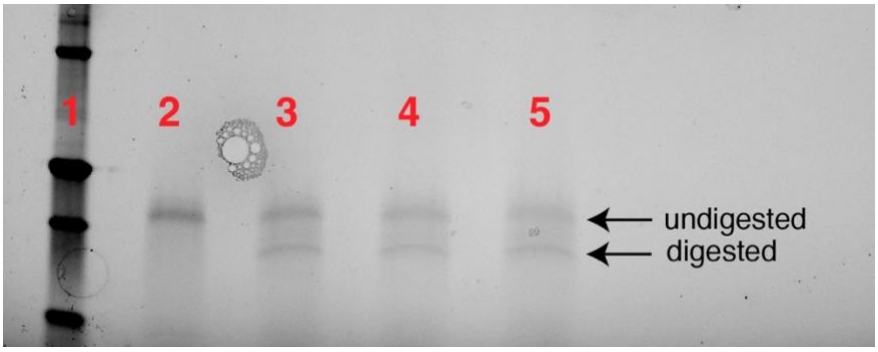


Figure S9: SDS-PAGE gel of trypsin digest products reveals random insertion of LHCII in liposome. The lanes are as follows: 1) Ladder. 2) LHCII proteoliposome pre-digestion. 3) Digestion products after 30 minutes. 4) Digestion products after 1 hour. 5) Digestion products after 2 hours.

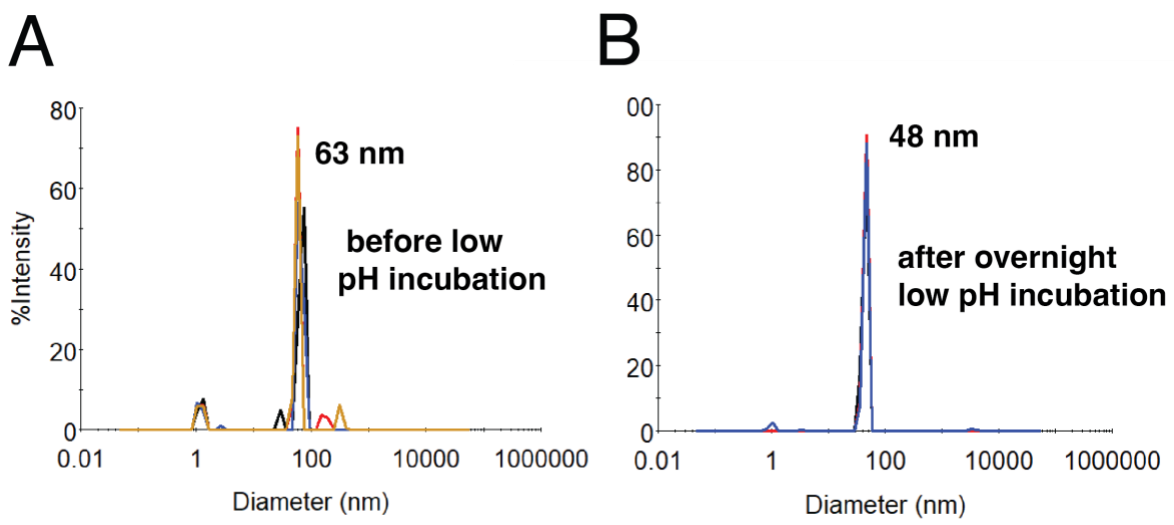


Figure S10: pH-dependent measurements of proteoliposome. Size of the proteoliposome (<N> = 5) sample (a) before and (b) after the overnight incubation as obtained by dynamic light scattering (DLS) measurements. This indicates that the liposomes remain intact after incubation. The slight difference in diameter may be due to a change in the viscosity of the solvent.

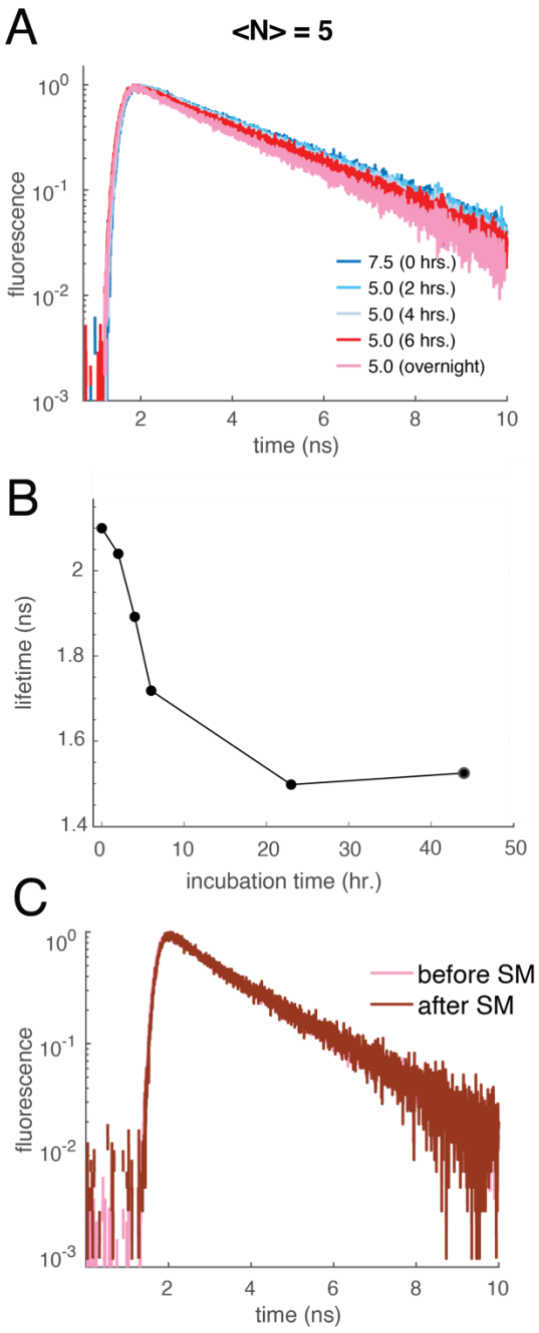


Figure S11: pH-dependent measurements of proteoliposome. (a,b) Fluorescence lifetime decays and the average lifetimes of the sample at the different stages of low pH incubation. (c) Single-molecule measurements of the incubated samples were made after ~10 hrs. of incubation. The overlay of the fluorescence decays of <N>=5 sample before and after the single-molecule measurements shows that the average lifetimes remain intact during the measurements.

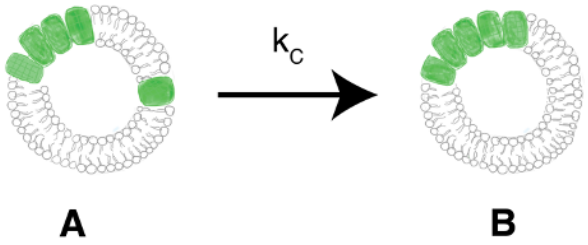


Figure S12: Clustering of proteins from configuration A to B happens at a rate of $k_c(1/\tau_c)$.

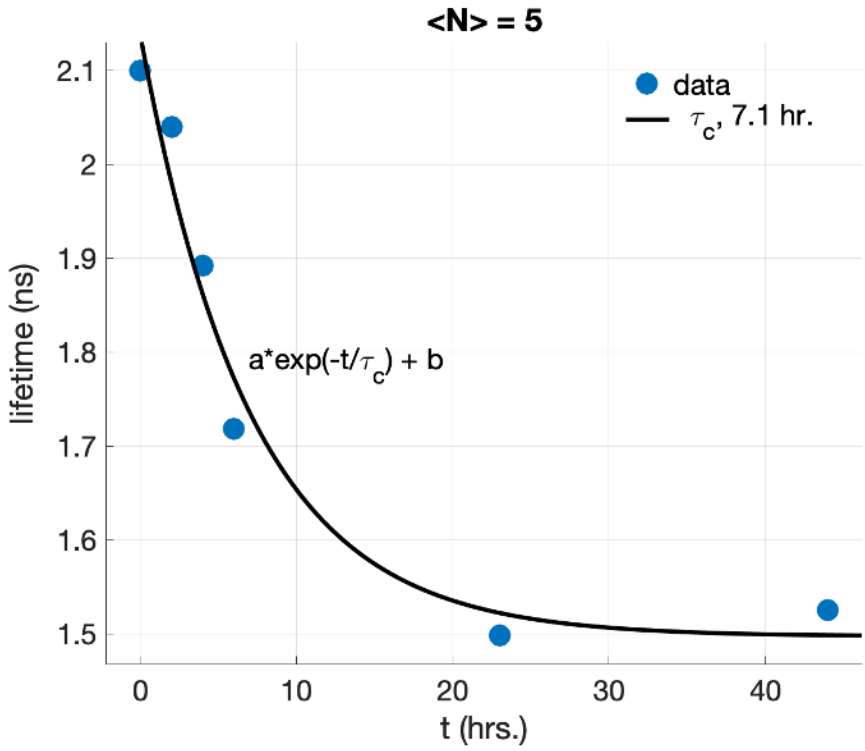


Figure S13: Lifetime drop in the sample with <N> = 5 sample with liposome diameter of 50 nm upon incubation at pH 5. The blue-filled circles are experimentally obtained lifetimes and the black line is the exponential fit.

Fig. S14.

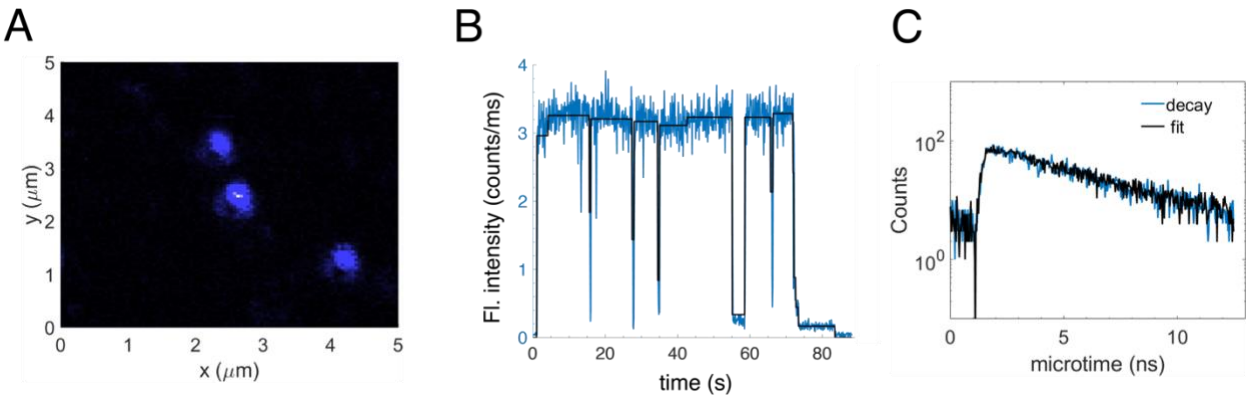


Figure S14: Single-molecule measurement of LHCII-proteoliposome.(a) A representative confocal image showing three LHCII proteoliposomes immobilized on a coverslip. (b) A typical fluorescence trace upon excitation of a single LHCII proteoliposome. The regions of constant fluorescence intensity or 'states' are identified by a change-point finding algorithm. (c) A Representative fit for the lifetime decay of a 'state' in single-molecule experiments based on maximum likelihood estimation (MLE) algorithm.

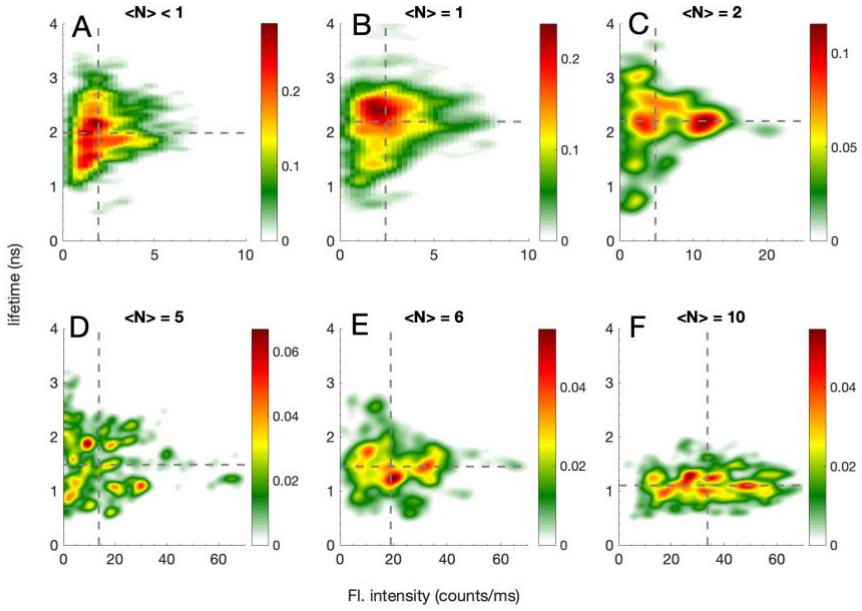


Figure S15: 2D density plots displaying lifetime and fluorescence intensity of LHCII-proteoliposome samples with an average number of proteins (<N>) ranging from less than one to ten. The dotted lines show the median lifetime and fluorescence intensity of the samples. The vertical color bars display probability densities.

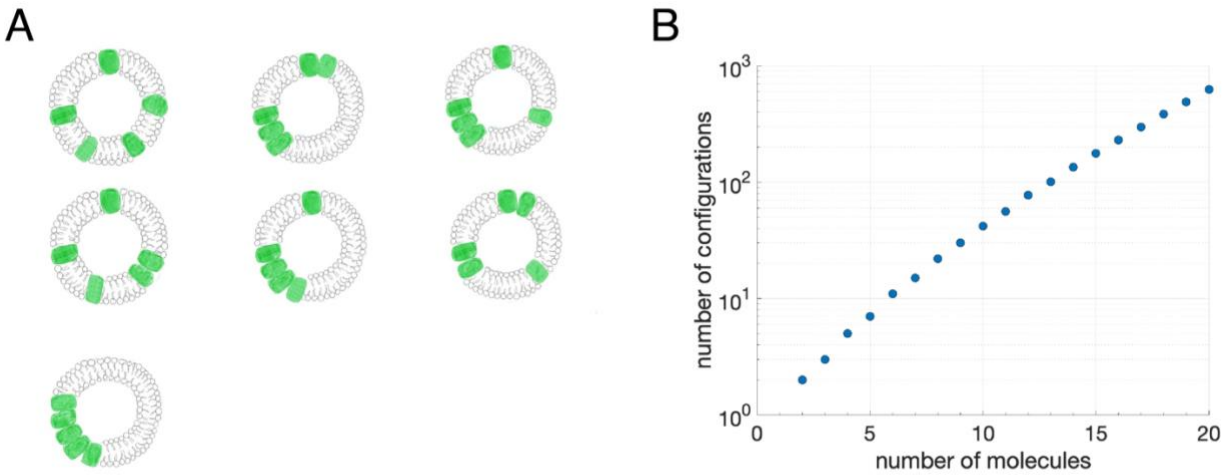


Figure S16: (a) Cartoons of the seven different configurations of clusters possible when five LHCII proteins are incorporated into a liposome. (b) The number of possible configurations increases with the number of proteins.

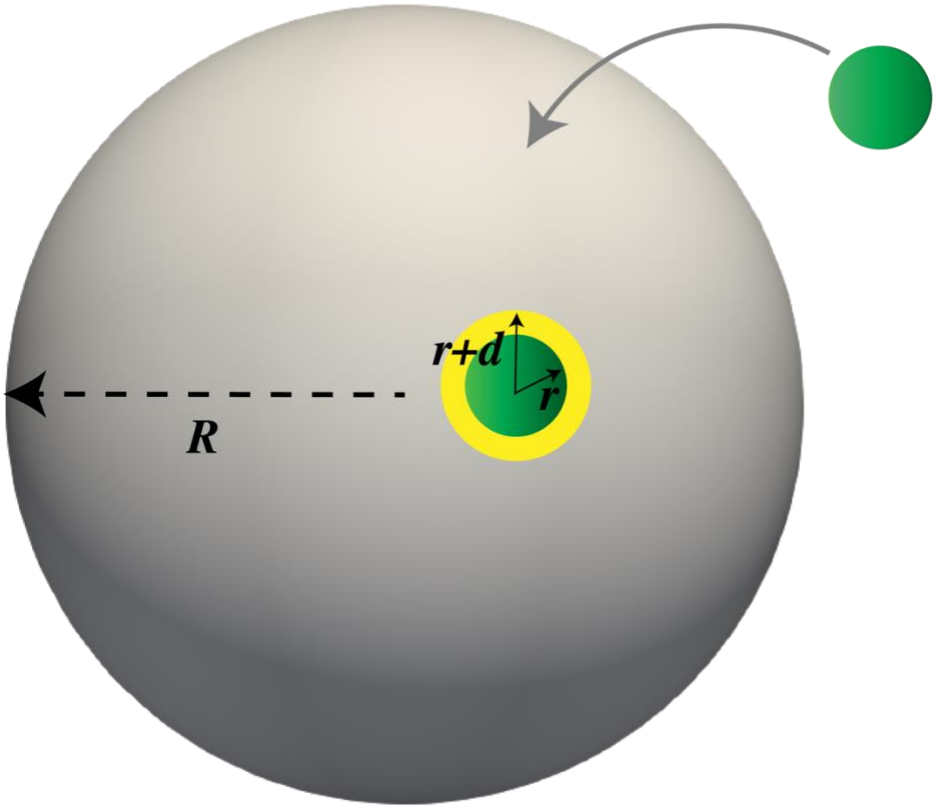


Figure S17: Probability of forming a cluster of two proteins on a spherical surface with radius R is proportional to the area shaded in yellow**.**

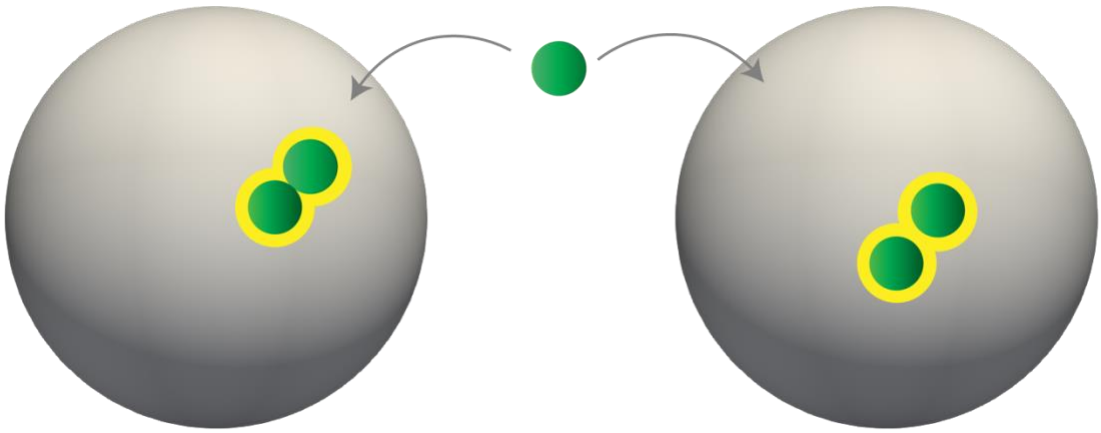


Figure S18: Probability of higher order cluster formation is proportional to the yellow shaded region. Two extreme configurations of clusters are shown in the left (minimum probability/compact arrangement) and right (maximum probability/non-compact arrangement) spheres.

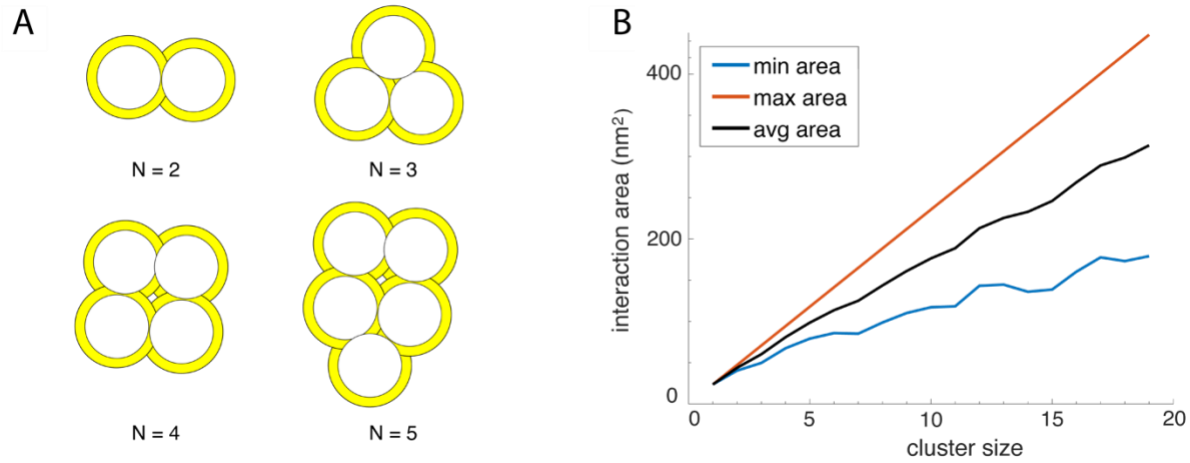


Figure S19: (a) Cluster of different sizes (N = 2,3,4,5) arranged in compact ways. The minimum area available for interactions is shown as a yellow-shaded region. (b) Minimum, maximum, and average area of interactions for different cluster sizes.

Fig. S20.

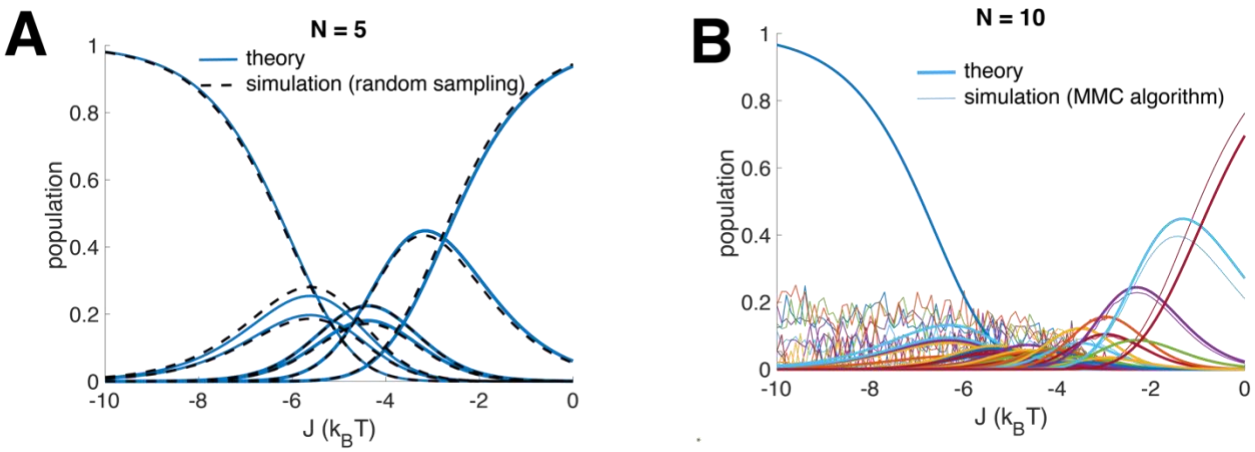


Figure S20: Comparison of the population of different configurations between theory (analytical) vs. simulation obtained using (a) random sampling and (b) MMC algorithms.

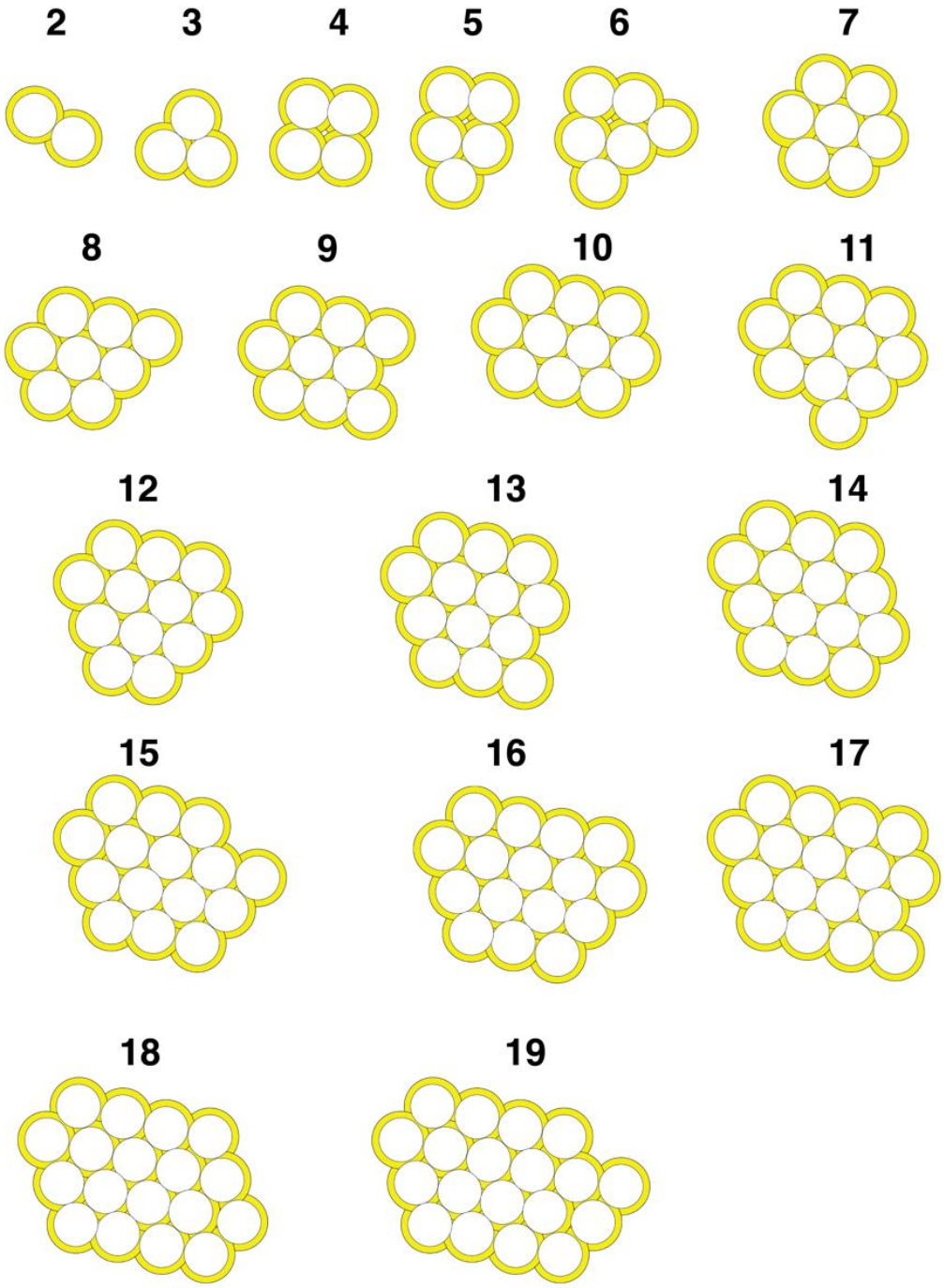


Figure S21: LHCII cluster geometry considered to obtain their analytical probability of formation.

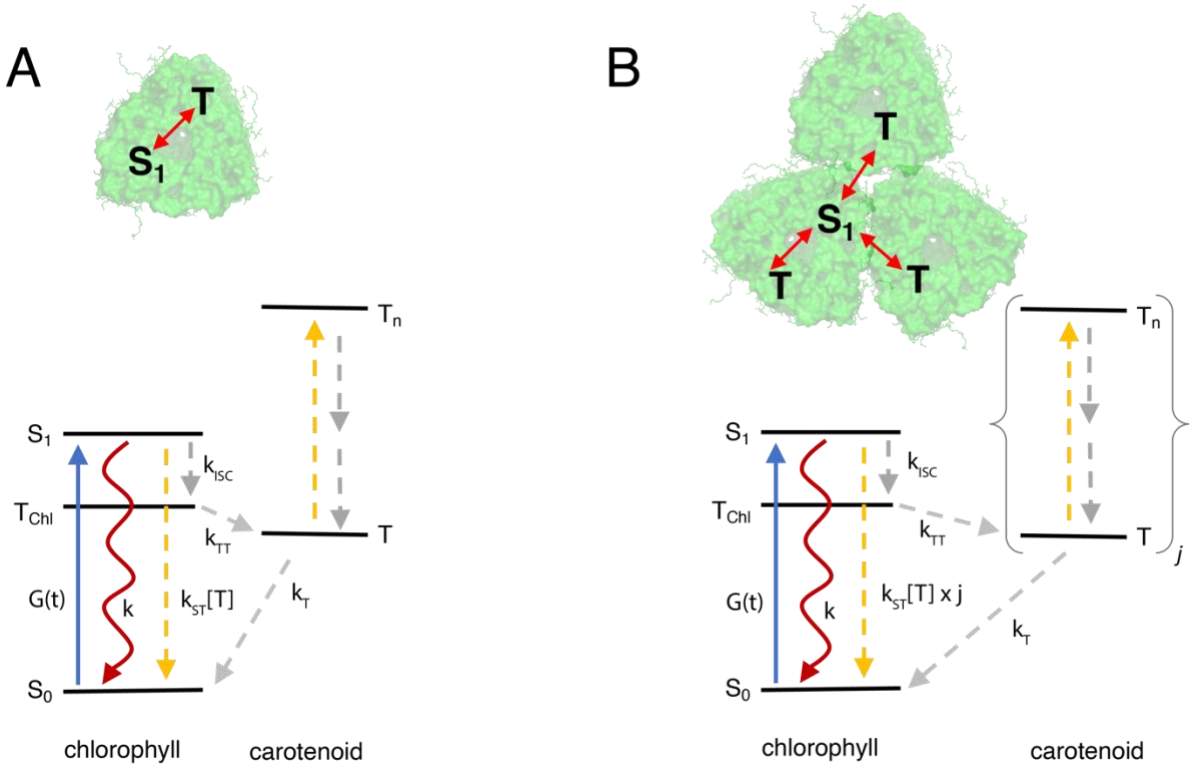


Figure S22: Kinetic model describing excited state relaxation pathways of isolated (a) and clustered LHCII complexes (b). In the clustered complex, the excited state of one LHCII complex can be quenched by the triplet states from other complexes via S-T annihilation process resulting in a faster decay.

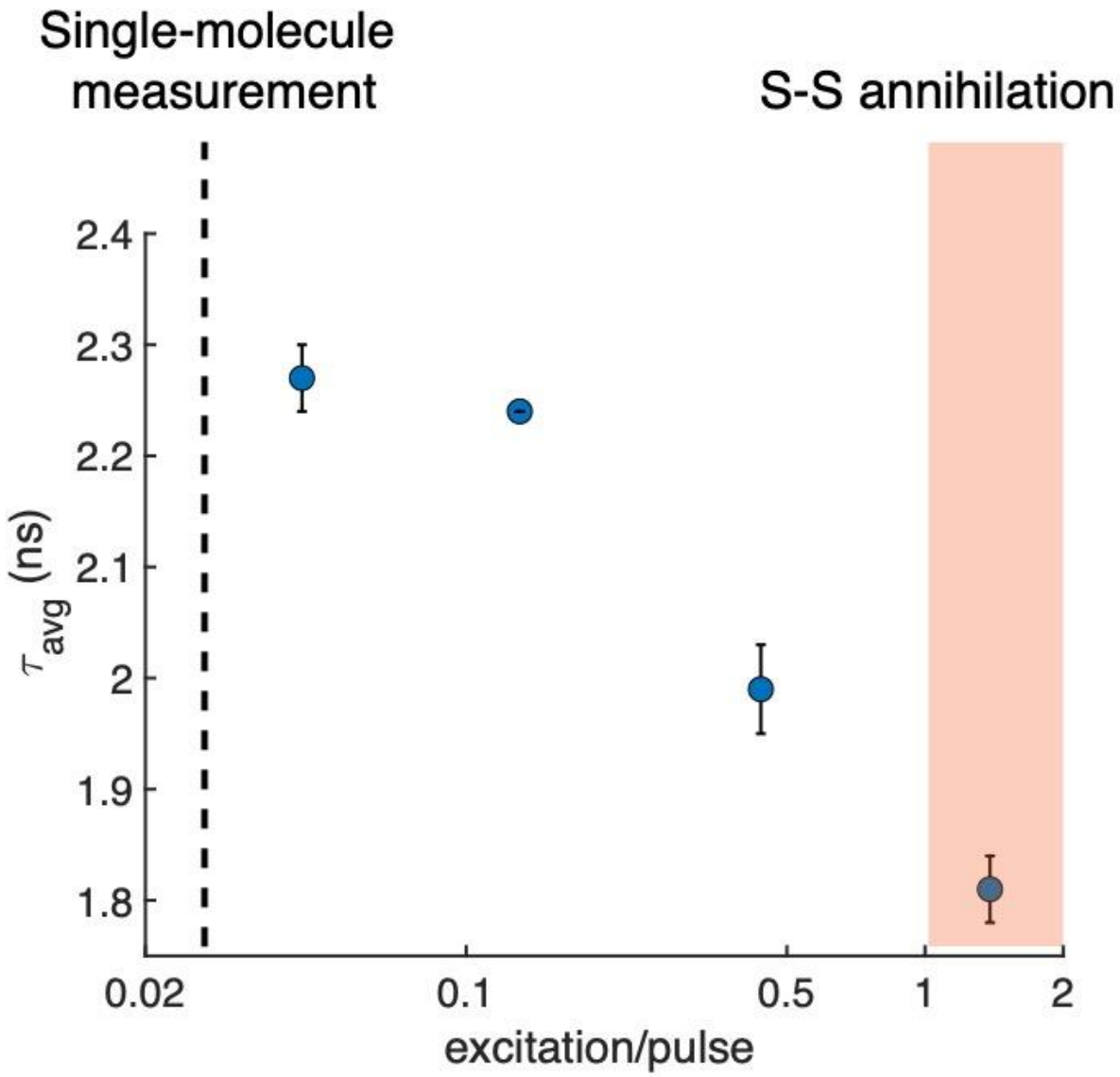


Figure S23: Average lifetime of LHCII proteoliposome with <N> = 5 at pH 7.5 as a function of excitation/pulse. The vertical dotted line shows the excitation/pulse used for single-molecule measurements. The orange-shaded region demonstrates the excitation/pulse > 1 where singlet-singlet annihilation is prominent. Error bars are the standard deviation.

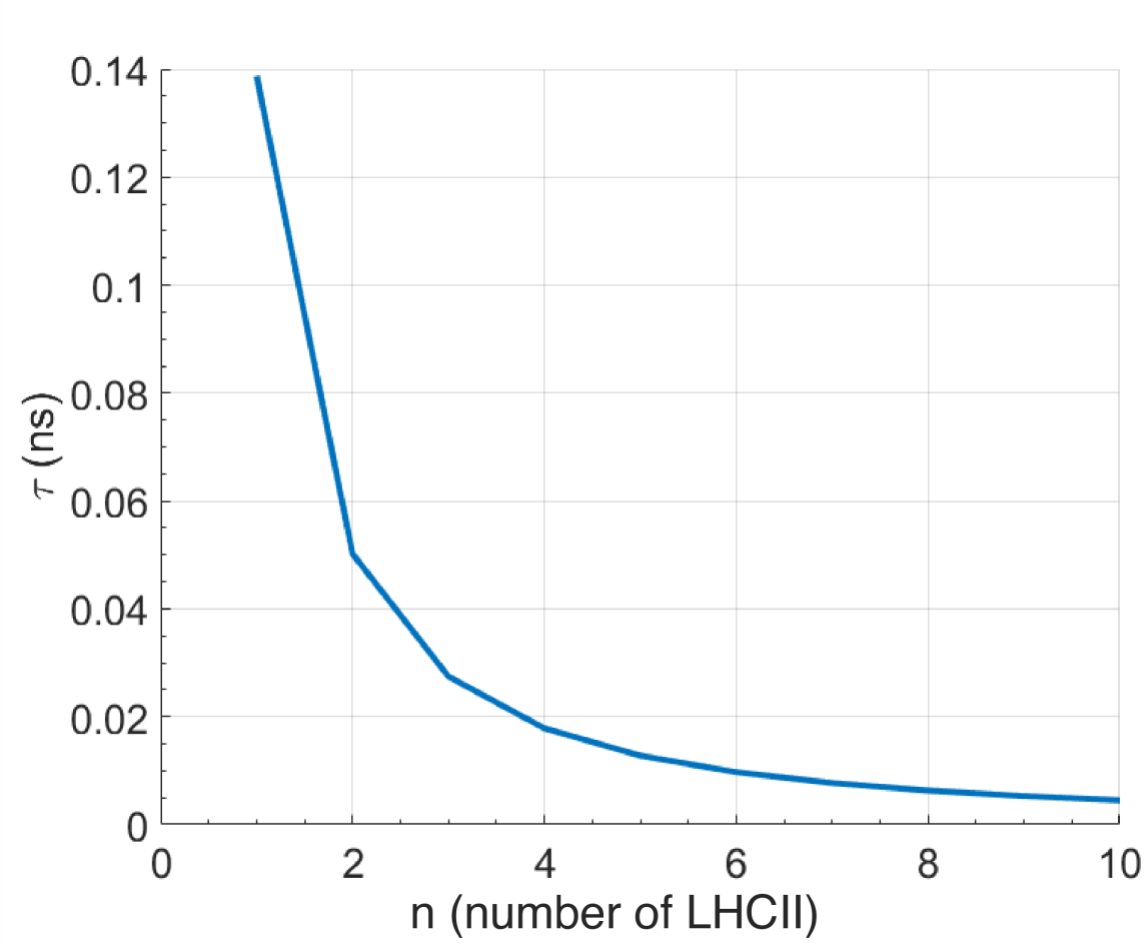


Figure S24: Excited state lifetime of LHCII aggregate as a function of cluster size computed from Eqn. S33.

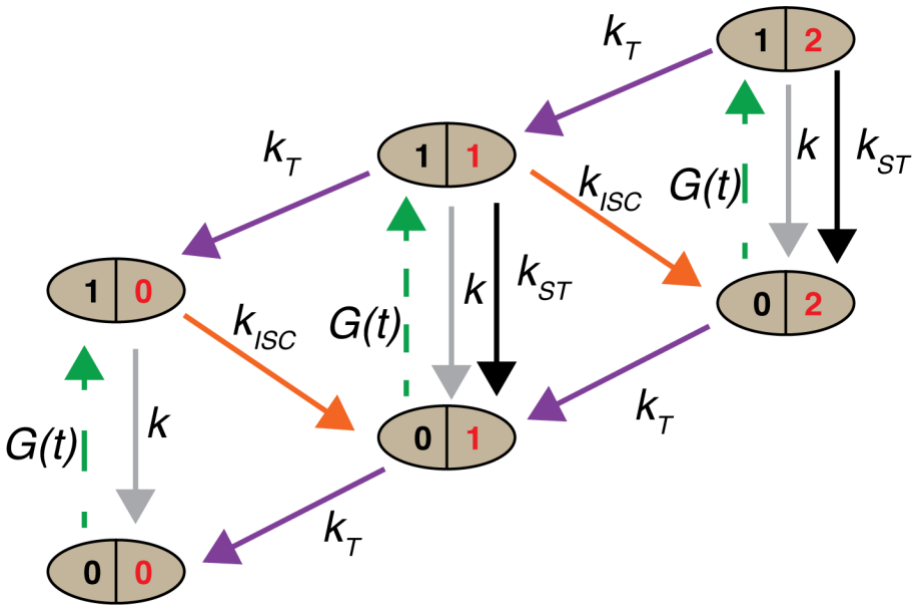


Figure S25: Stochastic model used to compute triplet state population. The ovals represent states with different numbers of singlets (black) and triplets (red) respectively. G(t) is the excitation rate; k and k_T are rate constants for linear de-excitation from singlet and triplet states, respectively. K_{ISC} and k_{ST} are the rate constants for inter-system crossing and ST annihilation, respectively. This figure is adapted from **(*27*)**.

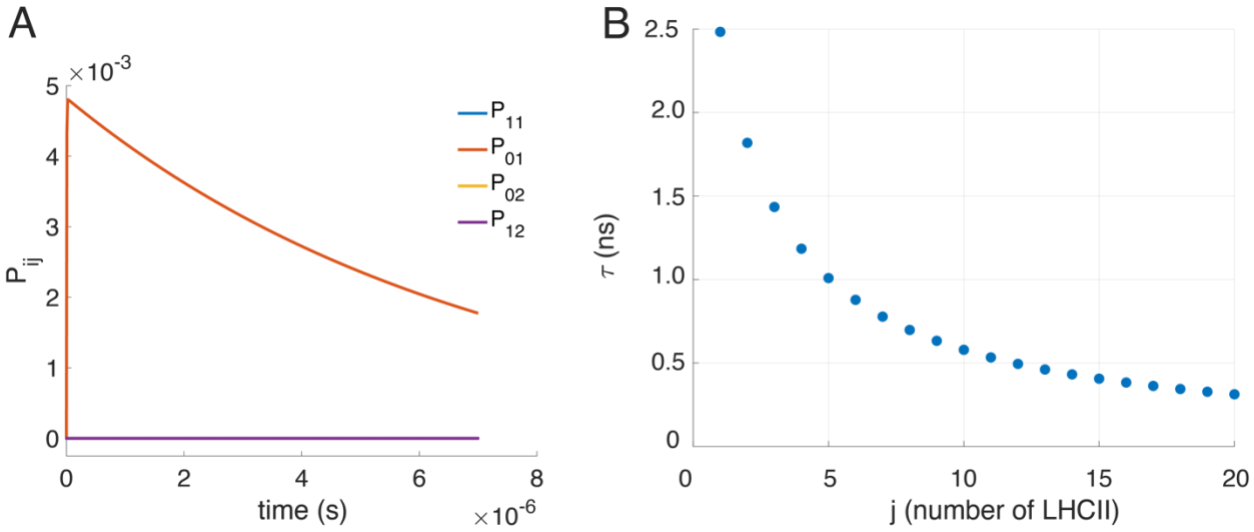


Figure S26: (a) Probability densities of states containing triplet populations as obtained from the numerical simulation. Except for P01, the contribution from other states is negligible. (b) Excited state lifetime of LHCII with different cluster sizes (j) as obtained from the modeling (Eqn. S33). The kinetic parameters used for the simulation are: $1/k_{ST}$ = 12 ps, 1/k = 5.81 ns, $1/k_T$ = 7 µs, $1/k_{isc}$ = 10 ns.

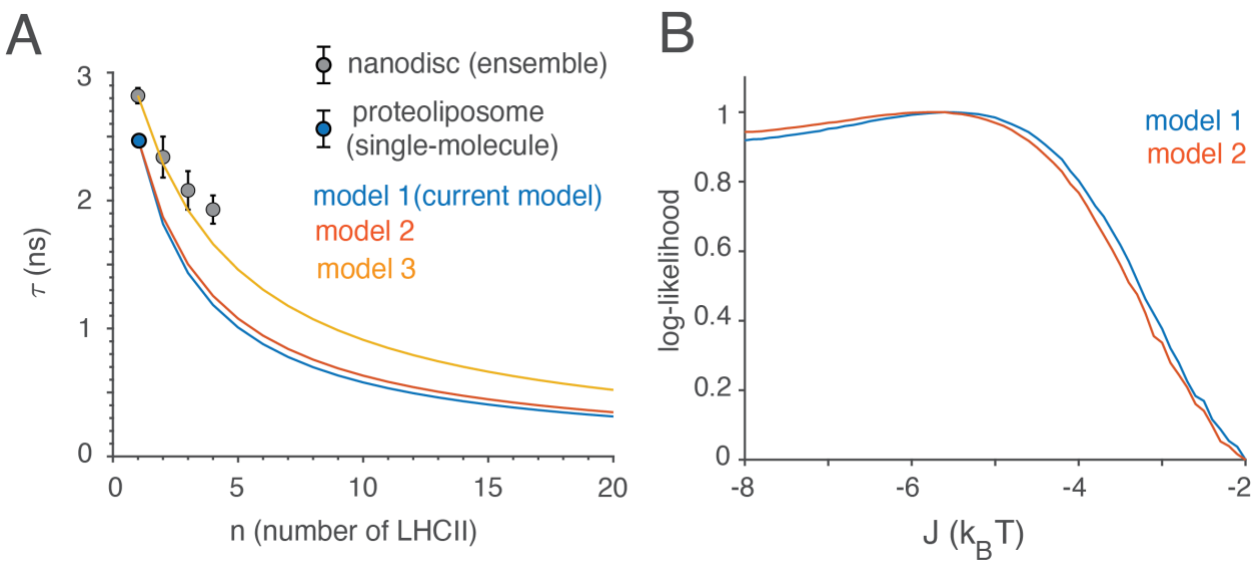


Figure S27: Alternative quenching pathway in LHCII clusters and the corresponding LHCII-LHCII interaction energy. (a) Excited state lifetime of LHCII aggregates ($\boldsymbol{\tau}$) as a function of cluster size (n). The circles represent the average lifetimes from LHCII-nanodisc (ensemble) and LHCII-proteoliposome (single molecule) at pH 7.5. Nanodisc data are obtained. from Son et al. (16). The error bars are the standard deviations. The lines represent the lifetime obtained using various models. Model 1 is used in the main text, i.e., $\boldsymbol{\tau = 1/(k + k_{isc} + n * k_{ST} * [T]_0)}$ where $\boldsymbol{k}$ ($\frac{1}{5.81\ ns}$), $\boldsymbol{k_{isc}}$ ($\frac{1}{10\ ns}$) and $\boldsymbol{k_{ST}}$ ($\frac{1}{12\ ps}$) are the rate constants for radiative, inter-system crossing and singlet-triplet annihilation, respectively. $[T]_0$ ($\boldsymbol{1.8 \times 10^{-3}}$) is the equilibrium triplet state population for our experimental condition as described in Sec S13. We adjusted $\boldsymbol{k_{ST}}$ from the previously reported (27) value of 1/36 ps to 1/12 ps to match our singly-loaded (<N>=1) LHCII-proteoliposome lifetime (~2.5 ns). For model 2, we used the expression, $\boldsymbol{\tau = 1/(k + k_{isc} + n * k_{ST} * [T]_0 + j * k_Q)}$. Here, we used ST annihilation rate as $\frac{1}{36\ ps}$ as reported by Gruber et al. (27) and added another quenching term $\boldsymbol{n * k_Q}$ where $\boldsymbol{k_Q}$ was taken to be $\frac{1}{12\ ns}$ ns to match singly-loaded (<N>=1) LHCII-proteoliposome lifetime (~2.5 ns) and n is the cluster size. Other parameters are the same as in model 1. For model 3, we used the same expression as model 2, i.e., $\boldsymbol{\tau = 1/(k + k_{isc} + n * k_{ST} * [T]_0 + n * k_Q)}$. Here, $\boldsymbol{k_{ST}}$ is kept as the published value ($\frac{1}{36\ ps}$) and $\boldsymbol{k_Q}$ is adjusted to $\frac{1}{31\ ns}$ to match ensemble-level lifetime (~2.8 ns) in LHCII-nanodisc sample with average number of proteins (<N>) as 1. (b) Normalized log-likelihood as obtained from the MLE fits for <N>=5 LHCII-proteoliposome sample at pH=7.5 using model 1 and model 2. The peak values from model 1 and 2 are -5.6 (-5.2 : -6.0) and -5.7 (-5.6 : -6.1) k_BT, respectively.

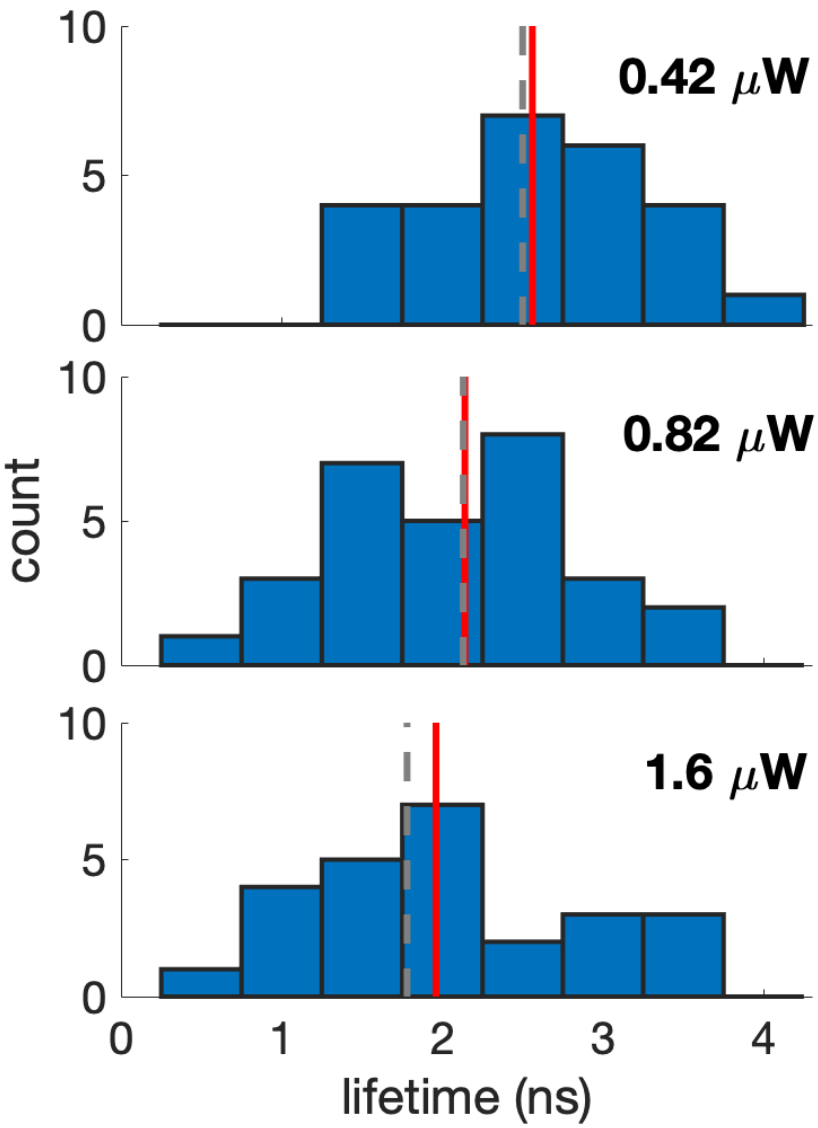


Figure S28: Lifetime distribution of single LHCII complexes immobilized in a PVA matrix at different laser power. The vertical red and dotted grey lines are the median of the distributions from the experiment and the simulation, respectively. The number of complexes recorded for each laser power are: 26, 29 and 25. The lifetime values are demonstrated in Table S7.

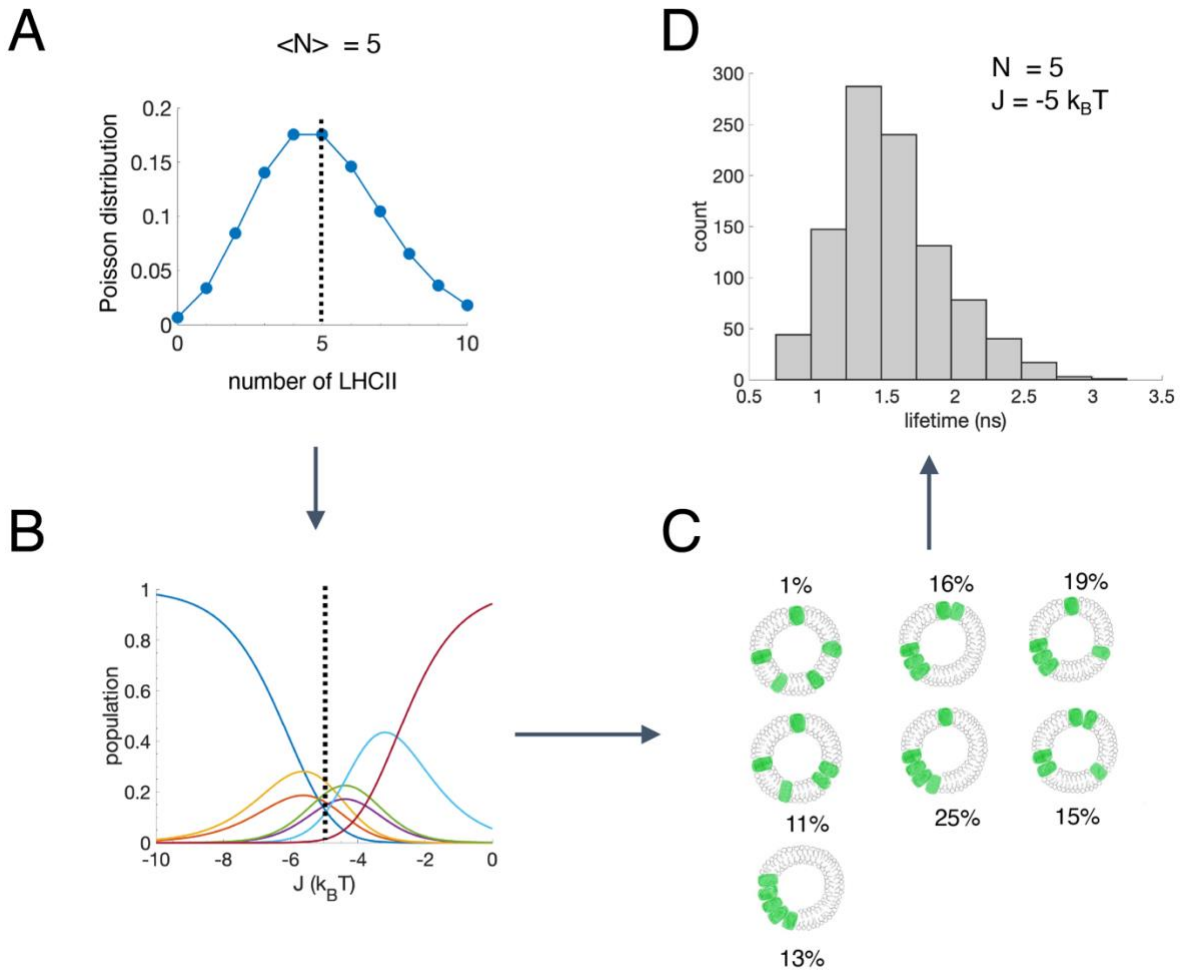


Figure S29: Extraction of LHCII-LHCII interaction energy from single-molecule lifetime data. (a) Distribution of the number of LHCII in liposomes for a sample containing on average five LHCII per liposome. In the first step of the simulation, the number of LHCII (N = 5 for example) selected based on this Poisson distribution (represented as a vertical dashed line) (b) Population of different configurations for N number of LHCII in liposome at different interaction energies as calculated in Section 11. (c) The relative population of different configurations and the lifetime of clusters based on the quenching model developed in Section 13 are used to compute the average lifetime of the sample at fixed J. Steps a-c are run many times to obtain a lifetime distribution for a specific N and J.

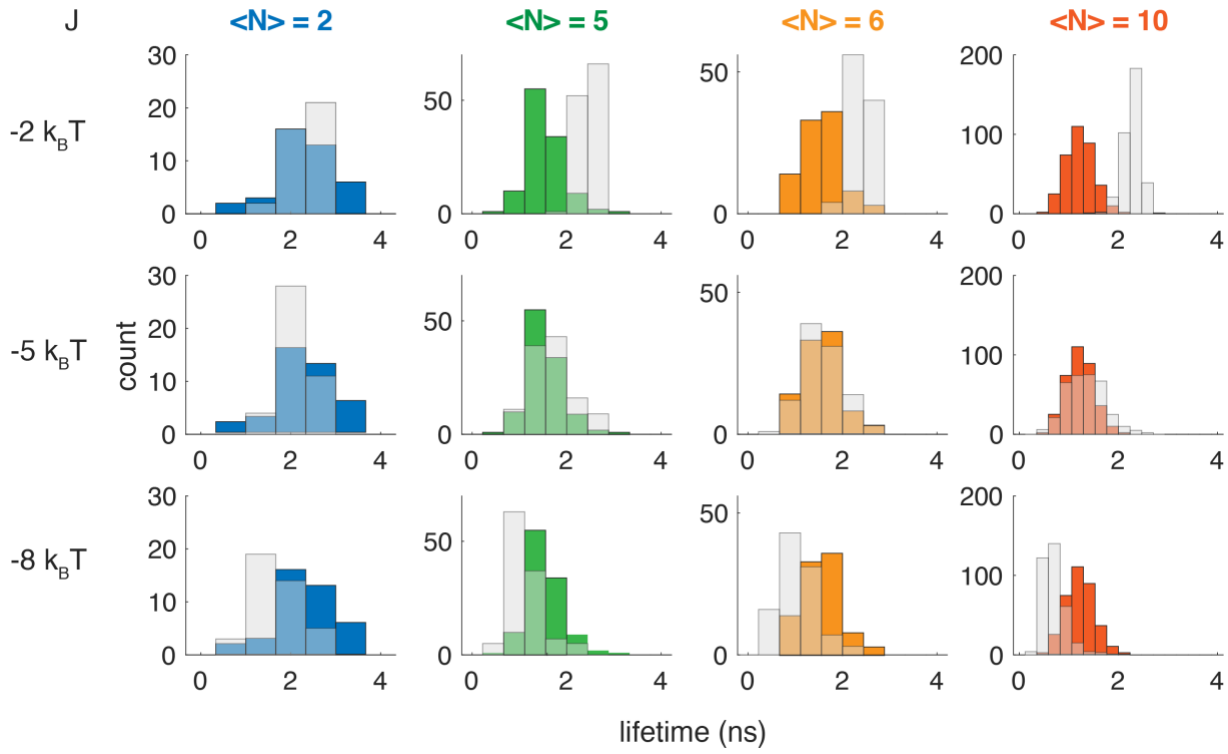


Figure S30: Comparison between experimental and simulated lifetime distribution at neutral pH. Lifetime distributions from single-molecule experiments at pH 7.5 are shown in blue, green, orange, and red histograms corresponding to <N> of 2, 5, 6, and 10, respectively. The simulated data (grey histograms) display the lifetime distributions at J = -2, -5, and -8 k_BT. The median lifetimes of the distributions are also shown in their corresponding figures.

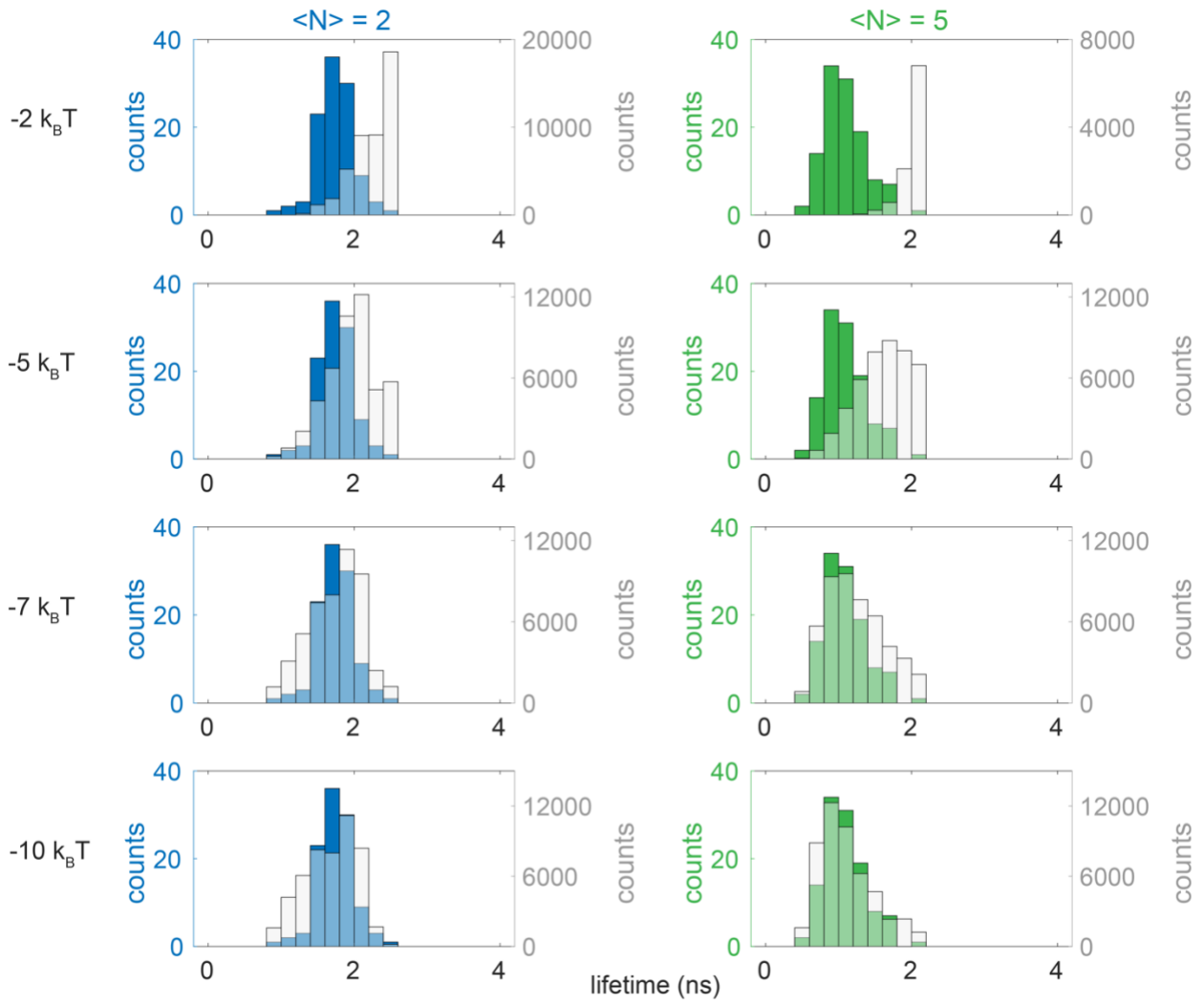


Figure S31: Comparison between experimental and simulated lifetime distribution at pH 5. The lifetime distributions from single-molecule measurements are shown in blue and green histograms for <N> = 2 and 5, respectively. The simulated lifetime distributions at J = -2, -5, -7, and -10 k_BT are displayed in grey histograms. At J = ~-7 k_BT. the simulated distribution matches well with the experimental ones.

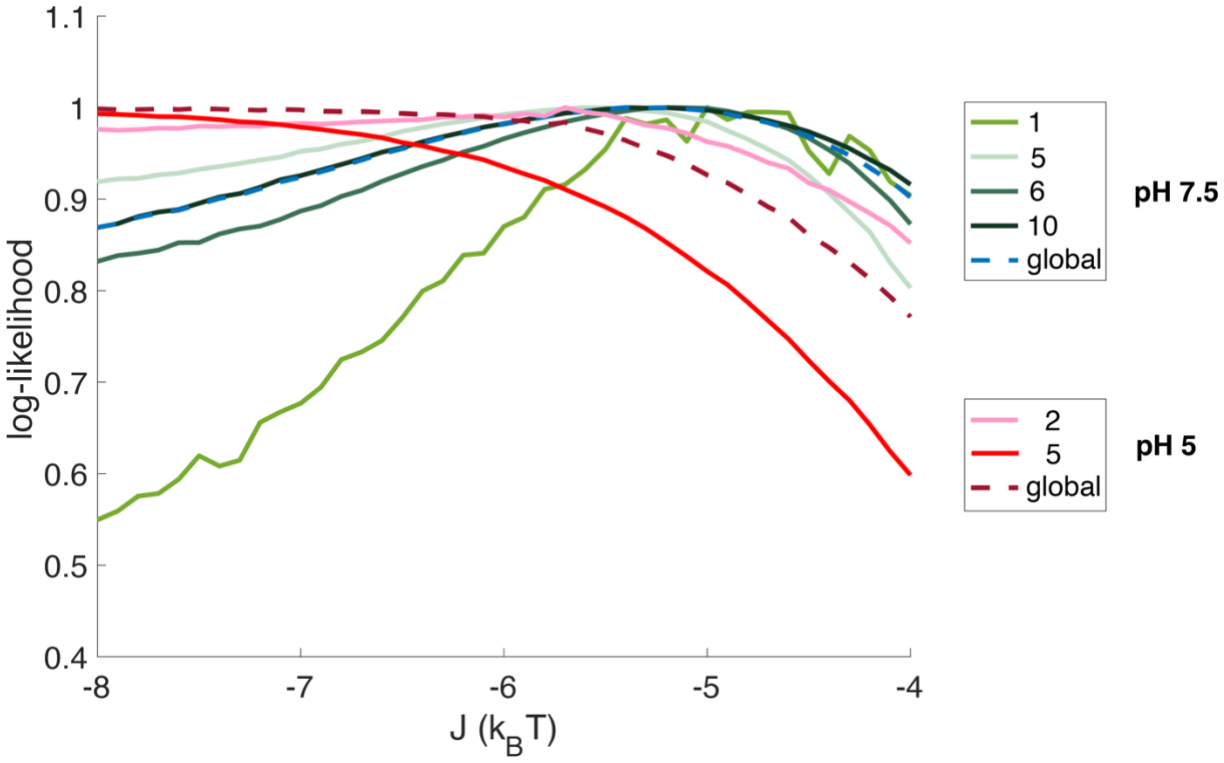


Figure S32: Normalized log-likelihood plots obtained from the MLE fits for the LHCII-proteoliposome samples at different protein densities (solid lines) at pH 7.5 and 5. The corresponding global fits are shown as dashed lines. The peak values of the log-likelihood are demonstrated in Table S8.

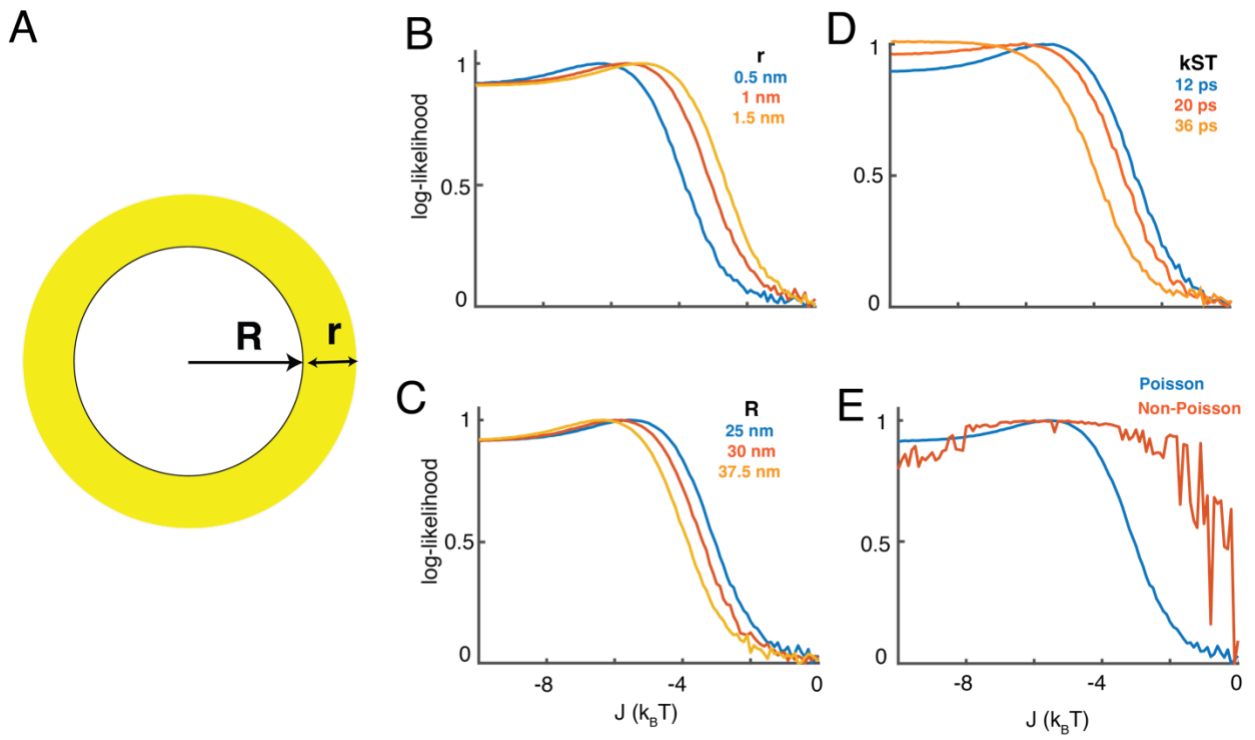


Figure S33: (a) The dependence of LHCII-LHCII interaction energy extracted from our model on the liposome radius (R) and interaction distance (r) is systematically investigated. Normalized log-likelihood plots at different (b) interaction distances (c) liposome sizes (d) S-T annihilation rate constants and (e) distribution types (Poisson and non-Poisson) are displayed. All the simulations are done for <N>=5.

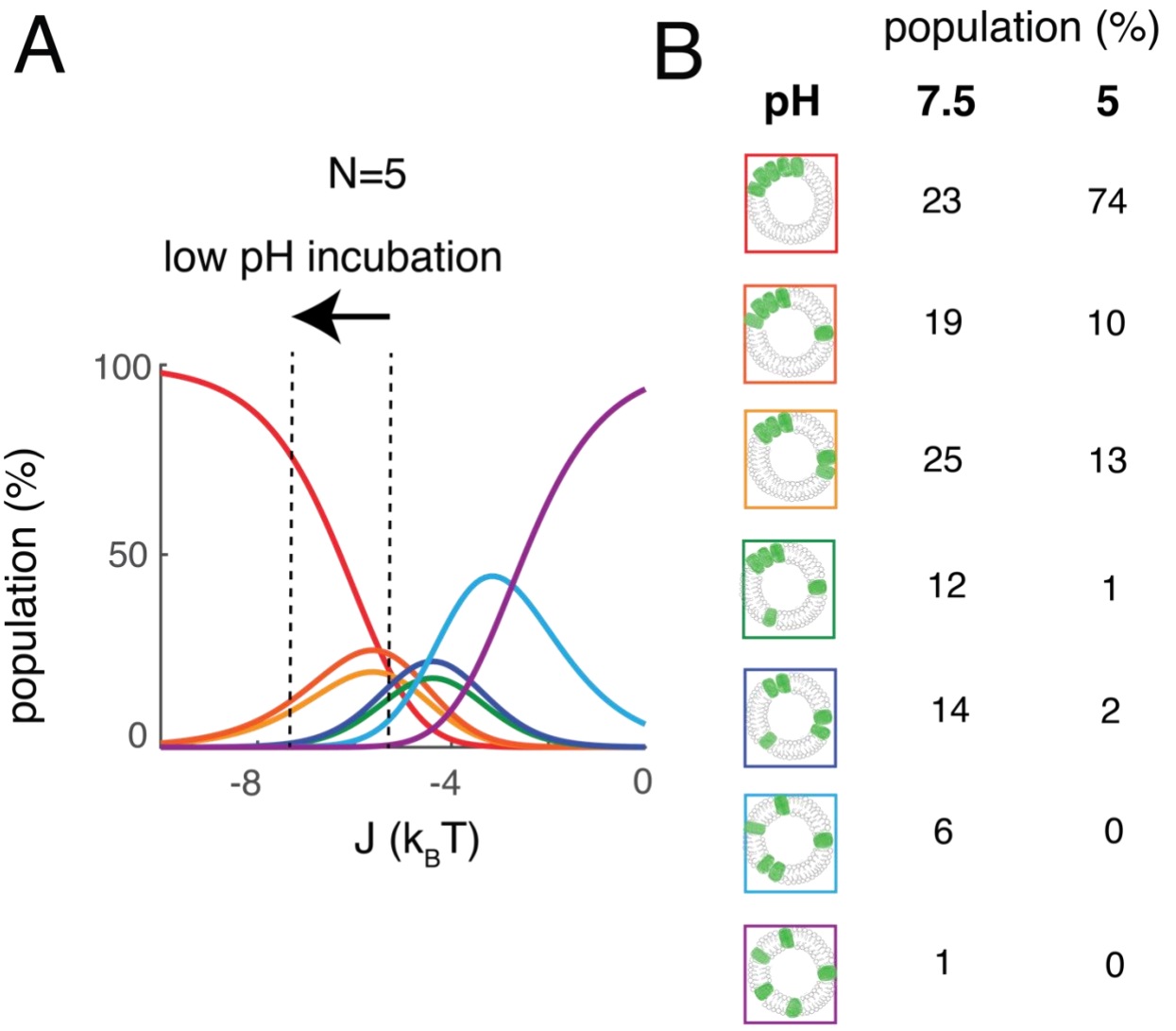
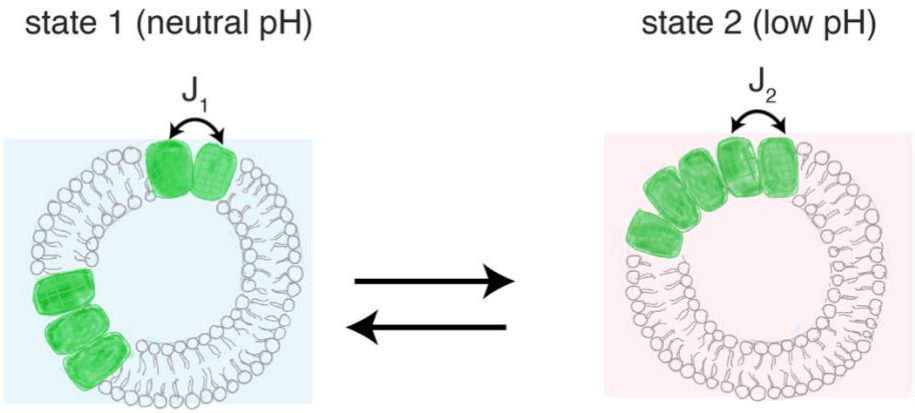


Figure S34: Shift in population configuration under low pH incubation. (a) Population of different configurations of the proteoliposome sample with N=5 as a function of LHCII-LHCII interaction energies (J). The arrow indicates the shift in J from -5.4 k_BT at pH of 7.5 to -7.2 k_BT at pH 5. (b) Cartoons representing the seven configurations for the N=5 sample and their population percentage under the pH 7.5 and 5. The cartoons of all the configurations are boxed in different colors matched to the population in (a). It is evident from the population percentage that although at neutral pH different configurations with clustered and unclustered proteins coexist, at low pH, the majority of the configurations are heavily clustered. For instance, the population of the configuration where all the proteins are clustered is 23% at pH 7.5. However, the population of this configuration is 74% once incubated at pH 5.



$$\Delta\Delta H = 4J_2 - 3J_1$$

$$\Delta\Delta S = k_B \log[w_2/w_1]$$

Figure S35: Schematic representation of pH-driven clustering in LHCII-proteoliposome. Based on the relative population of cluster configurations and interaction energies at neutral and low pH, the change in enthalpy (ΔΔH), entropy (ΔΔS), and therefore free energy change (ΔΔG) is calculated.

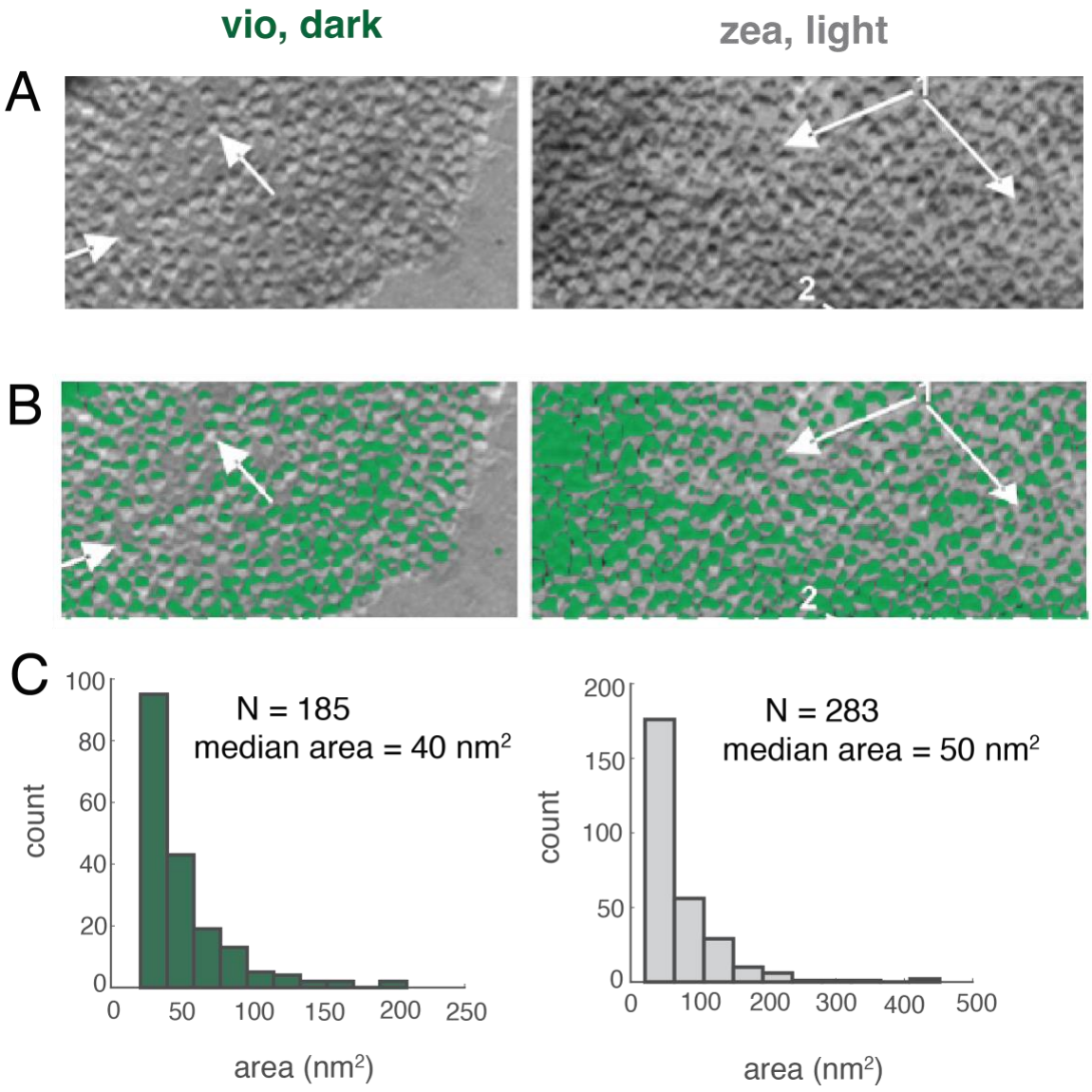


Figure S36: Image analysis of the freeze-fracture electron micrographs from the spinach chloroplast. (a). Freeze-fracture electron micrographs of the LHCII periplasmic fracture faces showing dark-adapted vio (left) and light-treated zea (right) from intact spinach chloroplast. This image is obtained from Johnson et al. **(*7*)** (Figure 5A, B). The arrows in the original image show the region of enhanced or reduced clustering of LHCII. (b) The original electron micrograph from **(*7*)** is cropped and the resulting image is analyzed using Image-Pro software (Media Cybernetics). A few regions of background and particles are identified and fed into the 'smart segmentation' feature of the software to produce a mask that returns the location and size of the LHCII clusters. The masks shown in green are overlayed with the electron micrographs. (c) The distribution of the LHCII cluster area is obtained from the segmented images of dark-adapted vio (left) and light-treated zea (right) samples. The median area of the light-treated zea sample is 25% (40 nm^2 vs. 50 nm^2) higher compared to the dark-adapted vio sample indicating an overall larger cluster size in the former sample. N = number of clusters in the segmented images.

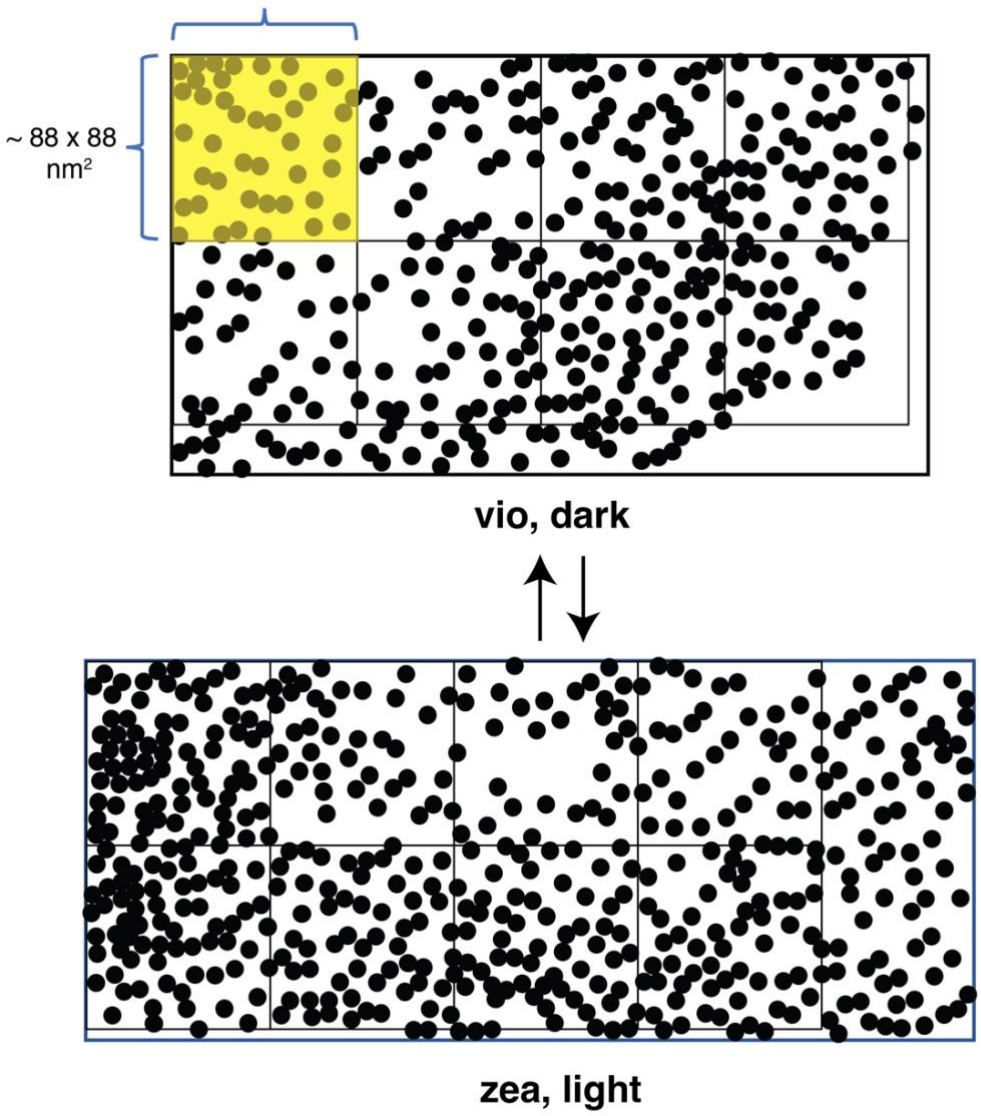


Figure S37: Free energy in light-driven clustering of LHCII in the chloroplast. Particle transforms from the segmented images of (a) vio-dark and (b) zea-light samples. The black-filled circles with an area of 50 nm^2 represent the LHCII trimers. The images are divided into square regions of ~ 88 nm × 88 nm to compute relevant thermodynamic parameters, i.e., the change in enthalpy, entropy and free energy. The computations from one such region (highlighted in yellow) are discussed in the text.

Table S1: Diameter of LHCII proteoliposomes with various protein-to-lipid ratios measured by dynamic light scattering. The numbers in the parenthesis are the standard deviations from independent measurements (n = 3).

protein:lipid	<N>	diameter (nm)
1:7000	<1	61 (1)
1:5000	1	60 (0.9)
1:2500	2	73 (0.9)
1:1000	5	79 (0.7)
1:880	6	96 (4)
1:510	10	99 (2)

Table S2: Ensemble lifetime of LHCII in liposome at various protein densities (<N>). The values in the parenthesis are the errors obtained from standard deviations of 2-3 technical replicates.

protein:lipid	<N>	pH	a_1 (%)	τ_1 (ns)	a_2 (%)	τ_2 (ns)	τ_{avg} (ns)
1:7000	<1	7.5	87	3.00	13	0.61	2.69
1:5000	1		81	3.24	19	0.24	2.67 *(0.10)*
1:2500	2		81	3.00	19	0.37	2.52 *(0.01)*
1:1000	5		78	2.70	22	0.61	2.24 *(0.08)*
1:880	6		78	2.64	22	0.44	2.17 *(0.18)*
1:510	10		76	2.37	24	0.35	1.88
1:5000	1	5.0	83	3.02	17	0.34	2.57 *(0.06)*
1:2500	2		73	2.65	27	0.39	2.04 *(0.06)*
1:1000	5		67	2.24	33	0.47	1.67 *(0.10)*

Table S3: Single-molecule lifetime of LHCII in liposome at various <N>. The values in the parenthesis are the errors obtained from standard deviations of the bootstrapped distributions.

protein:lipid	<N>	τ_{median} (ns), pH=7.5	τ_{median} , pH=5.0
1:7000	<1	2.29	2.02 *(0.02)*
1:5000	1	2.45 *(0.04)*	-
1:2500	2	2.27 *(0.08)*	1.75 *(0.02)*
1:1000	5	1.47 *(0.04)*	0.96 *(0.02)*
1:880	6	1.53 *(0.04)*	-
1:510	10	1.22 *(0.02)*	-

Table S4: Solutions of the Equation S10 with N = 5. The resultant m-matrix represents all the possible cluster configurations for this N with their geometric connectivity.

m_1	m_2	m_3	m_4	m_5
0	0	0	0	1
0	1	1	0	0
1	0	0	1	0
1	2	0	0	0
2	0	1	0	0
3	1	0	0	0
5	0	0	0	0

Table S5: The probability (w) of different configurations as computed from randomly distributing LHCII into liposome surface. The number of iterations used in the simulations is $10^7 - 10^9$.

configuration	N =2	N=3	N=4	N=5	N=6
a	0.006	6×10^{-5}	1.2×10^{-6}	2.0×10^{-8}	$<1.3 \times 10^{-9}$
b	0.994	0.017	99×10^{-6}	3.6×10^{-6}	4.9×10^{-8}
c		0.98	25×10^{-5}	5.5×10^{-6}	1.0×10^{-7}
d			0.033	0.0005	2.7×10^{-6}
e			0.97	0.0006	1.5×10^{-7}
f				0.055	2.2×10^{-5}
g				0.94	0.0014
h					0.0012
i					0.080
j					0.92

Table S6: Comparison of the probability of the configurations (w) obtained from analytical theory and random sampling code.

N = 2		N = 3		N = 4		N = 5		N = 6	
theory	sim	theory	sim	theory	sim	theory	sim	theory	sim
0.60	0.60	0.007	0.006	1.1e-4	1.2e-4	2.2e-6	2.0e-6	5.6e-8	<1.2e-7
99.4	99.4	1.807	1.686	11e-3	9.9e-3	4.1e-4	3.6e-4	4.6e-6	4.9e-6
		98.18	98.31	0.027	0.025	5.3e-4	5.4e-4	0.95e-5	1.0e-5
				3.615	3.346	0.05	0.05	3.3e-4	2.7e-4
				96.34	96.62	0.07	0.06	1.3e-4	1.5e-4
						6.02	5.50	2.4e-3	2.2e-3
						93.9	94.4	1.6e-3	1.6e-3
								0.16	0.14
								0.13	0.12
								9.04	8.10
								90.7	91.2

Table S7: Median excited state lifetime of LHCII (immobilized in a PVA matrix) obtained from the experiment and simulation at different laser power.

power (μW)	τ_{median} (ns, experiment)	τ_{median} (ns, simulation)
0.42	2.56	2.50
0.82	2.13	2.13
1.6	1.96	1.78

Table S8: LHCII-LHCII interaction energies (J) obtained from MLE fits. ND = not determined.

	J	J
<N>	pH,7.5	pH, 5.0
2	-5	-6.6
5	-5.6	-7.4
6	-5.2	ND
10	-5.4	ND
global	-5.4	-7.2

Table S9: Thermodynamic parameters for pH-driven aggregation in LHCII-proteoliposomes and chloroplast.

	<N>	$\Delta\Delta H(k_BT)$	$T\Delta\Delta S(k_BT)$	$\Delta\Delta G(k_BT)$
liposome	2	-3.6	-1.6	-2.0
	5	-12.7	-4.5	-8.2
	6	-15.5	-5.13	-10.4
	10	-26.2	-7.2	-19.0
chloroplast	50	-115		