Title
Pattern formation in photosynthetic membranes: a physical and statistical approach

Author
Schneider, Anna Rachel

Publication Date
2013

Peer reviewed|Thesis/dissertation

Pattern formation in photosynthetic membranes: a physical and statistical approach

By

Anna Rachel Schneider

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Biophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Phillip L. Geissler, Chair
Professor Jan T. Liphardt
Professor Naomi S. Ginsberg
Professor Berend Smit

Fall 2013

Pattern formation in photosynthetic membranes: a physical and statistical approach

Abstract

Pattern formation in photosynthetic membranes: a physical and statistical approach

by

Anna Rachel Schneider

Doctor of Philosophy in Biophysics

University of California, Berkeley

Professor Phillip L. Geissler, Chair

Photosystem II (PSII) and its associated light-harvesting complex II (LHCII) are highly concentrated in the stacked grana regions of photosynthetic thylakoid membranes from plants and green algae. PSII–LHCII supercomplexes can be arranged in disordered packings, ordered arrays, or mixtures thereof. The physical driving forces underlying array formation are unknown, and statistically robust methods of identifying arrays in micrographs are lacking, complicating attempts to determine a possible functional role for arrays in regulating light harvesting or energy conversion efficiency. This dissertation introduces new computational tools for studying the nano- and mesoscale organization of the membrane proteins that comprise the photosynthetic apparatus, and applies them to illuminate the origins of the diversity of local structural motifs in this adaptive biomaterial. First, we introduce a coarse-grained model of protein interactions in coupled photosynthetic membranes, focusing on a small number of particle types that feature simple shapes and potential energies motivated by structural studies. Reporting on computer simulations of the model's equilibrium fluctuations, we demonstrate its success in reproducing diverse structural features observed in experiments, including extended PSII–LHCII arrays. Free energy calculations reveal that the appearance of arrays marks a phase transition from the disordered fluid state to a system-spanning crystal. The predicted region of fluid-crystal coexistence is broad, encompassing much of the physiologically relevant parameter regime. Our results suggest that grana membranes lie at or near phase coexistence, conferring significant structural and functional flexibility to this densely packed membrane protein system. Upon extending this model to include simple representations of the membrane morphology and protein components found in thylakoid margins and stroma lamellae, we find that LHCII stacking and crowding in the grana are also primary determinants of the segregation of protein components between stacked and unstacked regions of the thylakoid membrane. In addition, we develop a statistical pipeline for identifying PSII crystals in nanometer-resolution micrographs of thylakoid membranes that would assist in future experimental tests of the predictions put forth in this dissertation; we validate our method on atomic force microscopy measurements of grana membranes isolated

from *Arabidopsis thaliana* wild-type and *soq1* mutant plants. As a whole, this work creates a foundation for future rigorous biophysical investigations of the structure, function, and dynamics of the photosynthetic apparatus.

To Mom, Dad, and all my other teachers.

Contents

List of Figures

List of Tables

Acknowledgments

It's probably gauche to quote Twitter in one's thesis, but I'm going to do it anyway. `@tressiemcphd`, in reference to academia: "I was a person when I got here. I'll be one when I leave." Grad school has given me the opportunity to explore many different ways to be a person, each aligned in a different way with the person I've been and the person I want to be. I have been fortunate to find so many nurturing, vibrant communities at Berkeley, and to maintain my connections elsewhere.

First thanks, of course, go to Mom and Dad: Not just for your 2.8 decades of love and support, or for the fact that I get to have more fun and substantive conversations about science and academia with my parents than almost anyone else I know does—in terms of things that have helped me through grad school, I might be most grateful that you raised me to believe in myself. And to Sarah and Daniel, for all the growing up together that we're still doing.

To my adviser, the great Phill Geissler: Thank you for making statistical mechanics a thing of excitement and beauty, in the classroom and in subgroup, by always striking the perfect balance of rigor, relevance, and panache. I am a better thinker for having seen the world through your scientific eyes. And I deeply appreciate your kindness and trust; I hope I've repaid it.

To my photosynthesis collaborators, especially Kapil Amarnath, Bibiana Onoa, Naomi Ginsberg, Helmut Kirchhoff, and the rest of the SISGR team: Thanks for all the long conversations, insightful questions, and complementary hard work. This project would not have been conceivable without your community. Thanks also to Jan Liphardt and Berend Smit for helpful comments and for hosting me early on.

To my boyfriend Lester Hedges, for your love, generosity, and encouragement. And to Laura Kogler, for being a consistently wonderful housemate for almost as long as I've ever lived anywhere.

To everyone who's come through the Pitzer Center over the years, especially Chris Ryan, Pat Shaffer, Ayelet Benjamini, Aaron Keys, Michael Grünwald, Katie Klymko, Todd Gingrich, Suri Vaikuntanathan, Patrick Varilly, Yael Elmatad, John Haberstroh, Toni Mey, Mike Levy, Alex Hudson, Kelsey Schuster, Dayton Thorpe, David Limmer, Jocelyn Rodgers, and Lutz Maibaum: It's been a joy to share the everyday excitements and challenges of science with you. I'll miss having such a receptive audience for baked goods. Special thanks to the members of the self-assembly and interfaces subgroups for all your helpful suggestions over

the years. And to Lisa Littlejohn, for caring so much and keeping us chugging along.

To John Chodera, Laura Rosen, Jaime Yassif, and Alan Lowe: Thank you for weaving me into your lives and letting me weave you into mine. Every step along the way was improved by your unlimited supply of support, inspiration, hugs, and craft beverages.

To the crew at the Berkeley Science Review: You are high-caliber humans, one and all, and it's been a pleasure crafting stories together. And to my fellow biophysics students and to the program leadership: It's true what we tell the recruits—the people are what make this such a wonderful program. There's nowhere I'd rather have done my PhD.

To all my Seattle lovelies, especially David Sunderland, Michelle Burce, Brandi Hair, Brandon Martin-Anderson, and the Librarians and Ministers: It's an honor every single time to get to make myself at home in your homes, celebrate at your weddings, and tag along to your adventures, after all these years of sporadic hellos. Next year in Wabi Town?

And last but very much not least, to Gavin McCormick, for helping me find the next ways to do better.

Chapter 1

Introduction

Photosynthesis powers life on Earth by converting a fluctuating supply of solar energy into chemical energy in the form of biomaterials that we use for food and fuel. Statistical physics probes the partitioning of energy and entropy, of order and disorder, and of work and dissipation in fluctuating systems. This dissertation marks one of the first attempts to bridge these two disparate fields while making significant contributions to both. Such interdisciplinary work is rewarding but challenging, both for the researcher and the reader. Therefore, to make your life easier, this introduction is divided into several sections targeted at different audiences: the first for statistical mechanicians looking for biological context, the second for photosynthesis biologists who want to learn modeling methods, and the third for those interested in previously published work at the intersection of these fields.

1.1 Molecular photosynthesis for soft matter physicists

1.1.1 Preliminaries

First, a note on scope. This dissertation is specifically focused on the molecular structures in plants that perform the very first steps of photosynthesis, namely light harvesting, exciton transport, charge separation, and electron transport—in other words, the part of a plant that acts like a solar panel. When I use the term the "photosynthetic apparatus," I refer to this structural and functional unit. Downstream processes in photosynthesis, e.g., carbon capture, are outside the scope of this dissertation, as are the structurally-divergent photosynthetic apparatuses in various photosynthetic bacteria (although these topics are also fascinating). Because the photosynthetic apparatus of green algae like *Chlamydomonas reinhardtii* is highly but not completely homologous to that of plants like the model organism *Arabidopsis thaliana*, much but not all of this material generalizes to green algae as well.

Also, a note on language. Photosynthesis involves a long list of biomolecules and biochemical processes that have been the subject of decades of study by structural and cell biologists. Therefore, whether we like it or not, the terminology in this field is dominated by alphabet soup and Greek morphemes. To get to the interesting questions, one is well served by a bit

of patience and stamina on this front.

For those looking to enter the field of photosynthesis, I especially recommend the ubiquitously cited review by Dekker and Boekema for a thorough introduction [48], the January 2012 special issue of Biochimica et Biophysica Acta (BBA) - Bioenergetics (Volume 1817, Issue 1) for a collection of modern reviews of many important topics, and the very recent review by Antal et al. for tabulations of useful numerical constants [9].

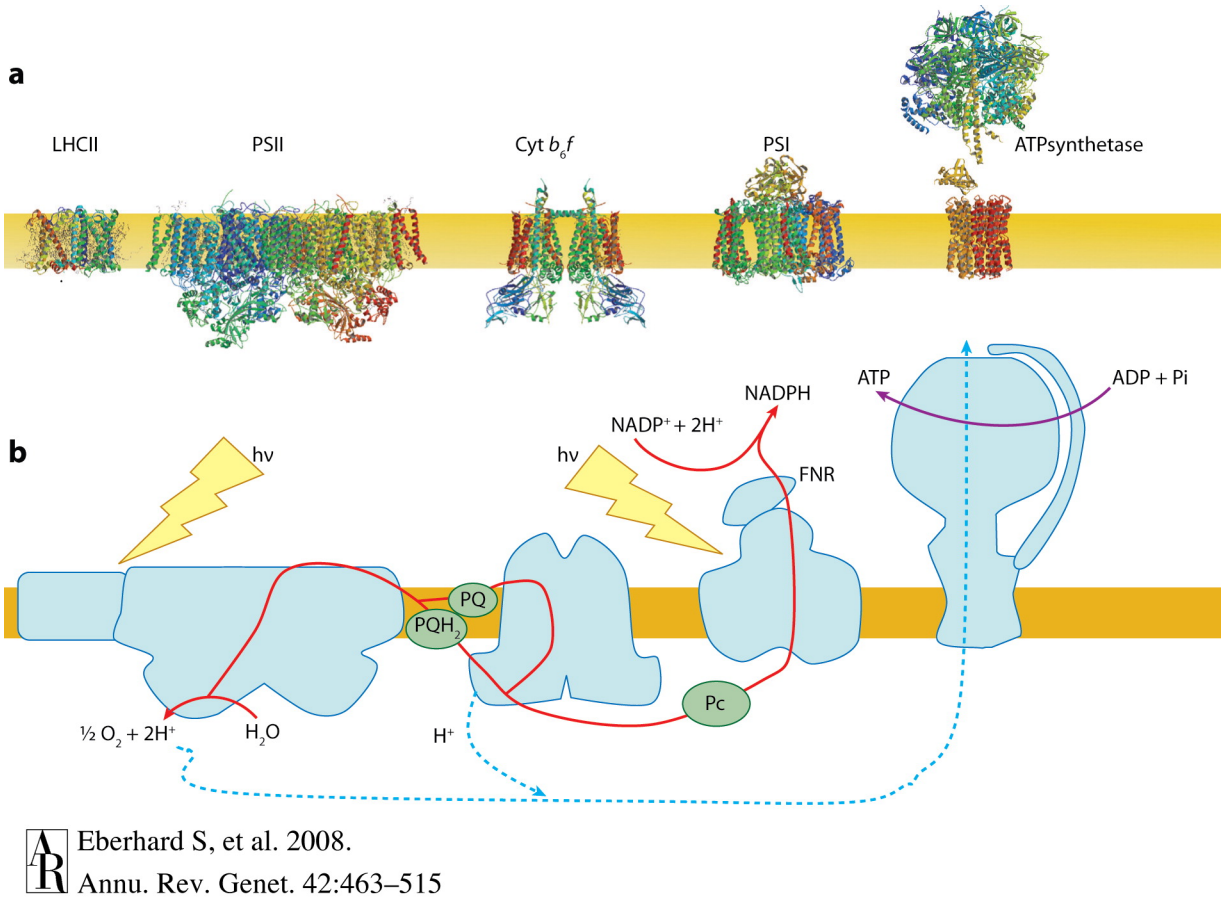
1.1.2 Protein structure and function

Photosynthesis is a driven four-photon process that consumes H_2O and produces O_2, ATP, and NADPH. There are five main protein complexes that carry out this process: light harvesting complex II (LHCII), photosystem II (PSII), cytochrome b_6f (cyt b_6f), photosystem I (PSI), and ATP synthase (Figure 1.1). All of these proteins are integral membrane proteins embedded in the thylakoid membrane inside chloroplasts, with distinct faces on the lumen (interior) and stroma (exterior) sides of the membrane.

LHCII is one of the most abundant membrane proteins in the biosphere [15]. The primary function of LHCII is to scaffold chlorophyll *a* and *b* pigments that become excited by visible photons and transfer that excitation energy toward photosystems. Because LHCII pigment-protein complexes increase the functional absorption cross-section of the photosynthetic apparatus, they are sometimes called "antenna" complexes. LHCII commonly refers to the "major" light harvesting complex, a trimer of Lhcb1, Lhcb2, and/or Lhcb3 gene products; other members of the LHC family are termed "minor" complexes (Lhcb4, Lhcb5, and Lhcb6 yield CP29, CP26, and CP24, respectively). X-ray crystal structures show that the LHCII trimer is roughly cylindrical with a diameter of approximately 6.5 nm in the plane of the membrane [120, 166]. The stromal face of LHCII features a negatively charged surface and positively charged N-terminal tails [166]. In vitro, high concentrations of LHCII form lamellar aggregates in the presence of various lipids or detergents, and these lamellae stack in the presence of divalent cations [161, 132].

PSII houses the oxygen-evolving complex (OEC), which splits water into protons and molecular oxygen via catalysis by a Mn_4CaO_5 cluster, and uses the extracted electrons to reduce the membrane-soluble electron carrier plastoquinone (PQ). The atomic structure of cyanobacterial PSII has been resolved in increasing detail [193, 94, 57, 175, 97], leading to advances in our understanding of the catalytic cycle [68]. The PSII core is a large complex composed of the D1 and D2 proteins, which perform charge separation and are essential for water splitting; CP43 and CP47, which form the core antenna; the extrinsic proteins of the OEC, which protrude on the lumenal side of the thylakoid membrane; and dozens of smaller polypeptides [158]. PSII is often found as a dimer and the two lobes of the OEC protrusion, separated by ≈9 nm, allow PSII to be easily identified in electron microscopy (EM) and atomic force microscopy (AFM) images of thylakoid membranes [101, 167, 35].

The protons and electrons separated by PSII follow different paths thereafter. The first step of linear electron transport is diffusion of plastoquinone from PSII to cyt b_6f, which transfers electrons to the water-soluble electron carrier plastocyanin (Pc); cyt b_6f also acts



Eberhard S, et al. 2008.
Annu. Rev. Genet. 42:463–515

Figure 1.1: Structure and function of the primary proteins in photosynthetic electron transport. Both panels show side views **a**: ribbon diagrams of crystal structures. **b**: cartoons showing linear electron flow through the system. The yellow band represents the lipid bilayer of the thylakoid membrane. Reproduced from [54].

as a sensor of the redox state of the membrane [114] and plays a key role in cyclic electron transfer [92]. Next, plastocyanin diffuses through the lumen to PSI, a 20 nm × 16 nm complex with several monomers of light harvesting complex I (LHCI) [24], which performs photon-catalyzed charge separation and transfers electrons to ferredoxin, ultimately reducing $NADP^+$ to NADPH [34]. Protons, in contrast, accumulate in the lumen to form a proton motive force across the thylakoid membrane, which powers F_1F_0 ATP synthase, a monomeric rotary motor with a head diameter of 16 nm [45, 127].

1.1.3 Membrane morphology

The thylakoid is a complex folded membrane system that partitions aqueous space inside the chloroplast into a single continuous interior lumen and an exterior stroma. This

partitioning is essential for photosynthetic function because it permits the formation of the proton gradient that powers ATP synthase (Figure 1.1). The volume of the aqueous stroma is much larger than that of the lumen, and ribosomes, RuBisCo, and most other globular chloroplast proteins reside in the former. Most thylakoid lipids are galactolipids, approximately half of which are the non-bilayer-forming lipid MGDG [188].

The thylakoid membrane is differentiated into stacks of flat discs of tightly appressed membrane called grana, tubes or sheets of unappressed membrane called stroma lamellae, and highly curved junctional regions called margins (Figure 1.2). Grana stacks have typical diameters in the range 300–600 nm and a vertical periodicity in the range 14–21 nm [48]. This repeat distance implies that the stromal faces of grana lamellae are separated by no more than a few nanometers of water, and that there are steric conflicts between PSII OECs in the lumen of at least some grana stacks [104]. Driving forces for three-dimensional stacking include electrostatics [14], interlayer interactions that can be disrupted by phosphorylation of LHCII or PSII [39, 87, 62, 79], protein-induced membrane curvature [10], and other generic membrane–membrane interactions [38]. The relative contributions of these interactions to thylakoid ultrastructure are not known [38].

Several groups have published electron cryo-tomography (ET) studies that put forth competing models for the overall topology of the thylakoid [159, 33, 134, 64, 45, 12]. While most ET studies try to interpret their findings within the context of a perfectly periodic model, Austin and Staehelin's work is notable for acknowledging the role of fluctuations, for example by enumerating the variability in the size of the junctions between stroma lamellae and grana membranes [12]. Consensus is emerging around variations of the basic helical model originally proposed in [142], although the structural details of the margins remain under debate [44, 137].

In the helical model, stroma lamellae wind around the grana stack in a right-handed helix. To the best of my knowledge, neither the physical nor the biochemical origins of this surprising chirality have been investigated. I speculate that clues may come from the so-called chiral macrodomains in grana membranes that have been studied by Garab and coworkers using circular dichroism (e.g., [80]), or from the cubic membrane structures called prolamellar bodies that can occur at the etioplast stage of thylakoid biogenesis (e.g., [156]), but these phenomena are not well understood at the nanoscale either. Overall, many open questions remain about thylakoid morphology and morphogenesis [2].

Isolated grana membranes are common model systems for the full thylakoid membrane. Grana can be isolated away from stroma lamellae by solubilization with various detergents: the original BBY protocol used Triton X-100 [22], while newer protocols use milder detergents such as α-DM in the hope of better preserving the protein organization within the grana [131]. Isolated grana are typically observed as stroma-associated pairs of lipid bilayers whose lumenal faces can be probed by AFM [103, 167]. Because pairing occurs between the "exterior" stromal faces and leaves the "interior" lumenal faces exposed, isolated grana are sometimes called "inside-out membranes;" this also suggests that net stroma-side attractions are stronger than net lumen-side attractions.

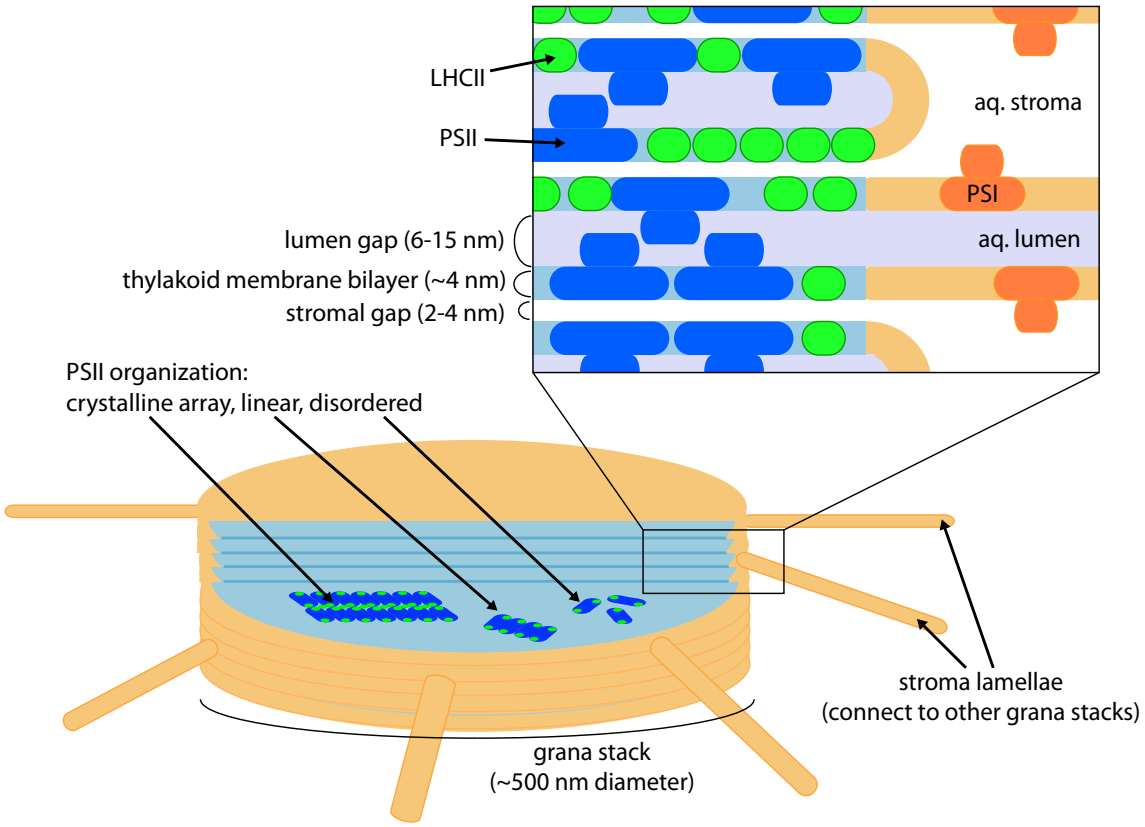


Figure 1.2: Cartoon of thylakoid membrane and protein organization. Inset design inspired by [48].

1.1.4 Flexible hierarchical assembly

PSII and LHCII exhibit self-assembled organization on a range of length scales [48, 137]. At the macromolecular scale, dimers of PSII core complexes associate specifically with 1–6 trimers of LHCII and up to 2 monomers of each of the minor light harvesting complexes (CP24, CP26, and CP29) to form a family of PSII supercomplexes that can be isolated by sucrose gradients and observed by single-particle electron microscopy [26, 35, 172]. PSII supercomplexes are named for the number of PSII core and LHCII trimer complexes they include: for instance, the large $C_2S_2M_2$ supercomplex includes a dimer of PSII Core complexes (red in Figure 1.3), two Strongly bound LHCII trimers (green), and two Moderately bound LHCII trimers (black) (as well as the minor antenna complexes that do not contribute to the naming scheme). The nomenclature of "strong" and "moderate" binding is due to the apparent hierarchical nature of assembly, and not due to direct determination of binding free energies in experiment or simulation. The presumed typical order of binding is (i) PSII core dimerization, then (ii) binding of CP26, CP29, and an LHCII trimer at the "S" position, then (iii) binding of CP24 and an LHCII trimer at the "M" position, although exceptions like C_2M

complexes can be found. In the model plant *Arabidopsis thaliana*, the $C_2S_2M_2$ complex is the most commonly observed, while smaller C_2S_2 and C_2S_2M complexes are more common in spinach [109]. Although CP24 appears to mediate LHCII binding at the M position in plants, the green alga *Chlamydomonas reinhardtii* can bind M complexes although it lacks the gene for CP24; *Chlamydomonas* PSII supercomplexes can also contain rarer loosely bound (L) complexes [172].

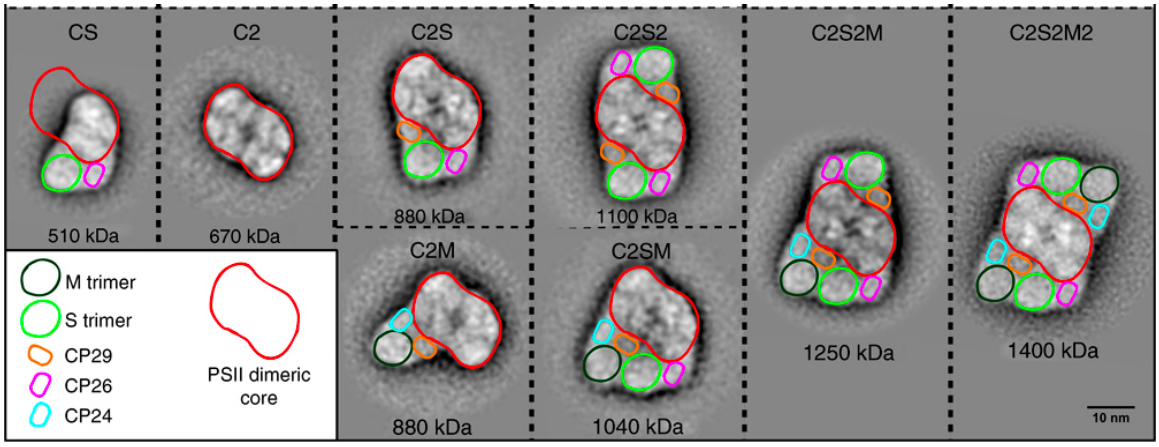


Figure 1.3: PSII supercomplexes in *Arabidopsis thaliana* characterized by single-particle EM. Adapted from [35].

PSII supercomplexes can further assemble into regular 2d patterns termed arrays. Since their early description in the 1960s [143], these striking motifs have frequently been observed in EM and AFM studies of stacked thylakoid membranes, and are composed of tens or hundreds of unit cells containing one PSII core dimer and variable quantities of LHCII (reviewed in [109, 48, 71]). Array extent and unit cell have been correlated with a number of experimental factors, including the use of strains that lack various light-harvesting complex [190, 150, 112, 46, 43], PsbS [96], or photosystem I [130] proteins; cold storage [157, 167]; and acclimation to low light [102, 110]. Yet, in part because observed arrays can vary widely between similar samples [25, 189], quantitatively predictive statements about the structure and function of PSII arrays have not emerged. Hypothesized functional rationales for PSII arrays primarily focus on predicted exciton transport effects, such as optimizing antenna size, excitation quenching site availability, or reaction center connectivity to match light conditions [41, 83], and on protein and electron carrier mobility arguments [103, 67], but no functional role has been proven. PSII arrays are the focus of Chapters 3 and 5.

Zooming out still further, many photosynthetic membrane proteins display dramatic lateral heterogeneity within the complex architecture of the thylakoid. Specifically, PSII and LHCII typically localize to grana membranes; photosystem I and ATP synthase are sterically excluded from appressed membranes by their large stromal protrusions and are thus predominantly found in unappressed membranes; and other proteins such as cytochrome b_6f are present in both regions [3, 165]. The packing fraction of proteins in grana can be

estimated by combining the number density of PSII (as counted in AFM or EM images) with the stoichiometries of other protein components to PSII (as measured by gel electrophoresis of proteins extracted from fractionated membranes), yielding typical packing fractions in the range $\approx$ 0.6–0.8 [133, 107, 70, 110]—much denser than a generic cell membrane. The protein packing fraction in stroma lamellae is harder to measure because the particle counting step is more difficult, but it is thought that stroma lamellae are less dense than grana. A functional rationale for high grana protein density is easy to guess—high chromophore density surrounding a high density of water-splitting catalysts—but the physical driving forces that maintain these density and composition gradients are not known. This question is the focus of Chapter 4.

Photosynthetic efficiency relies on precise spatial organization of sites for light harvesting, exciton transport, and charge separation [155]. Nanoscale maps of the relative arrangement of LHCII and PSII in native-like grana membrane environments will be necessary for building a complete picture of exciton flow during photosynthetic light harvesting and energy conversion. LHCII lacks a lumenal protrusion like that of the PSII OEC; this has prevented microscopists from simultaneously mapping the positions of LHCII and PSII in disordered membrane environments (i.e., outside of arrays). Super-resolution fluorescence imaging techniques are a promising alternative, but have not yet been successfully applied to grana membranes, which are already highly fluorescent. Computer simulation has the potential to complement experiment and address this important gap in structural and mechanistic understanding. Potential uses for simulated protein maps as platforms for other studies are discussed in Chapter 6.

1.1.5 The thylakoid as an active material

Thus far, I have framed the photosynthetic apparatus as a complicated but ultimately static system that accomplishes the basic functions of a solar panel and little more. Yet the natural photosynthetic apparatus is a much richer solar energy conversion device than its artificial counterparts—it is an active material that adapts to fluctuating photon inputs and downstream energetic demands on a variety of time scales, and is self-repairing when photodamage occurs. Because PSII is a key site of photodamage [179], many of these dynamic processes share the phenotype of a reduction in productive photochemistry at PSII and a concomitant increase in exciton relaxation via other pathways such as fluorescence. This has led to the use of the umbrella term nonphotochemical quenching (NPQ) to cover a family of processes with more or less archaic names, including energy-dependent quenching (qE), state transitions (qT), and photoinhibition (qI). While most of the research in this dissertation does not directly answer questions about NPQ, an awareness of NPQ is essential for understanding the motivations for, implications of, and potential extensions from this work, and could inspire the design of artificial biomimetic active materials.

The qE component of NPQ is a response to a rapid increase in light intensity, on the scale of seconds to minutes [192, 149]. The sequence of events in qE is as follows: (i) increased water-splitting activity at PSII decreases the pH in the lumen; (ii) the xanthophyll cycle

enzyme violaxanthin deepoxidase (VDE) becomes protonated and catalyzes the conversion of violaxanthin to zeaxanthin; (iii) PsbS [116, 117], a difficult-to-characterize member of the LHC family, also becomes protonated; (iv) molecular rearrangements occur; (v) excitations of antenna chlorophylls are quenched. Although a quantitatively predictive mathematical model of qE has been developed based on current knowledge [191], many details of this process remain unclear, and in particular, many proposed models of step (iv) are unsatisfying from a soft matter perspective. PsbS may undergo a dimer-to-monomer transition upon protonation [20], and PsbS expression levels have been correlated (via the *npq4* knockout and *L17* overexpression mutants) to PSII array formation [96], grana stacking [108], fluorescence lifetime [81], dissociation of a LHCII-CP29-CP24 complex from PSII [23], and changes in typical inter-PSII distances [23]. In addition, the NPQ phenotype of qE has been correlated to changes in typical inter-PSII and inter-LHCII distances [91], changes in bulk protein mobility [91], and various spectroscopic signatures of LHCII aggregation (e.g., [148, 177]). These findings are often interpreted within the context of xanthophyll-mediated LHCII aggregation [69, 84], although more work is needed to characterize these putative aggregates in vivo.

The other two well-characterized components of NPQ, state transitions and photoinhibition, are both dependent on phosphorylation of the stromal faces of grana-localized proteins, leading to PSII supercomplex disassembly, diffusive transport of the phosphorylated proteins from grana to stroma lamellae, and a change in the degree of grana stacking [170, 181]. Yet the two processes have little else in common. State transitions occur when the redox balance of the plastoquinone pool is upset in moderate or fluctuating light conditions [114], involve phosphorylation of LHCII by the kinase STN7 [18] and detachment of LHCII from PSII [87, 51, 88, 49], and result in an increase in LHCII mobility [52, 40, 86], the formation of physical and excitonic connections between LHCII and PSI [122, 63], and ultimately in the smoothing of structural and composition gradients between grana and stroma lamellae [39]. It is often stated that 80% of major LHCII complexes participate in state transitions in *Chlamydomonas* while only 20% do in plants; the classic references for these two claims ([49] and [6, 16], respectively) use somewhat outdated experimental methods. On the other hand, in the PSII photodamage-and-repair cycle, damaging high light conditions trigger phosphorylation of PSII core proteins by the kinase STN8, leading to complete disassembly of the supercomplex [171], increased bulk pigment-protein mobility [79], transport of the damaged D1 protein to the stroma lamellae for degradation and resynthesis [13], and reduction in the diameter of grana stacks [62, 79]. From a physical perspective, open problems in state transitions and PSII repair include constructing clear framings of their out-of-equilibrium character (i.e., is each better thought of as a driven process or as relaxation after a perturbation), determining the nature of the driving forces for protein and membrane rearrangement induced by phosphorylation, and dissecting the protein-specific mobility patterns that are convoluted within the current state-of-the-art fluorescence recovery after photobleaching (FRAP) experiments.

1.2 Coarse-grained simulation for plant biologists

1.2.1 Motivation for modeling

Photosynthesis spans disparate length and time scales, from femtosecond quantum events to seasonal canopy-scale processes. Many of the key phenomena that regulate the efficiency of light-harvesting and charge separation in the thylakoid membrane occur on intermediate scales—nanometers to microns in space, and milliseconds to minutes in time. In this regime, it has been experimentally challenging to simultaneously probe structure and function. On one hand, atomic force microscopy (AFM) and electron microscopy (EM) and tomography (ET) have proven to be powerful tools for elucidating the organization of the thylakoid membrane architecture and of its intrinsic pigment-proteins, but they typically lack sufficient temporal resolution to resolve dynamic structural fluctuations in detail. On the other hand, the popular biochemical, spectroscopic, and fluorescence microscopy methods are effective for exploring protein behavior within and trafficking between grana and stroma lamellae, but obscure functional distinctions between the many local environments within each membrane region. In addition, while isolated grana are useful model systems, the complexities of thylakoid morphology and pigment-protein synthesis have made it difficult to construct minimal reconstituted systems that would enhance experimental control over the membrane's structure and contents.

Coarse-grained computer simulation has the potential to bridge the gap between these structural and functional techniques. Computer simulation offers control over every input to the system (e.g., the number and identity of particles), and access to every output (e.g., particle locations over time), enabling the researcher to test hypotheses and make predictions about relationships between thermodynamic parameters and biologically-relevant phenotypes. Here, we specifically consider coarse-graining at the nanoscale, where soft matter physics is often more illuminating than quantum-mechanical or even atomistic chemical perspectives. Coarse-grained modeling is powerful and widely applicable, yet, like any scientific technique, requires attention to a host of subtleties if it is to yield meaningful insights [59].

1.2.2 Energetics

Coarse-grained models derive their strength from the clarity and broad applicability of their assumptions, rather than from the precision of their details. Therefore, the potential energy function governing a coarse-grained model should include the simplest set of particles, interactions, and material properties that captures the phenomenology of interest and is motivated by the underlying physics.

For instance, consider the nanoscale properties of a generic protein: it occupies space; it may have a compact shape; it may form specific contacts with other proteins; it may aggregate or crystallize. In many cases, these properties can be represented mathematically by pairwise energetic potentials between protein particles.

The basic features of appropriate potentials between particles representing proteins (or

strongly bound complexes) are often clearly dictated by their scale and molecular nature. For example, forces between proteins with well-defined internal structure should include a steric contribution, establishing the volumes they occupy. This repulsive interaction is harsh, acting over short range, and can be reasonably caricatured as a singular "hard core." Continuous potentials that achieve the same effect, such as the Weeks-Chandler-Andersen (WCA) potential [182], are sometimes preferred for practical reasons. Purely repulsive hard spheres were among the earliest model systems studied with molecular simulation techniques and, despite their simplicity, display a rich phase behavior [4]. Phase diagrams have also been explored for nonspherical hard shapes, including rods [28], ellipsoids [140], cubes [162], and exotic polyhedra [42, 78].

Cohesion between proteins in solution can emerge from many sources, e.g., hydrogen bonding, screened Coulomb interactions, hydrophobic effects, and other solvent-mediated effects. Despite this variety, however, the attractions effected by these forces are similar in nature at the coarse-grained level. In particular, they typically attenuate over distances that are short compared to the scale of proteins themselves. They thus act primarily as contact potentials, which reward close approach of coarse-grained particles. A minimal model of this behavior is a "square-well" potential that has a constant strength within a cutoff radius and is noninteracting beyond the cutoff radius. Other models use the Lennard-Jones (LJ) potential to accurately capture the $r^{1/6}$ decay of van der Waals interactions [115], the Yukawa potential to capture screened electrostatics [55], or custom functional forms to capture other phenomenology (e.g., [154, 144]). Square-well and Yukawa potentials are typically used in concert with a hard-core repulsion; the LJ potential includes its own volume exclusion term, from which the WCA potential is derived. Any attractive interaction can be made to act isotropically, favoring nonspecific aggregation, or between interaction sites on so-called "patchy" particles, often favoring self-assembly of specific structures [75].

Lipid bilayer membranes are modeled in different ways depending on the desired range of bending fluctuations. If the membrane is essentially flat and bending fluctuations are not expected to affect the phenomenon of interest, it can be modeled simply as a static planar surface through which particles representing intrinsic membrane proteins can travel. Because grana lamellae appear flat in most electron micrograms and tomograms (reviewed in [48, 44, 137]), this approach has been used for coarse-grained simulations of grana proteins [154, 107, 174, 173, 52].

In more generality, a membrane can be well described by the Helfrich Hamiltonian for an elastic sheet [151]. When only small-amplitude fluctuations about a planar equilibrium state are considered, it is convenient to represent the membrane in the Monge gauge (i.e., each point (x, y) in the plane is associated with a height $h(x, y)$ above the plane), and to linearize the Hamiltonian to decouple the Fourier modes [151]. If significant changes in membrane curvature or topology are essential to the research question, then the membrane can be represented as a triangulated set of tethers and nodes [65], or as a collection of "membrane patch" particles governed by a many-body potential that mimics the hydrophobic forces that drive lipid bilayer cohesion [144]. To represent the internal structure of lipids at higher resolution, various models have been developed (e.g., [47, 183, 89]).

1.2.3 Dynamics

Coarse-grained biophysical models forego an explicit solvent for computational and conceptual efficiency. They instead include the random buffeting of an implicit solvent via stochasticity and friction in the algorithm that generates new configurations in a trajectory. Two common classes of algorithms for this purpose are Langevin dynamics and Metropolis Monte Carlo, which we sketch here; see the excellent textbooks [60, 8] for thorough explanations and implementation guidelines.

Overdamped Langevin dynamics, also known as Brownian dynamics, propagate a system by integrating equations of motion that include deterministic forces derived from the potential energy function, as well as random forces parameterized by diffusion coefficients. For systems of particles, the integration is typically performed in real space; for a membrane in the Monge gauge, it is often easier to perform the integration in Fourier space [32]. Because the equations of motion involve gradients of the potential, standard algorithms for Brownian dynamics require the interaction energies to be differentiable.

Metropolis Monte Carlo takes a conceptually different approach: each new configuration is generated from the previous one by proposing a random perturbation or "move" of the system, then accepting or rejecting the move according to the Metropolis-Hastings acceptance criterion. The researcher is permitted considerable leeway in selecting the types of moves that are proposed. These moves can be physically motivated, such as small displacements or rotations of single particles, or they can be starkly nonphysical, such as the insertion or deletion of entire particles; in either case, moves should be designed to efficiently sample the most important regions of state space. The main requirement is that the proposal–acceptance scheme must obey detailed balance so that the process creates a Markov chain with a well-defined stationary distribution. Advanced Monte Carlo methods include algorithms for free energy calculations (e.g., umbrella sampling with MBAR [160]), equilibration on rough landscapes (e.g., replica exchange [53]), and rare event sampling (e.g., transition path sampling [27]).

The choice of simulation dynamics algorithm often hinges on whether it is more important to fully characterize the system's equilibrium properties, or to most realistically capture its dynamics. Brownian dynamics and Monte Carlo dynamics can both accurately simulate time correlations in the low-Reynolds-number systems of biology in some limits, although hydrodynamic effects can be difficult to compute correctly [56, 32, 21]; both can require large computational resources to explore the system's equilibrium fluctuations. In Monte Carlo, cleverly-chosen move sets can dramatically reduce the computational time necessary to sample the equilibrium distribution—yet these moves are often ones that create highly unphysical dynamics. However, compromises exist; for instance, virtual-move Monte Carlo is a computationally-efficient Monte Carlo method that yields the correct dynamic behavior both for single particles undergoing free diffusion, and for large clusters undergoing collective motion [185]. We note that when investigating an inherently time-dependent question such as the kinetics of a system's relaxation after a perturbation, it is often illustrative to characterize the equilibrium ensembles before and after the perturbation as a prelude to studying the

dynamics in between.

1.2.4 Emergent behavior

In equilibrium simulations with Monte Carlo or Brownian dynamics, as in reality, it is not the minimum energy classes of configurations that are the most probable ones—it is the minimum *free* energy classes, which account for the effects of entropy as well as energy. The complex interplay between energetic and entropic forces can give rise to striking self-assembled structures and counterintuitive collective phenomena, even in seemingly simple systems. Discovery, characterization, and prediction of such emergent properties are frequent goals of coarse-grained modeling.

Several classes of entropic forces are likely to be important in photosynthetic protein–membrane systems. One is depletion-attraction, an effective attraction that brings some components of a system closer together in order to maximize the entropy of other (often smaller) components [11]. Depletion-attraction is ubiquitous in biological systems [123], and it is the driving force behind the crystallization of hard particles (described above). Another entropic force acts between layers of a stack of membranes, whose out-of-plane fluctuations are suppressed by steric constraints imposed by neighboring layers [77]. An accurate accounting of this force is necessary to understand the adhesion of membrane stacks [118].

Other emergent behaviors can arise from coupling between membranes and membrane proteins. For instance, membrane-mediated forces between intrinsic membrane proteins are caused by hydrophobic mismatch [73, 152], membrane curvature [169], and lipid composition [47]. Conversely, membrane-associated proteins can sculpt the membrane's morphology and composition fluctuations [163, 124].

Phase transitions are the paragon of collective behavior in statistical mechanics. Characterized by nonanalytic change (e.g., a discontinuity) in an observable quantity, they arise not only in simple molecular substances (e.g., water freezing or boiling) but also in a variety of biophysical contexts, including membrane binding-unbinding transitions [118] and protein crystallization [154]. At phase coexistence, a system can stably exist in each of two very different states, but is unstable as a mixture of the two (unless coexistence is allowed by the thermodynamic ensemble) [37]. Even if only one phase is stable at a time, the kinetics of phase transitions can nevertheless be complex. In classical nucleation theory, it is first slow to overcome the free energy barrier to nucleation of the new phase inside the old, then fast for nuclei to begin to grow, and finally (for high-symmetry phases) slow for defects to anneal; even more sluggish and exotic dynamics can be observed in practice [184, 76].

1.3 Previous applications of coarse-grained modeling to photosynthesis

Membrane morphology The mechanism of membrane curvature generation in the photosynthetic purple bacterium *Rhodobacter sphaeroides* has been investigated by Monte Carlo

simulation [61]. In that work, four physical features of the biological system were included in the coarse-grained model: the flexibility and fluidity of the membrane, the high packing fraction of reaction center-light harvesting 1 (RC-LH1) complexes and light harvesting 2 (LH2) complexes in the membrane, the size disparity between RC-LH1 and LH2 complexes, and the wedge shape of the complexes. Thus, the coarse-grained model consisted of a fluctuating triangulated network representing the membrane, with hard spheres at the vertices representing protein complexes. By varying the spontaneous curvatures and diameters of the hard sphere vertices in equilibrium simulations, the authors found support for the hypotheses that protein shape (via the Helfrich energy) and crowding (via depletion-attraction) both influence the organization of the bacterial photosynthetic membrane.

Protein self-assembly Thylakoid membrane proteins, particularly photosystem II (PSII) and light-harvesting complex II (LHCII), display a variety of self-assembled structural motifs (reviewed in [48, 137, 109]). Kirchhoff, Tremmel, and coworkers investigated crowding effects on PSII and LHCII organization by conducting Monte Carlo simulations of a coarse-grained model in which the protein complexes were represented by hard particles with structurally-detailed shapes, and the membrane was represented by a discretized static 2d sheet [174, 107]. This approach was computationally challenging because the discretization of space introduced artifacts when a short-range attraction was added to the model, and may have exacerbated the inherent difficulty of equilibrating a dense system using single-particle Monte Carlo moves [174]. These studies concluded that a richer model would be necessary to capture key structural motifs observed in vivo [107].

Mobility Particle mobility due to Brownian motion is characterized by a diffusion coefficient and can be studied experimentally by techniques like fluorescence recovery after photobleaching (FRAP) [103, 66, 91, 79, 106], single-particle tracking [40], or fluorescence correlation spectroscopy (FCS) [86]. After a perturbation that brings a system out of equilibrium, diffusion leads to fluxes of particle density that act to smooth out gradients in chemical potential, whether or not the perturbation changed the intrinsic particle diffusion coefficients. In contrast, in an equilibrium system, Brownian motion does not lead to net transport of particles, even if diffusion coefficients are large.

Plastoquinone mobility was considered using simulations of the Kirchhoff-Tremmel model, yielding a percolation threshold for plastoquinone at 60–70% protein packing fraction [173, 174]. LHCII diffusive transport from grana to stroma lamellae during state transitions was studied via Monte Carlo simulation of a mixture of hard discs; by fitting to time series of membrane fractionation data, the authors concluded that phosphorylated LHCII has a higher diffusion coefficient than unphosphorylated LHCII [52].

Chapter 2

A coarse-grained model of the photosynthetic apparatus

This Chapter lays out the model and methods that are used in Chapters 3 and 4. Our model was developed to recapitulate key features of thylakoid membrane protein organization described in Chapter 1, particularly in stacked grana membranes, at a spatial resolution comparable to that achieved by AFM and EM (i.e., at the nanoscale). We extend the previous computational approaches described in Chapter 1 in two key ways: we simulate multiple, coupled membrane layers in a grana stack; and we employ computational methods capable of exploring highly cooperative transitions. This Chapter contains published work from [154].

2.1 The model

2.1.1 Geometry

Our model consists of two coupled 2d layers, representing stroma-paired grana membranes, which comprise the minimal system for PSII array formation [45]. These two layers can be thought of as an isolated membrane pair, or as two stroma-paired membranes in a larger grana stack where lumen-side interactions are assumed to be negligible [48].

We represent protein complexes within membrane layers of a grana stack as greatly simplified particles that can move in only two directions (x and y) and rotate only about the axis z perpendicular to these layers. Our simulations sample configurations of a pair of these two-dimensional systems, periodically replicated in x and y, coupled by stacking interactions between particles in different layers. The model includes two particle species: isotropic disc-shaped particles represent trimeric LHCII complexes, and discorectangle-shaped particles represent PSII supercomplexes (specifically the so-called C_2S_2 supercomplex in *Chlamydomonas* [172], spinach [139], and *Arabidopsis* [35]) (Fig. 2.1).

Particle shapes are simple approximations of the solved protein complex structures [35, 138], and particle sizes are assigned to be consistent with the protein structures and with previous coarse-grained models [173]. Specifically, each LHCII particle has a diameter $\sigma_{\mathrm{L}} =$

6.5 nm, and each PSII particle has a rectangle width and cap diameter $D_P = 12.0$ nm and rectangle length $L_P = 14.5$ nm (Fig. 2.2). PSII particles, like PSII protein supercomplexes, have two constituent embedded LHCIIs that move with the PSII particle as a rigid body. The embedded LHCIIs are constrained to their physiological locations on either side of the PSII axis [138], making the PSII particles chiral.

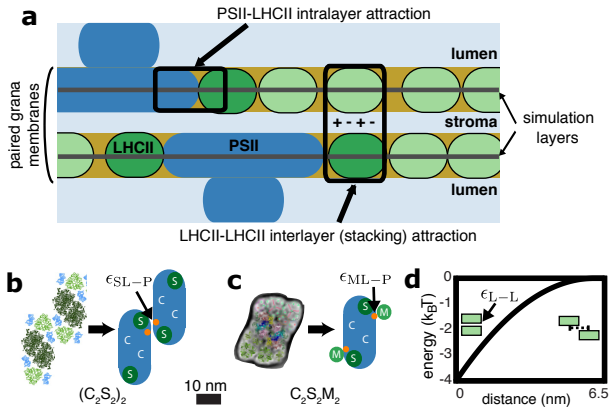


Figure 2.1: Graphical summary of the model. **a**: Side view cartoon of protein geometry and interactions. PSII (blue) and LHCII (green) particles reside in two coupled layers, each of which represents a lipid bilayer grana membrane (tan) surrounded by stromal and lumenal aqueous regions (pale blue). **b** and **c**: Top view from the lumenal side of the coarse-graining procedure. Solved protein structures of LHCII trimers are roughly circular; model LHCII particles are hard discs with diameter 6.5 nm (dark green: S-LHCII; medium green: M-LHCII). PSII C_2S_2 supercomplexes are roughly rod-shaped; model PSII particles are hard discorectangles with diameter 12 nm and tip-to-tip length 26.5 nm (blue). The orange dots show LHCII binding sites that allow formation of supramolecular complexes, including $(C_2S_2)_2$ (**b**, governed by ϵ_{SL-P}) and $C_2S_2M_2$ (**c**, governed by ϵ_{ML-P}). Protein structures adapted from [35, 48]. **d**: Potential energy of the LHCII interlayer stacking interaction. Insets: side view cartoons of unbound LHCII (light green) in stacked (left) and unstacked (right) configurations.

2.1.2 Energetics

In addition to excluding area within the layer that they occupy, these particles attract one another in three distinct ways.

The first two attractions act between an interaction site on PSII and an LHCII particle residing in the same layer. These interactions are motivated by experiments where, when protein complexes are isolated from thylakoid membranes of *Arabidopsis*, spinach, or *Chlamydomonas*, LHCII complexes are found to be distributed among at least three different local environments: they can be strongly bound within a supercomplex (embedded

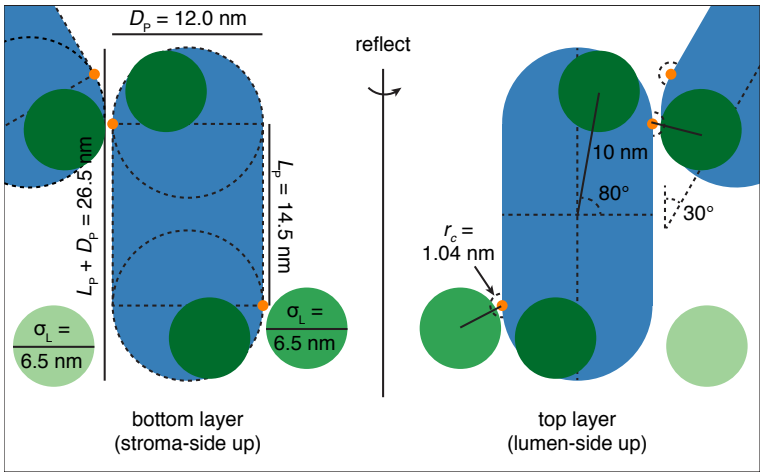


Figure 2.2: Detailed geometry of the particle sizes, shapes, and intralayer interactions. (left) Non-embedded LHCII particles (light or medium green) are discs with diameter $\sigma_\mathrm{L} = 6.5$ nm, excluding an area $a_\mathrm{L} = \frac{\pi}{4}\sigma_\mathrm{L}^2 \approx 33$ nm^2. PSII particles (blue) are discorectangles with width $D_\mathrm{P} = 12.0$ nm and length $L_\mathrm{P} = 14.5$ nm, yielding a total tip-to-tip length of 26.5 nm and excluding a total area $a_\mathrm{P} = \frac{\pi}{4}D_\mathrm{P}^2 + L_\mathrm{P}D_\mathrm{P} \approx 287$ nm^2. Within PSII particles, embedded LHCII particles (dark green) and in-plane PSII–LHCII binding sites (orange) participate in attractive interactions but do not exclude area. (right) LHCII particles are considered bound to an in-plane binding site if the LHCII particle penetrates a disc of radius $r_c = 1.04$ nm around the in-plane binding site. If the LHCII particle is embedded, then, in addition to satisfying the distance cutoff, the angle difference between the long axes of the two PSII particles hosting the interaction sites must be $\leq 30°$. PSII particles in the bottom (left) and top (right) layers of the two-layer simulation geometry are mirror images because PSII protein supercomplexes are chiral and have distinct stromal and lumenal faces (see Fig. 2.1a).

or S-LHCII), moderately bound to the edge of a supercomplex (bound or M-LHCII), or structurally detached from PSII (free LHCII) [35, 48, 189, 172]. Motivated by the apparent equilibrium between moderately bound and free LHCIIs, we introduce two interaction sites on the periphery of each PSII particle, as illustrated by the orange dots in Fig. 2.1b and c and Fig. 2.2. These interaction sites attract LHCIIs in the same layer over a short range ($\approx$1 nm), such that modeled M-LHCIIs in the "bound" state are constrained to locations relative to PSII that are consistent with published $C_2S_2M_x$ complexes [35, 172].

Experiments suggest that the intralayer interactions between PSII supercomplexes and free LHCIIs are similar to yet distinct from those between a PSII supercomplex and the S-LHCII of another supercomplex; for instance, binding of M-LHCII (Fig. 2.1c) appears to require PSII subunit CP24 in *Arabidopsis*, while association of S-LHCII with the same site (Fig. 2.1b) may not [112]. We therefore define separate energy scales for these interactions ($\epsilon_{\mathrm{ML-P}}$ and $\epsilon_{\mathrm{SL-P}}$, respectively). Based on measurements of intramembrane associations between other protein complexes in photosynthetic membranes [119] and on other coarse-grained membrane protein models [126], these energies are expected to be modest, on the order of one or a few

$k_{\mathrm{B}}T$; we focus on $\epsilon_{\mathrm{SL-P}} = \epsilon_{\mathrm{ML-P}} = 2k_{\mathrm{B}}T$.

We model both PSII–LHCII interactions as square-well potentials, with a distance cutoff of $0.66\sigma_{\mathrm{L}}$ between the LHCII center and the binding site, or equivalently a distance cutoff between the LHCII edge and the binding site of $0.16\sigma_{\mathrm{L}} = 1.04$ nm; this short-range distance cutoff ensures that each binding site can bind at most one LHCII at a time and is comparable to other coarse-grained membrane protein models [126]. This square well completely describes the $\epsilon_{\mathrm{ML\text{-}P}}$ interaction between PSIIs and nonembedded LHCII particles. The attraction between PSIIs and embedded LHCIIs, $\epsilon_{\mathrm{SL\text{-}P}}$, has an additional constraint: the angle difference between the long axes of the two PSII particles hosting the interaction sites must be $\leq 30°$, ensuring that only approximately-parallel PSIIs may participate [48]. Temperature T is fixed at 1 throughout, such that $k_{\mathrm{B}}T$ is our reduced unit of energy.

The other attraction involves two LHCII particles (free or embedded) in different layers. Electrostatic interactions between LHCII stromal faces contribute to grana stack integrity [14] and are implicated in PSII array formation in experiments [45, 25]. Standfuss and coworkers have proposed a "velcro-like" qualitative model for the LHCII stacking interaction [166], but a quantitative understanding is lacking. We have chosen a simple phenomenological form for the LHCII–LHCII interlayer attraction that favors LHCII face-to-face contact (Fig. 2.1a,d). This interaction has an energetic minimum, with a value of $\epsilon_{\mathrm{L\text{-}L}}$, when one disc completely eclipses another as viewed from above the layers. It vanishes continuously at $r = \sigma_{\mathrm{L}}$, where r is the lateral distance between the centers of two LHCII discs. Specifically, let r be the lateral distance between LHCII particle centers, i.e., the magnitude of the disc-disc separation vector once projected into the plane of a membrane sheet. Then the form of the model LHCII–LHCII interlayer interaction, u_{stacking}, is

$$u_{\mathrm{stacking}}(r) = \begin{cases} -\epsilon_{\mathrm{L-L}} \left(\frac{r-\sigma_{\mathrm{L}}}{\sigma_{\mathrm{L}}}\right)^2 & \text{if } r < \sigma_{\mathrm{L}} \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

and is plotted in Fig. 2.1d. All LHCII particles can participate in this interaction.

We set the strength of the attraction to $\epsilon_{\mathrm{L-L}} = 4\ k_{\mathrm{B}}T$, except where noted, which creates strong but reversible binding. This is consistent with the electrostatic attraction expected between two correspondingly charged parallel plates in an aqueous solution of counterions. According to the approximate approach of Ref. [129], this energy can be estimated as $E_{\mathrm{elec}} = -2\pi a\sigma^2 d/(4\pi\epsilon_0\epsilon)$, where a is the plate area, σ is the surface charge density on each plate, d is the distance between the plates, ϵ_0 is the permittivity of free space, and ϵ is the dielectric constant of water. Here, the plates represent the negatively-charged stromal surfaces of LHCII, and the counterions represent positively-charged LHCII N-terminal domains (which are flexible and extend into the aqueous gap [166]) as well as dissolved cations. Therefore, we assign $a = a_{\mathrm{L}}$, the area of an LHCII particle (which has been matched to the protein structure), and $\sigma = -0.02\ C/\mathrm{m}^2$, within the range of measured surface charge densities on the stromal side of grana membranes [14]. When two such plates are separated by a typical stromal gap distance $d = 1.8$ nm [48], we obtain $E_{\mathrm{elec}} = -3.94\ k_{\mathrm{B}}T$.

We emphasize that this model includes only the short-ranged intramembrane protein-protein attractions that are most conclusively established by single-particle structural studies.

Longer-range patterns in our simulations can only arise from emergent correlations involving many particles. We also emphasize that the values of the interaction strengths that we have selected are intended to be representative of the physiology of a generic grana membrane, and not to represent any specific species or growth condition. Experimental measurements of the interaction strengths, which would be difficult but not inconceivable (e.g., along the lines of Ref. [119]), could support future efforts to tailor the model to precisely match specific experimental conditions.

2.2 Monte Carlo simulations

Using standard Metropolis Monte Carlo methods, we sampled equilibrium configurational distributions of systems comprising N_{P} PSII particles and a variable number N_{L} of free LHCII particles. Trial moves included translational displacements of a randomly chosen LHCII or PSII particle, and rotational displacements of a randomly chosen PSII particle. Translational moves were restricted to periodically replicated 2D surfaces, each with area A per layer. x- and y-components of translational moves were chosen from Gaussian distributions with zero mean and variances of 0.6 nm^2 for free LHCII displacements or 0.2 nm^2 for PSII displacements. PSII rotations were chosen uniformly on the interval $(-10°, 10°)$. An "MC sweep" consisted on average of N_{L} LHCII displacements, N_{P} PSII displacements, and N_{P} PSII rotations.

To scrutinize phase behavior, we performed simulations in which A and N_{L} were allowed to fluctuate (i.e., sampling an osmotic ensemble [125]). These fluctuations were regulated by specified values of a two-dimensional pressure p and a chemical potential μ_{L}, respectively. In addition to particle displacement trial moves, osmotic ensemble simulations included trial moves that change A (box area moves) and trial moves that change N_{L} (LHCII insertion and deletion moves). Acceptance probabilities for the latter moves were constructed to satisfy detailed balance, with the area per layer A scaled by the number of layers n_l where necessary.

2.3 Analysis of simulated data

2.3.1 Clustering algorithm

PSII arrays were identified using connected component analysis [82] with three neighbor criteria, all of which must be satisfied for two PSII particles to be "connected": (i) the angle difference between particle axes must be $< 15°$; (ii) the component of the center-to-center separation vector parallel to a particle axis must be < 26.5 nm; (iii) the component of the center-to-center separation vector perpendicular to a particle axis must be < 14 nm.

2.3.2 Free energy calculations

We determined free energy profiles $F(\rho)$ as a function of total packing fraction ρ by exploiting the fundamental relationship $F(\rho) = -k_{\mathrm{B}}T \ln P(\rho)$ to the probability distribution

$P(\rho)$ for spontaneous packing fraction fluctuations. Statistics of ρ were in turn determined by maximum likelihood estimation from osmotic ensemble simulations with imposed harmonic bias potentials of the form $u_{\text{bias}}(\rho) = k(\rho - \rho_0)^2$. The MBAR method as implemented in the PyMBAR package [160] was used to pool and reweight these umbrella sampling results as follows.

2.3.3 Thermodynamic reweighting

In the osmotic ensemble, a microscopic state (i.e., a particular realization of all fluctuating variables) is specified by $\Gamma = \left(A, N_{\text{L}}, \mathbf{r}^N\right)$, where $\mathbf{r}^N$ is a vector detailing the coordinates of all $N = N_{\text{L}} + N_{\text{P}}$ particles. Defining n_l as the number of membrane layers ($n_l = 2$ for all calculations described in this work), the equilibrium probability distribution $P(\Gamma)$ of such microstates is given by

$$P(\Gamma) \propto \frac{1}{N_{\text{L}}!\ell^{2N_{\text{L}}}} e^{-\beta\left(U(\Gamma) + n_l A p - (\mu_{\text{L}} - \mu_{\text{L}}^{(0)})N_{\text{L}}\right)} \tag{2.2}$$

where $\beta = (k_{\text{B}}T)^{-1}$, $\ell = 1$ nm is a reference length scale defining the standard state, and $\mu_{\text{L}}^{(0)}$ is the chemical potential of noninteracting LHCII particles at the standard state density ℓ^{-2}. We report all chemical potential values relative to $\mu_{\text{L}}^{(0)}$, defining $\bar{\mu}_{\text{L}} = \mu_{\text{L}} - \mu_{\text{L}}^{(0)}$. The proportionality coefficient in Eq. 2.2 is determined by normalization, $\sum_\Gamma P(\Gamma) = 1$, where $\sum_\Gamma(\cdots) = \int dA \sum_{N_{\text{L}}} \int_A d\mathbf{r}^N (\cdots)$ implies summation/integration over all fluctuating variables.

The packing fraction $\rho = (n_l A)^{-1} \sum_i a_i N_i$ is proportional to a linear combination of the excluded areas a_i for each particle type i. Its probability density $P(\rho)$ sums the statistical weights of all microstates sharing a given value of ρ, say $\rho = \varrho$:

$$\begin{aligned} P(\varrho) &\propto \sum_\Gamma P(\Gamma)\delta\left(\rho - \varrho\right) \\ &\propto \exp\left[-\beta F(\varrho)\right], \end{aligned}$$

Now consider probability distributions $\tilde{P}(\Gamma)$ and $\tilde{P}(\rho)$ at a different set of thermodynamic parameters, $\tilde{\mu}_{\text{L}} = \mu_{\text{L}} + \Delta\mu_{\text{L}}$ and $\tilde{p} = p + \Delta p$. The simple connection between weights of a particular microstate at different thermodynamic parameters,

$$\tilde{P}(\Gamma) \propto P(\Gamma) \exp\left[-\beta\left(n_l A \Delta p - N_{\text{L}}\Delta\mu_{\text{L}}\right)\right],$$

allows us to determine $\tilde{P}(\rho)$ at $\rho = \varrho$ by "reweighting" density statistics at the original parameters μ_{L} and p:

$$\begin{aligned} \tilde{P}(\varrho) &\propto \sum_\Gamma P(\Gamma)\delta\left(\rho - \varrho\right) \exp\left[-\beta\left(n_l A \Delta p - N_{\text{L}}\Delta\mu_{\text{L}}\right)\right] \\ &\propto \left\langle \exp\left[-\beta\left(n_l A \Delta p - N_{\text{L}}\Delta\mu_{\text{L}}\right)\right]\right\rangle_{\varrho,\{p,\mu_{\text{L}}\}} \\ &\propto \exp\left[-\beta\tilde{F}(\varrho)\right]. \end{aligned}$$

The subscript $\varrho, \{p, \mu_{\mathrm{L}}\}$ on angled brackets indicates a conditional average over configurations with $\rho = \varrho$, at chemical potential μ_{L} and pressure p. We exploited these relationships to compute free energy profiles for a range of values of μ_{L} and p, using data from simulations at a single thermodynamic state.

2.3.4 Drawing the phase diagram

Constructing coexistence curves in the ϕ,ρ-plane requires computing the average composition $\langle\phi\rangle_\alpha$ and packing fraction $\langle\rho\rangle_\alpha$ for each phase α at each state $\{p^*, \mu_{\mathrm{L}}^*\}$ of phase equilibrium. To do so, we associate with phase α a continuous interval of ρ that: (a) straddles the corresponding minimum in free energy at ρ_α, (b) does not cross the local free energy maximum at intermediate packing fraction $\rho^\ddagger$, and (c) is bounded on either side by values ρ_{bound} satisfying $F(\rho_{\mathrm{bound}}) = F(\rho_\alpha) + \Delta F$. The coexistence curves we present were obtained with $\Delta F = 10\ k_{\mathrm{B}}T$; results are quite insensitive to this choice provided that ΔF is significantly larger than $k_{\mathrm{B}}T$ but smaller than the barrier height $F(\rho^\ddagger) - F(\rho_\alpha)$.

Integrating over these intervals, we computed average order parameters at $p^* = p + \Delta p$, $\mu_{\mathrm{L}}^* = \mu_{\mathrm{L}} + \Delta\mu_{\mathrm{L}}$ according to

$$\begin{aligned}\langle\rho\rangle_\alpha &= \int_\alpha d\rho \tilde{P}(\rho)\rho, \qquad \text{and} \\ \langle\phi\rangle_\alpha &= \frac{1}{N_{\mathrm{P}}}\int_\alpha d\rho \tilde{P}(\rho)\langle N_{\mathrm{L}}\rangle_{\rho,\{p^*,\mu_{\mathrm{L}}^*\}}.\end{aligned}$$

Using the same reweighting scheme described above, $\langle N_{\mathrm{L}}\rangle_{\rho,\{p^*,\mu_{\mathrm{L}}^*\}}$ was calculated from statistics at the reference state $\{p, \mu_{\mathrm{L}}\}$:

$$\langle N_{\mathrm{L}}\rangle_{\rho,\{p^*,\mu_{\mathrm{L}}^*\}} = \frac{\langle N_{\mathrm{L}} \exp\left[-\beta\left(n_l A\Delta p - N_{\mathrm{L}}\Delta\mu_{\mathrm{L}}\right)\right]\rangle_{\rho,\{p,\mu_{\mathrm{L}}\}}}{\langle \exp\left[-\beta\left(n_l A\Delta p - N_{\mathrm{L}}\Delta\mu_{\mathrm{L}}\right)\right]\rangle_{\rho,\{p,\mu_{\mathrm{L}}\}}}$$

Systems with fixed density and composition that lie inside the coexistence region will of course phase separate, yielding at equilibrium a fluid phase with parameters (ϕ_{f},ρ_{f}) and a crystalline phase with parameters (ϕ_{c},ρ_{c}), separated by an interface. For a given net packing fraction ρ and composition ϕ, the end points of its tie line in Fig. 3.10 indicate the properties ϕ_{f}, ρ_{f}, ϕ_{c}, and ρ_{c} of these consitituent phases, which lie along the corresponding coexistence curves. The tie lines were calculated parametrically, by varying the possible proportions of coexisting states. Let $f = A_{\mathrm{f}}/A$ be the fraction of total area partitioned to the fluid phase, so that $1 - f = A_{\mathrm{c}}/A$ is the fraction partitioned to the crystal phase. The net packing fraction is a simple linear combination $\rho(f) = \rho_{\mathrm{f}} f + \rho_{\mathrm{c}}(1-f)$ of the coexisting values. Net composition is more conveniently written in terms of $z = f/(1-f)$ and the ratio $r = (a_{\mathrm{L}}\phi_{\mathrm{c}} + a_{\mathrm{P}})/(a_{\mathrm{L}}\phi_{\mathrm{f}} + a_{\mathrm{P}})$ of packed area per PSII in the two phases,

$$\phi(z) = \frac{\phi_{\mathrm{f}}\rho_{\mathrm{f}} z + r\phi_{\mathrm{c}}\rho_{\mathrm{c}}}{\rho_{\mathrm{f}} z + r\rho_{\mathrm{c}}}.$$

Given coexisting values of ϕ_f, ρ_f, ϕ_c, ρ_c (from which r can be easily determined), we vary f from 0 to 1 (thus varying z from 0 to ∞). The set of paired values (ϕ, ρ) generated in this way comprise a single tie line.

2.4 Comparison to experimental data

Let ϕ' be an experimentally measured mole ratio of LHCII trimers (including free, S-LHCII, etc.) to PSII core monomers, e.g., as measured by Western blots or spectroscopy. Because there are two PSII core monomers and two LHCII trimers per C_2S_2 supercomplex, $\phi = 2\phi' - 2$.

Let η_P be a number density of PSII, typically reported in units of PSII per square micron. Then $\rho = \eta_P (a_P + \phi a_L)$.

To calculate the experimental data points in Fig. 3.10, we took values of η_P and ϕ' from Tables 2 and 4, respectively, in Ref. [102], and used $a_L = 33.18$ nm^2 and $a_P = 287.1$ nm^2 (Fig. 2.2). For the low-light growth condition, $\phi' = 5.42, \eta_P = 1436 \pm 57/\mu\text{m}^2 \implies \phi = 8.84, \rho = 0.833 \pm 0.013$. For the ordinary-light growth condition, $\phi' = 3.28, \eta_P = 1720 \pm 63/\mu\text{m}^2 \implies \phi = 4.56, \rho = 0.754 \pm 0.028$. If a standard deviation or standard error were available for ϕ, it could be included when calculating the error in ρ.

Chapter 3

Coexistence of fluid and crystalline phases of proteins in photosynthetic membranes

In this Chapter, we use the model and simulations described in Chapter 2 to examine PSII array formation in system sizes comparable to full grana discs (hundreds of nanometers in diameter). By thoroughly examining thermal fluctuations in our model of a PSII–LHCII binary mixture, we reveal a phase transition between a relatively dilute, disordered PSII–LHCII fluid and a dense, ordered PSII–LHCII crystal. Physiological protein concentrations are at fluid-crystal phase coexistence or near the boundary between the fluid and coexistence regions, where small changes in protein density or interactions can lead to dramatic shifts in the observed degree of array formation and in any array-dependent functionality. This Chapter contains published work from [154].

3.1 Simulations capture experimentally-determined intramembrane protein organization

We performed Monte Carlo simulations to sample equilibrium particle configurations at a wide range of particle concentrations within experimentally relevant parameter regimes. Two standard experimental measures of grana protein content are the LHCII:PSII ratio (specifically, the mole ratio ϕ of free LHCII trimers to PSII C_2S_2 supercomplexes) and the number density of PSII (typically PSII per square micron of membrane). These metrics can be combined to estimate the total protein packing fraction ρ (i.e., the fraction of grana membrane area occupied by proteins). Published values of these quantities vary significantly based on plant growth conditions and membrane preparation protocols: ϕ is typically in the range $\approx$ 2–6 [90, 102, 180], and ρ is typically in the range $\approx$ 0.6–0.8 [133, 107, 70, 110]. We performed simulations with the parameters given in Table 3.1. For each set of structural parameters; for each set of parameters, at least 7 simulations were initialized from independent

fluid or partially crystalline configurations.

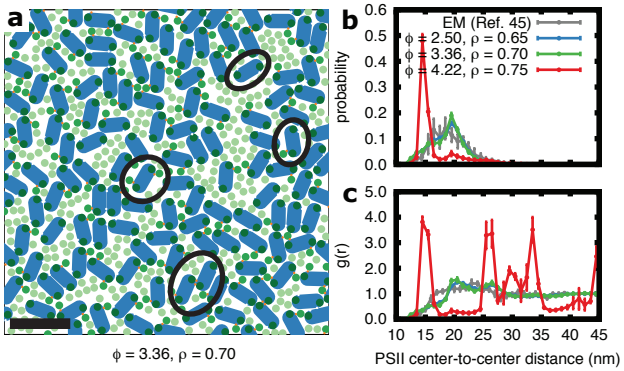


Figure 3.1: Intramembrane structure of a coupled pair of model grana membranes at 1750 PSII/μm^2. **a**: Snapshot of the lumen-side-up membrane layer from a simulation at $\rho = 0.70$, $\phi = 3.36$ showing a representative disordered configuration. Example supramolecular complexes are circled, clockwise from top: C_2S_2, C_2S_2M, $(C_2S_2)_2$, $C_2S_2M_2$. Color scheme as in Fig. 2.1; scale bar = 50 nm. **b** and **c**: Comparison between experimental and simulated data for the statistics of intramembrane PSII center-to-center separation distances. **b** shows the distribution of nearest-neighbor distances, and **c** shows the radial distribution function $g(r)$. Freeze-fracture EM data on isolated spinach grana membranes [107], with 1700 PSII/μm^2 on average, are consistent with these simulations of the disordered state. In contrast, simulations at $\rho = 0.75$ show PSII aggregation and ordering.

Simulations of moderately dense systems ($\rho = 0.70$, $\phi = 3.36$) showed good agreement with a variety of experimental probes of intramembrane protein organization. Fig. 3.1 shows data from lumen-side-up layers of simulated membrane stacks, the perspective from which PSII positions are routinely detected experimentally [109]. LHCII and PSII particles were distributed uniformly and isotropically throughout the layer, with PSII orientations correlated over about 30 nm (2.5 PSII widths). Statistics of PSII–PSII separation distances matched well with experimental results, with computed and experimental nearest-neighbor PSII center-to-center distances of 18.7 ± 2.9 nm and 19.0 ± 4.1 nm, respectively, and weakly structured radial distribution functions $g(r)$ (Fig. 3.1b and c, [107]); small offsets in peak positions could be due to our simplified particle shapes. These correlations arise both from direct association between PSII pairs and from effective forces mediated by LHCIIs and by particles in the opposing layer. Contributions from LHCII fluctuations include "depletion" attractions [11], which tend to cluster PSIIs in order to maximize the space available to the smaller LHCII species.

Though simple in form, our model of PSII–LHCII intralayer attraction was sufficient to stabilize significant populations of several known supramolecular complexes—PSII supercomplex dimers ($(C_2S_2)_2$), PSII supercomplex with one additional LHCII (C_2S_2M), and PSII supercomplex with two additional LHCIIs ($C_2S_2M_2$)—in equilibrium with a pool of lone PSII

Table 3.1: Structural parameters for closed (canonical ensemble) simulations.

N_{P} per layer	N_{L} per layer	box side length (nm)	packing fraction ρ	LHCII:PSII ratio ϕ	PSII/μm^2
128	100	270	0.550	0.78	1756
128	240	270	0.613	1.88	1756
128	320	270	0.650	2.50	1756
128	430	270	0.700	3.36	1756
128	540	270	0.750	4.22	1756

supercomplexes (C_2S_2). The balance of this equilibrium can be tuned by changing the strength of the intralayer attraction; at $\epsilon_{SL-P} = \epsilon_{ML-P} = 2\ k_BT$, LHCII binding and unbinding is facile, and approximately 25% of LHCII particles are bound to PSII supercomplexes under these conditions.

3.2 LHCII stacking creates interlayer PSII correlations

Our simple model of LHCII stacking interactions similarly proved sufficient to generate a range of structural motifs that have been inferred from experiment. Specifically, we frequently observed (1) pairs of stacked, nearly parallel PSIIs, in which the LHCIIs chirally embedded in one PSII are both aligned with the embedded LHCIIs of the opposing PSII (of opposite chirality) (Fig. 3.4a); (2) nearly perpendicular pairs, in which all embedded LHCIIs stack over free LHCIIs in the other layer (Fig. 3.4b); and (3) rows of rotated PSIIs in one layer stacked atop a parallel but oppositely rotated row in the other layer, in which the two LHCIIs embedded in a given PSII are aligned with those of two different PSIIs in the opposing layer (Fig. 3.4c). In the absence of LHCII stacking interactions (i.e., when $\epsilon_{L-L} = 0$), these PSII correlations disappeared (Fig. 3.2).

All three of these motifs have been reported in separate experimental studies: Fig. 3.4a resembles single-particle electron micrographs of isolated PSII pairs [139], Fig. 3.4b resembles cryo-electron tomograms of PSII arrays in thylakoids [45], and Fig. 3.4c resembles transmission electron micrographs of PSII arrays in BBY membranes [25]. From those observations, however, it was not clear whether different motifs could be simultaneously abundant; nor could those authors establish LHCII stacking as a sufficient driving force for various interlayer correlations. Our results indicate that LHCII stacking can indeed drive all of these associations among PSIIs in opposing layers.

3.3 Simulated PSII arrays depend on packing fraction and attraction strength

Above a packing fraction of $\rho \approx 0.7$, simulations exhibited sizable ordered arrays, featuring alternating rows of PSIIs and LHCIIs (Fig. 3.4d-f). Interestingly, PSIIs in these configurations do not engage in direct intralayer attractions. Instead, adhesion between PSII rows is provided by intralayer attractions to the interspersed M-LHCIIs that bridge between PSII rows, similar to the bridging seen in the C_2S_2M arrays in Ref. [25]. Although arrays with only C_2S_2 unit cells have been observed, especially in the absence of CP24 [112], arrays with additional LHCIIs are more common in wild-type plants [110], suggesting that M-LHCIIs also play a role in array formation in vivo.

Each row of PSIIs is stabilized by stacking interactions with a PSII row in the opposing layer, as in Fig. 3.4c (and by the less geometrically specific depletion attractions). This key role for stacking of embedded LHCIIs is highlighted by an absence of ordered arrays

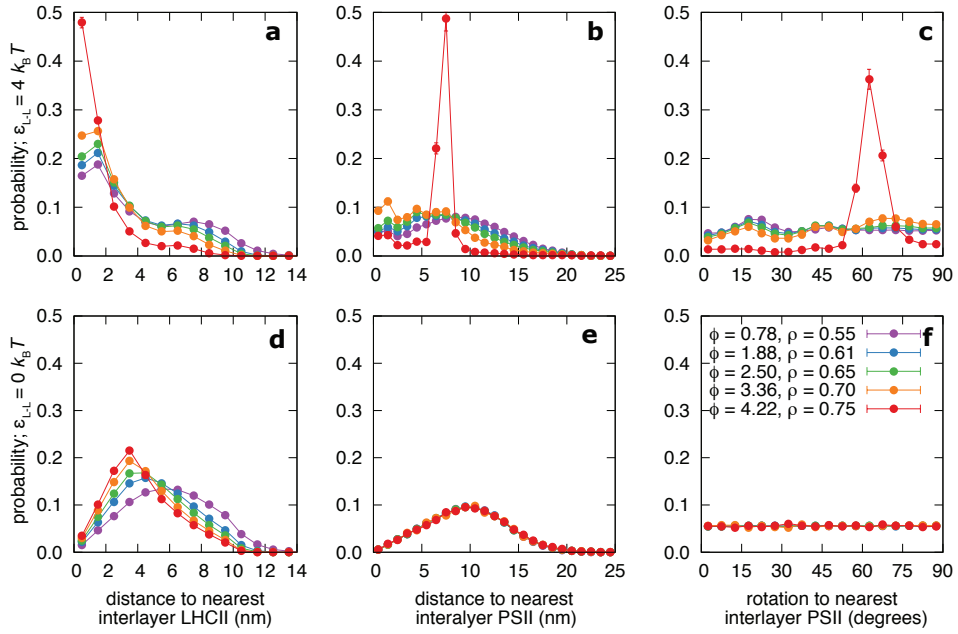


Figure 3.2: Probability distributions of interlayer distances and angles in canonical ensemble simulations with 1750 PSII particles per μm^2. Interlayer distance is the projected distance between particle centers in opposing layers and does not include any implicit vertical separation. **a-c**: Probability distributions in the presence of LHCII stacking interactions ($\epsilon_{\mathrm{L-L}} = 4\ k_{\mathrm{B}}T$). **d-f**: Probability distributions in the absence of LHCII stacking interactions ($\epsilon_{\mathrm{L-L}} = 0\ k_{\mathrm{B}}T$). In all conditions, stacking increases the probability of LHCII closer than 2 nm (**a** vs **d**) and PSII closer than 5 nm (**b** vs **e**). The large peaks in the $\rho = 0.75$ data in panels **a-c** are due to interlayer LHCII contacts in crystals.

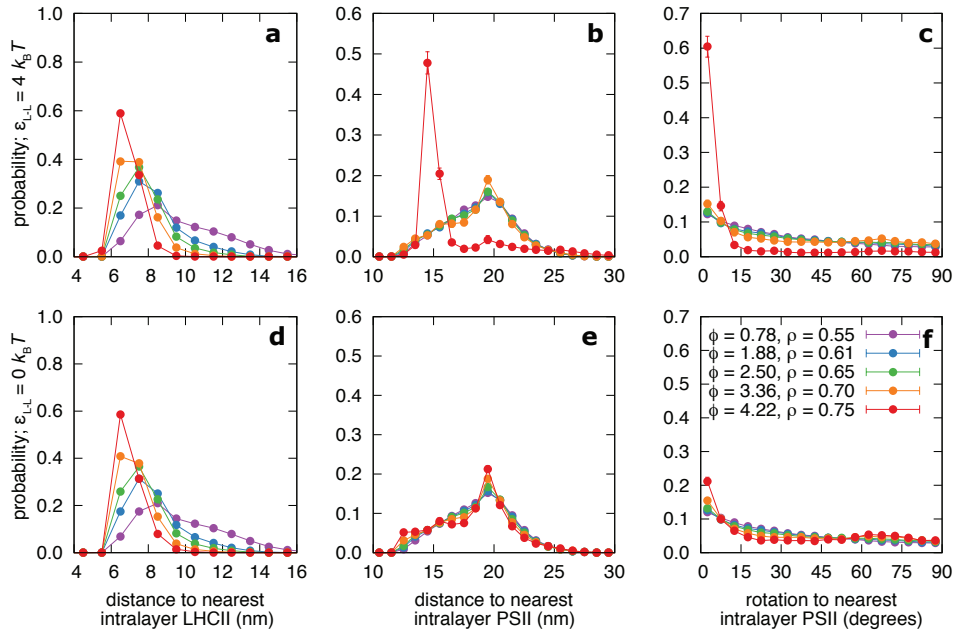


Figure 3.3: Probability distributions of intralayer distances and angles in canonical ensemble simulations with 1750 PSII particles per μm^2. Intralayer distance is the distance between particle centers in the same layer; hard-core replusions prohibit LHCII–LHCII separations of less than 6.5 nm and PSII–PSII separations of less than 12 nm. **a-c**: Probability distributions in the presence of LHCII stacking interactions ($\epsilon_{\mathrm{L-L}} = 4\ k_{\mathrm{B}}T$). **d-f**: Probability distributions in the absence of LHCII stacking interactions ($\epsilon_{\mathrm{L-L}} = 0\ k_{\mathrm{B}}T$). Stacking has little effect on the nearest-neighbor distance and angle distributions in the absence of crystals ($\rho < 0.75$). The large peaks in the $\rho = 0.75$ data in panels **b** and **c** are due to protein packing in crystals.

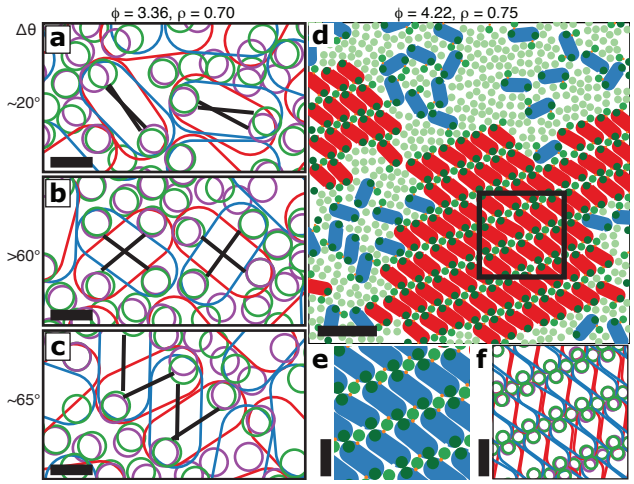


Figure 3.4: LHCII stacking effects on PSII organization. **a**-**c**: PSII interlayer motifs, as described in the text. Green and blue outlines: LHCII and PSII particles in the upper (lumen-side-up) layer; purple and red outlines: LHCII and PSII particles in the lower (stroma-side-up) layer. Black lines drawn parallel to the long axis of selected rods highlight orientational relationships between PSII particles in different layers. Orientation correlations $\Delta\theta$ also appear in Fig. 3.2c. Scale bars = 10 nm. **d**: Snapshot of the top layer from a simulation at $\phi = 4.22$, $\rho = 0.75$ showing a representative array. Color scheme as in Fig. 2.1, except arrayed PSII are colored red (see Chapter 2). Scale bar = 50 nm. **e** and **f**: Magnified views of the boxed region in panel **d**. Scale bars = 20 nm. The role of intralayer attractions is highlighted in **e** by showing only particles in the top layer and indicating the locations of PSII–LHCII interaction sites on each PSII. The stabilizing role of stacking is highlighted in **f** by showing particles from both layers in outline form (as in panels **a**-**c**).

in simulations with $\epsilon_{\mathrm{L-L}} = 0$ (Fig. 3.3). The experimental correlation between arrays and stacking [45] supports the realism of our coarse-grained model, as well as the conclusion that both modes of attraction in our model are essential for ordering.

3.4 A fluid-crystal phase transition is manifested in osmotic ensemble simulations

Configurations in which tens or hundreds of PSIIs cluster tightly together, such as in Fig. 3.4 and in many EM images, reflect strong emergent forces of association, sufficient to offset the entropic cost of sequestering the constituent complexes from the surrounding disordered environment. Computer simulations allow us to address whether these large arrays further signify a more profound underlying phenomenon, namely, a true phase transition from a disordered "fluid" of relatively low PSII density to a system-spanning crystal of tightly

packed PSIIs. The simulations we have described thus far are not suited to address this question, nor would be experiments probing isolated grana membranes with fixed protein content. In these situations crystallization could not proceed to completion, simply because the system's net composition and total area are fixed at values inconsistent with the crystalline phase. Indeed, in simulations of closed systems like those shown in Fig. 3.4, array growth halted before all PSII had been incorporated, creating a dynamic equilibrium between arrayed and disordered PSII (Fig. 3.5).

Phase transitions can be more readily identified by studying open systems such as the osmotic ensemble, in which area and composition can fluctuate subject to external fields at fixed temperature [125]. First, we imposed a 2D "pressure" p that regulates changes in total area. More precisely, we fixed the osmotic pressure that particles experience parallel to the plane of the membrane; simulated area fluctuations implicitly represent addition or subtraction of lipids from the membrane. Second, we allowed the population of one component (we chose LHCII) to fluctuate at fixed chemical potential μ_{L}. Corresponding number fluctuations in a real system involve exchanging material with a very large bath. In intact thylakoids, the stroma lamellae and other connected grana stacks could play the role of a bath.

Our examination of detailed phase behavior focused on the type of C_2S_2M array shown in Fig. 3.4d by restricting the model interactions. In the simulations described below, intralayer attractions to free LHCIIs were retained by maintaining $\epsilon_{\mathrm{ML-P}} = 2\ k_{\mathrm{B}}T$, while attractions between PSIIs and embedded LHCIIs were omitted by setting $\epsilon_{\mathrm{SL-P}} = 0$. This simplification did not affect the internal energy of the C_2S_2M array and ensured that $C_2S_2M_x$ complexes were the only single-layer supramolecular structures directly stabilized by model energetics. The prevalence of $C_2S_2M_x$ complexes in *Arabidopsis* [35] and *Chlamydomonas* [172] suggests that the restricted model may be particularly appropriate for these organisms. We expect the fully interacting model to exhibit qualitatively similar phase behavior, although there may be an additional C_2S_2 phase.

Varying pressure at fixed chemical potential μ_{L} produced a sharp change in average packing fraction (Fig. 3.7a), indicating a highly cooperative transition. The concomitantly sudden appearance of a system-spanning PSII array (Fig. 3.7c, Fig. 3.6) suggests that the degree of cooperativity would grow with system size, as in a first-order phase transition. The low- and high-density phases are distinguished as well by substantially different compressibilities $\rho^{-1}\,(\partial\rho/\partial p)_{T,\mu_{\mathrm{L}},N_{\mathrm{P}}}$; we therefore liken them to a low-density fluid and a tightly packed crystal, respectively.

For the finite, micron-scale system that we simulated, the jump in packing fraction ρ is necessarily rounded and could only become discontinuous in the thermodynamic limit of an infinitely large system. Demonstrating true phase behavior would require an analysis of scaling as this limit is approached. Given the limited spatial extent of natural thylakoids, we instead scrutinized remnant hallmarks of phase coexistence in a finite system, specifically bistable free energy profiles and the presence of stable interfaces.

We computed the free energy $F(\rho)$ as a function of packing fraction for specific values of the thermodynamic parameters (p, μ_{L}, T) using umbrella sampling (Fig. 3.8). Free energy profiles

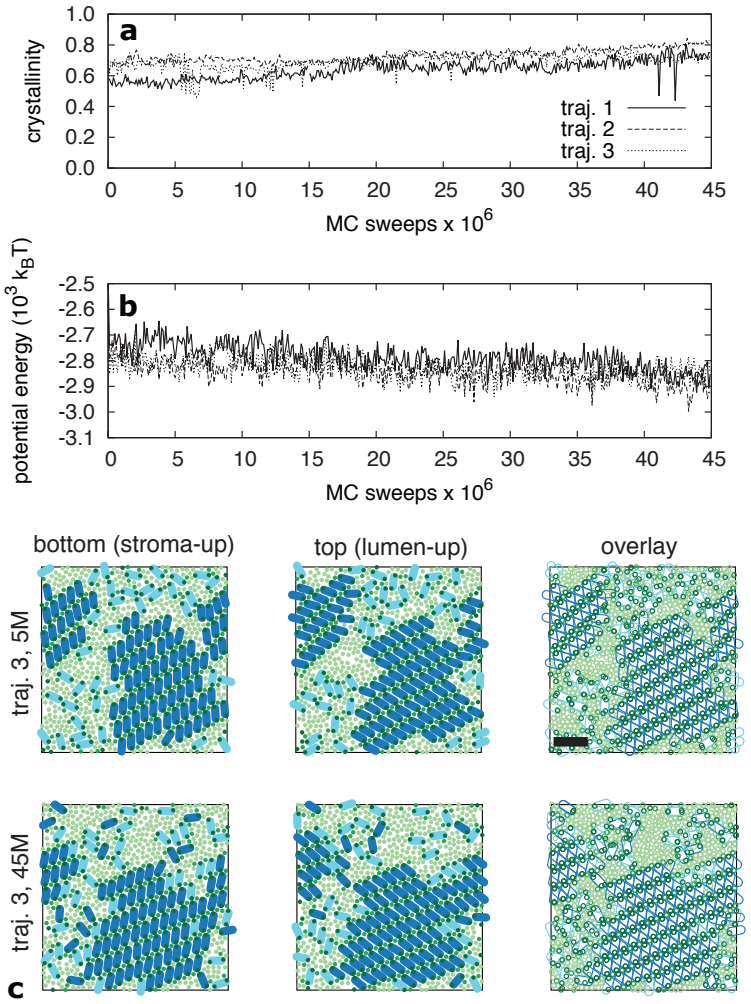


Figure 3.5: Trajectories of long-lived crystalline order in simulations with 1750 PSII/μm^2. **a** and **b**: Trajectories of crystallinity and potential energy from three representative simulations at $\rho = 0.75$, $\phi = 4.22$. In **a**, crystallinity is defined as the fraction of PSII particles in the largest crystalline array, as determined with the clustering algorithm. Crystallinity and potential energy are both stable over the course of 45 million MC sweeps, with a slight trend toward higher crystallinity and lower potential energy. **c**: Snapshots from trajectory 3 after 5×10^6 (first row) and 45×10^6 (second row) MC sweeps, with periodic boundary conditions in x and y. In both snapshots, PSII particles that are part of the largest crystalline array at the earlier time point are dark blue, and other PSII particles are light blue. At the later time point, some formerly crystalline PSIIs have dissociated and some formerly non-crystalline PSIIs have joined the crystal, showing that the system is in dynamic equilibrium. Scale bar = 50 nm.

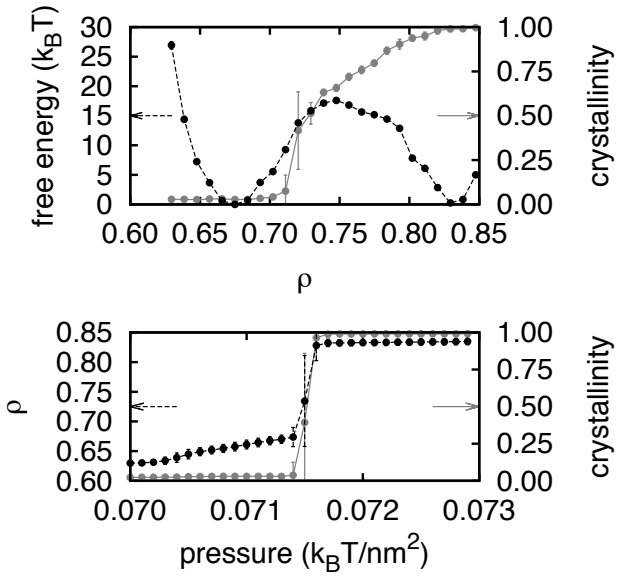


Figure 3.6: Crystallinity in reweighted osmotic ensemble simulations. Crystallinity is the fraction of PSII particles in the largest crystalline array, as determined with the clustering algorithm. Top panel: Free energy (dotted black line) and crystallinity (solid gray line) as a function of packing fraction, at $\bar{\mu}_L = 0.2\ k_BT$, $p = 0.0715\ k_BT/nm^2$. Bottom panel: Packing fraction ρ (dotted black line) and crystallinity (solid gray line) as a function of pressure, at $\bar{\mu}_L = 0.2\ k_BT$. Whiskers are standard deviations. Crystallinity is ≈ 0 in the fluid phase, intermediate on the free energy barrier (top) or at coexistence (bottom), and ≈ 1 in the crystal phase.

at many other values of these parameters were then calculated by thermodynamic reweighting [160]. These profiles exhibit two distinct basins, with minima at $\rho_f < 0.7$ and $\rho_c > 0.8$, over a range of pressures (Fig. 3.7b). The low-packing fraction minimum corresponds to a disordered PSII–LHCII two-component fluid (bottom of Fig. 3.7c); configurations representative of the high-packing fraction minimum are nearly perfect PSII–LHCII co-crystals (top left of Fig. 3.7, Fig. 3.6). Constraining the packing fraction to lie midway between ρ_f and ρ_c produces heterogeneous structures in which fluid and crystal coexist, separated by a system-spanning interface (Fig. 3.7c). Together with observations of hysteresis when pressure is cycled above and below the coexistence pressure p^* (Fig. 3.9), these observations point strongly to a first-order phase transition.

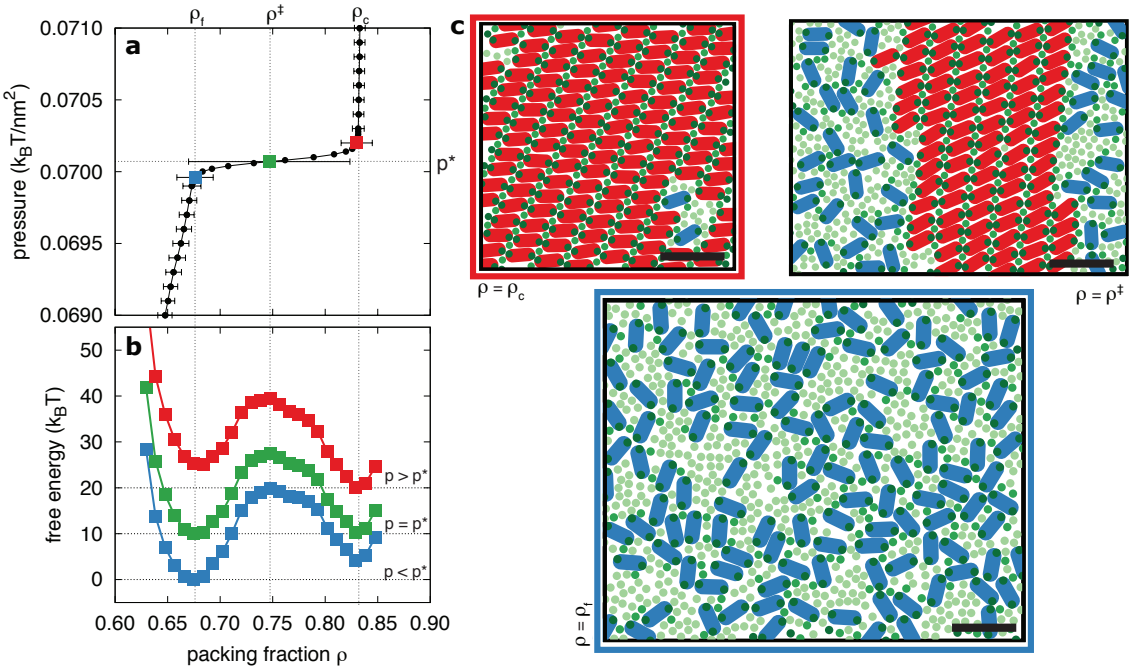


Figure 3.7: Umbrella sampling simulations provide evidence for a phase transition. **a**: Applied pressure p versus packing fraction ρ along a line of constant chemical potential shows a sharp crossover at p^* from a low-pressure, low-packing fraction regime to a high-pressure, high-packing fraction regime. Means (line and points) and root-mean-squared fluctuations (whiskers) of ρ at each pressure are calculated from probability distributions derived from free energy surfaces like those in panel **b**. Fluctuations, and therefore whiskers, are large in the vicinity of p^*; for clarity we show only one whisker in this region. **b**: Free energy as a function of ρ for a system at relative chemical potential $\bar{\mu}_L = 0.1\ k_BT$ and three values of pressure near coexistence: within the fluid phase (blue, $p = 0.06996\ k_BT/nm^2$), within the crystalline phase (red, $p = 0.07020\ k_BT/nm^2$), and at coexistence (green, $p^* = 0.07007\ k_BT/nm^2$). Error bars estimated from the MBAR method are smaller than the symbols. Because the zero of free energy is arbitrary at each pressure, curves are vertically offset for clarity. Dotted lines are guides to the eye. **c**: Snapshots taken from umbrella sampling simulations biased to the stated packing fractions. Color scheme as in Fig. 3.4d. Scale bars = 50 nm.

3.5 Phase coexistence region includes physiological conditions

From free energy profiles like those of Fig. 3.7b, we constructed a phase diagram in the plane of packing fraction ρ and composition ϕ. For a given chemical potential μ_L^*, the coexistence pressure p^* can be uniquely identified as the pressure that maximizes the variance of packing fraction fluctuations; note the large standard deviation at p^* in Fig. 3.7a. Values of

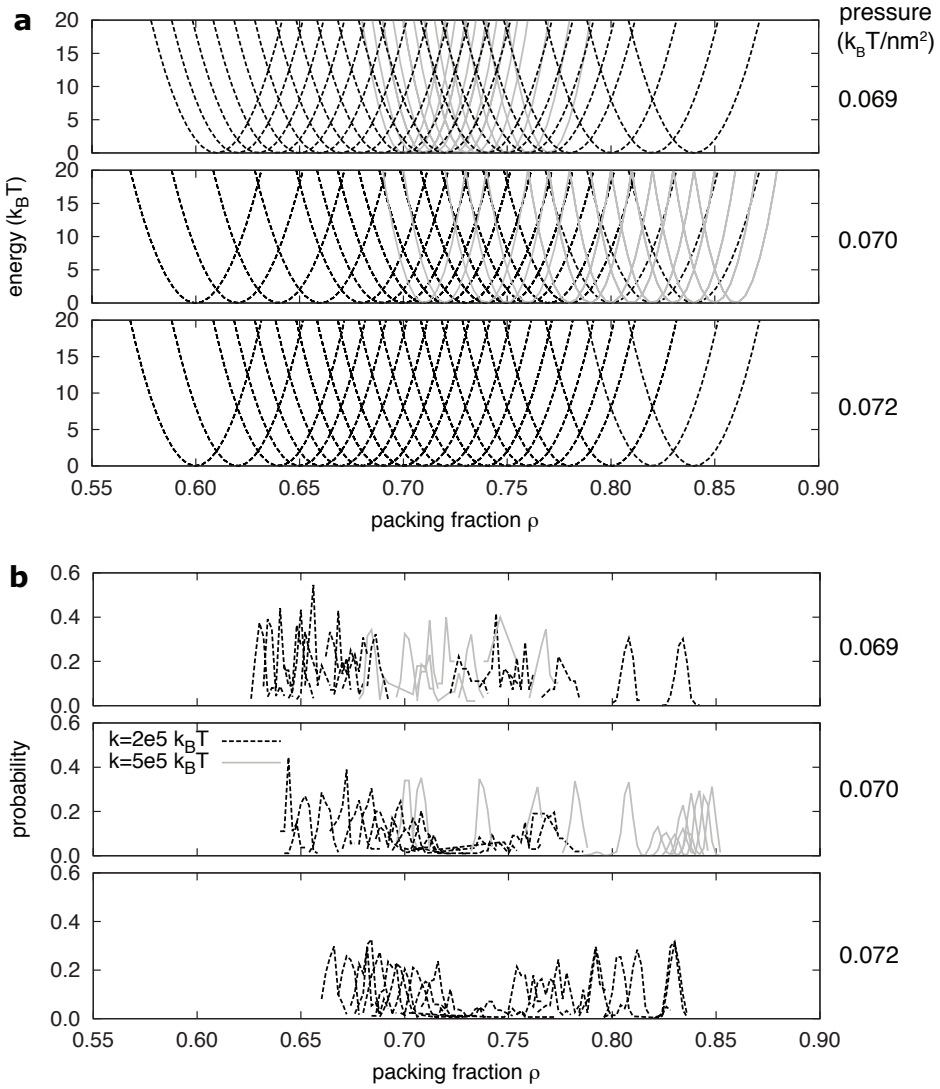


Figure 3.8: Details of the umbrella sampling simulations. Statistics of ρ were determined by maximum likelihood estimation from a set of 139 osmotic ensemble simulations ($\bar{\mu}_{\mathrm{ex,L}} = 0.1$ $k_{\mathrm{B}}T$ relative to a standard state; $p = 0.069, 0.070, 0.072$ $k_{\mathrm{B}}T/\mathrm{nm}^2$). **a**: Harmonic potentials of the form $u_{\mathrm{bias}}(\rho) = k(\rho - \rho_0)^2$ used to bias simulations and enhance sampling along the order parameter ρ. Values of ρ_0 and k were chosen to focus sampling on a series of small intervals spanning the range $\rho = 0.625$ to $\rho = 0.852$. **b**: Normalized histograms of uncorrelated samples of ρ obtained from the umbrella sampling simulations, with $\bar{\mu}_{\mathrm{L}} = 0.1$ $k_{\mathrm{B}}T$. Dotted black lines: $k = 2 \times 10^5$ $k_{\mathrm{B}}T$; solid gray lines: $k = 5 \times 10^5$ $k_{\mathrm{B}}T$. Top, middle, and bottom panels in each subfigure are for simulations at $p = 0.069$, 0.070, and 0.072 $k_{\mathrm{B}}T/\mathrm{nm}^2$, respectively.

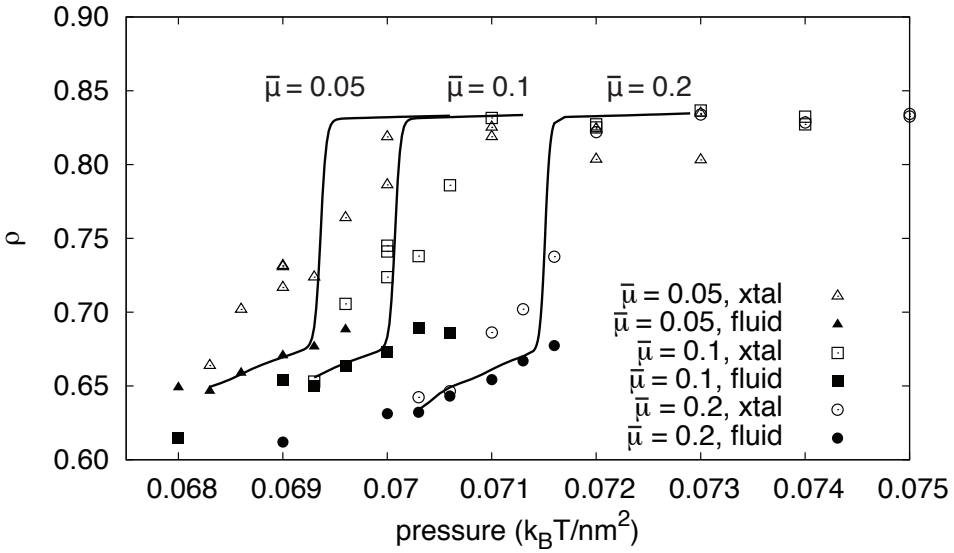


Figure 3.9: Comparison between straightforward and umbrella sampling simulations in the osmotic ensemble. Axes are reversed compared to Fig. 3.7a to put the independent variable (p) on the x-axis. Open and filled symbols: straightfoward osmotic ensemble simulations initialized from a crystalline or fluid configuration, respectively, at the stated values of $\bar{\mu}_{\mathrm{L}}$ (in units of $k_{\mathrm{B}}T$). Hysteresis in the straightforward simulations is present even after tens of millions of Monte Carlo sweeps. Umbrella sampling (lines) gives excellent agreement with straightforward simulations in the low and high density regimes along each line of constant $\bar{\mu}_{\mathrm{L}}$, and exhibits a sharp transition in between.

ρ and ϕ for the two phases at the thermodynamic state $(\mu_{\mathrm{L}}^*, p^*)$ determine a pair of points at the boundaries of the crystal–fluid coexistence region. Repeating this procedure for different values of μ_{L}^*, we traced out coexistence curves bounding regions where homogeneous fluid and crystal phases are thermodynamically stable (Fig. 3.10).

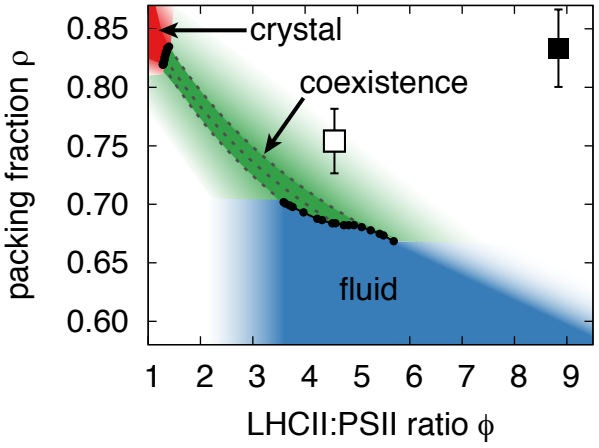


Figure 3.10: Phase diagram of crystal–fluid coexistence in the (ϕ, ρ) plane. Each thermodynamic state $(\mu_{\mathrm{L}}^*, p^*)$ of phase equilibrium maps in this plane to two points and a corresponding tie line. One point lies at the boundary of the homogeneous fluid phase (low packing fraction and high LHCII content, shaded blue), the other at the boundary of the fully crystalline state (high packing fraction and low LHCII content, shaded red). Example tie lines are drawn as dashed lines connecting these boundaries for selected values of p^*. Coexistence (shaded green) and pure-phase regions extend beyond the shaded areas determined by our limited data. In particular, the coexistence region is expected to span the entire upper-right-hand quadrant. Filled and open symbols are low-light and ordinary-light phase points, respectively, calculated from data in Ref. [102] (see Chapter 2).

The dashed lines in Fig. 3.10 trace tie lines, i.e., lines along which the relative extent of the two pure phase regions varies while the packing fraction and composition of each phase remains constant (as determined by the endpoints). These tie lines thus enable straightforward predictions for thermodynamic states in the coexistence region, requiring no characterization beyond the physical properties of the endpoints.

Remarkably, the resulting coexistence region encompasses the grana packing fractions and compositions reported from many in vivo and in vitro experiments. For instance, the filled and open squares in Fig. 3.10 are experimental measurements from Ref. [102]. Other reports suggest that ϕ is typically in the range $\approx$ 2–6 [90, 102, 180], and that ρ is typically in the range $\approx$ 0.6–0.8 [133, 107, 70, 110]. Thus, our model of the LHCII–PSII protein system supports coexistence in some physiologically relevant conditions.

3.6 Discussion

We have demonstrated that simulations of a simple model of PSII and LHCII in stacked grana membranes, when configured to represent a generic grana membrane, can recapitulate many disparate and nontrivial experimental observations. These behaviors emerge spontaneously from the model's short-ranged interactions without the need for us to presuppose any particular target assembled structures. Specifically, we observe a distribution of C_2S_2, C_2S_2M, and $C_2S_2M_2$ complexes (Fig. 3.1a); LHCII-mediated intermembrane associations between PSII supercomplexes (Fig. 3.4a–c); and co-occurence of disordered and crystalline-ordered regions in PSII- and LHCII-rich membranes (Fig. 3.4d). Importantly, all of these qualitative features appear to be common to grana membranes from many photosynthetic organisms and many growth conditions, although the quantitative details of course vary. To the best of our knowledge, this work marks the first reported computational investigation of grana membranes to share these commonalities with experiment.

Our principal prediction from numerical studies of this model is that the appearance of finite arrays of PSII and LHCII signals thermodynamic coexistence of disordered (fluid) and ordered (crystalline) phases. The phase boundaries we have computed further suggest that many physiological conditions lie at or near such coexistence. The experimental data reported in Ref. [102], correlating degree of array formation with the packing fraction and composition of grana membranes, offer the most concrete opportunity for comparison. Membranes at conditions just inside our coexistence region, from plants grown in so-called "ordinary" light (300 μmol photons/m^2/s), were found to exhibit a low degree ($< 2\%$) of crystallinity. Membranes corresponding to conditions deep within our coexistence region, grown in low light (30 μmol photons/m^2/s), showed significantly enhanced (22%) crystallinity. These consistent observations are indicated by the filled and open symbols in Fig. 3.10. The tie line we have computed suggests that a fully equilibrated system in ordinary light would possess greater crystallinity than observed in experiment, possibly reflecting high nucleation barriers or the slow growth characteristic of dynamics near coexistence (see Fig. 3.6). Active processes, such as phosphorylation [170], protein translation, or lipid exchange [58], could also prevent the system from achieving an equilibrated crystalline state. A recent study of low-light-acclimated membranes [110] reported arrays at protein packing fractions ($\rho < 0.6$) much lower than previously measured. Our calculations do not provide a direct way to resolve this apparent discrepancy, unless the different organisms feature substantially different energy scales and/or modes of protein association.

Experiments that systematically quantify dependence of crystallinity upon packing fraction at various protein compositions could further test or exploit our predictions. For example, grana membranes could be isolated from plants grown at different light intensities, generating samples over a range of values of ϕ. Diluting these membranes with additional lipid [70] and measuring crystallinity via AFM or EM would allow construction of experimental phase diagrams analogous to those we have computed. Matching experimental and theoretical results in detail could determine appropriate values of the interaction parameters in our model.

The well-understood phenomenology of phase transitions helps explain why the extent of PSII array formation varies so dramatically between similar samples in experiment. The presence, extent, and number of arrays in the laboratory may be influenced by the characteristically slow dynamics associated with phase transitions. Nucleation of one phase from the other will be governed by rare structural fluctuations (although biological signalling processes could favor the growth of precritical nuclei [93]). Crowding within grana membranes [103, 104] will likely make the subsequent repartitioning of material between phases slow as well. Moreover, grana isolation protocols that increase the protein packing fraction beyond the freezing transition densities for 2d hard discs ($\rho \approx 0.7$ [85]) or hard rods ($\rho \approx 0.8$ [17]), such as BBY [70], could trap the isolated grana in jammed nonequilibrium configurations, further complicating experimental determinations of the equilibrium distribution of PSII arrays. Further work will be required to elucidate these dynamic effects.

Proximity to phase coexistence could also contribute to substantial changes in thylakoid function observed to accompany modest changes in protein content and interactions. In addition to the low-light acclimation scenario discussed above, many regulatory processes associated with non-photochemical quenching, photoprotection, and repair shift the system's position relative to phase boundaries in the ρ, ϕ-plane: (1) State transitions involve transport of LHCII from PSII-rich to PSI-rich membranes [88], reducing the local protein density and LHCII:PSII ratio. (2) Photoinhibition-induced phosphorylation decreases the diameter of grana stacks and breaks up PSII supercomplexes [62, 79], changing the system size and composition. (3) The qE component of nonphotochemical quenching may introduce an additional intramembrane attraction among LHCII [91], affecting the relative stability of the fluid and crystal phases. These spatioregulatory processes are often interpreted as acting primarily over short length scales, tuning exciton fate by changing the relative distances between light harvesting sites, quenchers, and reaction centers. We suggest that, in addition, these processes may directly regulate global protein organization within the thylakoid membrane by inducing cooperative structural transitions.

Suitably adapted, the model and computational framework developed in this study may help to clarify the mechanisms of such spatioregulatory changes. Direct analyses of grana-scale LHCII organization are difficult and rare, with the notable recent exception of Ref. [91]. With a few experimental parameters as input (namely ϕ and ρ or their equivalents), our methods can generate physiologically reasonable LHCII configurations in the presence of PSII in a physically grounded and unbiased fashion, and in sufficient quantity to enable statistical comparisons (e.g., Figs. 3.2 and 3.3). These ensembles of configurations could serve as the foundation for future studies of the many phenomena in which LHCII plays a key role.

Chapter 4

Vignettes from beyond the grana, or, extended models capture some structural features of NPQ

The composition and total packing fraction of membrane proteins vary between the grana and stroma lamellae regions of the thylakoid membrane [165]; this is often termed "lateral heterogeneity" in the literature. The physicochemical driving forces that maintain these density and composition gradients are not well characterized [38]. Compelling hypotheses have been put forward for some aspects of the spatial variations in composition—PSI and ATP synthase are likely to be sterically excluded from tightly appressed grana membranes by their bulky stroma-side protrusions (Figure 1.1), while attractive stacking interactions between LHCII complexes (Chapter 3) are likely to favor their accumulation in the grana. Yet these hypotheses are insufficient to explain the density gradient between grana and stroma lamellae, or why PSII complexes accumulate in the denser grana region. A better understanding of lateral heterogeneity would shed light on regulatory processes in which proteins redistribute between thylakoid membrane regions, namely state transitions and PSII repair.

In this Chapter, we approach this problem by extending the model developed in Chapter 2 to include coarse-grained representations of the key thylakoid membrane proteins that were neglected in the grana-only model, and to include effective external potentials with spatial variation between grana, stroma lamellae, and marginal regions of the simulation box. Using interaction strengths similar to those in Chapter 3 in simulations parameterized by data from Helmut Kirchhoff's group, we find that LHCII stacking energetics and entropic effects in the grana are sufficient to account qualitatively for most of the composition gradients observed in vivo both before and after photodamaging high-light treatment; external potentials can be tuned to achieve better quantitative agreement. Surprisingly, these simulations do not display gradients in the total density between subsystems; further work is needed to resolve this discrepancy. In the last section of this Chapter, we return to the grana and investigate a simple model of LHCII aggregation during qE, which also displays partial agreement with

experiments. This work was informed and guided by discussions with Helmut Kirchhoff, Kapil Amarnath, and as always Phill Geissler.

4.1 An extended model supports composition gradients within the thylakoid

In Chapters 2 and 3, we focused on the two protein complexes that make up the vast majority of the grana protein content, LHCII and PSII. A full model of the thylakoid should also include the other proteins in Figure 1.1: cyt b_6f, PSI, and ATPase. To represent these complexes in our coarse-grained model, here we introduce three additional species of disc-shaped particles, with hard-core diameters $\sigma_{\mathrm{PSI}} = 15.5$ nm (with LHCIs [24]), $\sigma_{\mathrm{ATPase}} = 10.0$ nm [1], and $\sigma_{\mathrm{cytb6f}} = 8.0$ nm [113]; these particle sizes are similar to those in used in the models of [105, 52]. To the best of our knowledge, the only co-complex of these proteins that has been isolated and characterized via single-particle EM is the PSI-LHCII complex involved in state transitions [122, 63]; we neglect this specific binding interaction here and treat all cyt b_6f, PSI, and ATPase particles as being subject to no pairwise potentials besides hard-core repulsions.

Physiologically-relevant composition gradients between the stacked grana G and the unstacked margin M and stroma lamellae S regions of the simulation box were favored by modifying LHCII stacking interaction in Eq. 2.1 such that it was only active if both LHCII particles were located in the grana region. Specifically, for two LHCII particles located at $\mathbf{r}_i$ and $\mathbf{r}_j$ in opposing simulation layers,

$$u_{\mathrm{stacking,L}}(\mathbf{r}_i, \mathbf{r}_j) = \begin{cases} -\epsilon_{\mathrm{L-L}} \left(\frac{|\mathbf{r}_i - \mathbf{r}_j| - \sigma_{\mathrm{L}}}{\sigma_{\mathrm{L}}} \right)^2 & \mathbf{r}_i, \mathbf{r}_j \in G \text{ and } |\mathbf{r}_i - \mathbf{r}_j| < \sigma_{\mathrm{L}} \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

The most direct effect of this interaction is to favor the localization of LHCII (either free or embedded in PSII particles) to the grana. This modification captured the fact that grana are defined by the presence of tight stroma-side appression, while the other membrane regions (margins and stroma lamellae) are defined by its absence. With this model alone, the margin and stroma subsystems are energetically identical; later we add external potentials that distinguish between them.

With this extended model in hand, the challenge is now to identify the range of parameters for which the model recapitulates experimentally-observed compositions within each of the three thylakoid subsystems. This is a nontrivial task due to the large number of parameters (Table 4.1). A canonical ensemble simulation is specified by 3 subsystem areas A_i or area fractions f_i, 5 particle counts N_α or particle number densities η_α, and the 3 energetic parameters of the original model ($\epsilon_{\mathrm{L-L}}$, $\epsilon_{\mathrm{SL-P}}$, and $\epsilon_{\mathrm{ML-P}}$). Our approach is to estimate f_i and η_α from experiment (as described below), equilibrate simulations under a range of potentials, and identify Hamiltonians for which the 3 subsystem packing fractions ρ_i and 15 mole fractions $m_{\alpha,i}$ best approximate the experimental values.

variable	description	definition	reported in expt?	fixed in sim?
$N_{\alpha,i}$	number of particles of type α in subsystem i		no	no
A_i	area of subsystem i		no	yes
N_α	total number of particles of type α	$\sum_i N_{\alpha,i}$	no	yes
A	total area	$\sum_i A_i$	no	yes
f_i	fraction of total area in subsystem i	A_i/A	yes	yes
η_α	overall number density of particle α	N_α/A	no	yes
$\eta_{\alpha,i}$	number density of particle α in subsystem i	$N_{\alpha,i}/A_i$	yes	no
$\phi_{\alpha:\beta}$	overall stoichiometry of particle α to particle β	N_α/N_β	yes	yes
$\phi_{\alpha:\beta,i}$	stoichiometry of particle α to particle β in subsystem i	$N_{\alpha,i}/N_{\beta,i}$	yes	no
$m_{\alpha,i}$	mole fraction of particle α in subsystem i	$N_{\alpha,i}/N_\alpha$	yes	no

Table 4.1: Quick reference for variables in the thylakoid model.

To parameterize our simulations, we collaborated with the group of Helmut Kirchhoff at Washington State University. Table 4.2 lists their experimental measurements of f_i, $m_{\mathrm{PSII},i}$, and $\phi_{\alpha:\mathrm{PSII},i}$ for $i \in (G, M, S)$ and $\alpha \in$ (LHCII, PSI, cytb6f), and $\eta_{\mathrm{PSII,G}}$, for *Arabidopsis* thylakoids before and after a brief period of high-light illumination that induced the PSII damage-and-repair cycle. In detail, they measured the ratios Φ of monomeric PSI, monomeric cyt b_6f, and total trimeric LHCII to monomeric PSII, while our simulations require the ratios ϕ of monomeric PSI, dimeric cyt b_6f, and free trimeric LHCII to PSII C_2S_2 supercomplexes; this conversion is accomplished by

$$\phi_{\alpha:\mathrm{PSII},i} = \begin{cases} 2\Phi_{\alpha:\mathrm{PSII},i} - 2 & \alpha = \text{LHCII} \\ 2\Phi_{\alpha:\mathrm{PSII},i} & \alpha = \text{PSI} \\ \Phi_{\alpha:\mathrm{PSII},i} & \alpha = \text{cyt } b_6f \end{cases} \tag{4.2}$$

This data set is sufficient to calculate the number density of each particle species α in the full canonical ensemble simulation box, i.e.,

$$\eta_{\mathrm{PSII}} = \eta_{\mathrm{PSII},G} f_G / m_{\mathrm{PSII},G} \tag{4.3}$$

$$\eta_\alpha = \eta_{\mathrm{PSII}} \sum_{i \in (G,M,S)} m_{\mathrm{PSII},i} \phi_{\alpha:\mathrm{PSII},i} \tag{4.4}$$

The values of $\eta_{\mathrm{PSII},G}$ were measured in isolated grana, which are denser than grana in vivo [70]; after correcting for this systematic error, we selected the parameters in Table 4.3 as a good consensus for both the control and high-light conditions. All simulations were run using these parameters in a two-layer stroma-paired simulation geometry, where each rectangular layer was divided into three regions—grana (left), margin (middle), and stroma lamellae (right)—such that the total area was apportioned according to f_G, f_M, and f_S. All particles could freely diffuse between all regions within a layer via single-particle Monte Carlo moves as in Chapter 2. Simulations were initialized with a C_2S_2M crystal nucleus in the grana

parameter	control	high-light	source
f_G	0.60	0.48	HK
f_M	0.20	0.35	HK
f_S	0.20	0.17	HK
$\Phi_{\text{LHCII:PSII,G}}$	3.76	4.85	HK
$\Phi_{\text{PSI:PSII,G}}$	0.19	0.17	HK
$\Phi_{\text{cytb6f:PSII,G}}$	0.14	0.15	HK
$\Phi_{\text{LHCII:PSII,M}}$	2.92	3.07	HK
$\Phi_{\text{PSI:PSII,M}}$	0.87	0.72	HK
$\Phi_{\text{cytb6f:PSII,M}}$	1.58	1.04	HK
$\Phi_{\text{LHCII:PSII,S}}$	-	1.74	HK
$\Phi_{\text{PSI:PSII,S}}$	1.69	0.86	HK
$\Phi_{\text{cytb6f:PSII,S}}$	0.84	0.44	HK
$\eta_{\text{PSII,G}}$	$1732/\mu\text{m}^2$	$1437/\mu\text{m}^2$	HK
$\eta_{\text{ATPase,G}}$	0	0	assumption
$\eta_{\text{ATPase,M}}$	0	0	low estimate
$\eta_{\text{ATPase,S}}$	$1200/\mu\text{m}^2$	$1200/\mu\text{m}^2$	estimate, [164, 45]
$m_{\text{PSII,G}}$	0.77	0.60	HK
$m_{\text{PSII,M}}$	0.14	0.27	HK
$m_{\text{PSII,S}}$	0.09	0.12	HK

Table 4.2: Experimentally measured parameters of thylakoids in control and high-light conditions. Φs are based on monomeric PSII, PSI, and cytb6f, and trimeric LHCII. ηs are based on C_2S_2 PSII, monomeric PSI, dimeric cytb6f, and trimeric free LHCII. HK: personal communication from Helmut Kirchhoff and Sujith Puthiyaveetil, 2013.

region to aid equilibration; configurations after several million MC sweeps of equilibration are shown in Figure 4.1.

First, we examined the performance of this model with $\epsilon_{\text{L-L}} = 4\ k_{\text{B}}T$ (as in Chapter 3) for the "control" scenario (Figure 4.1a). Figure 4.2a shows how the five particle species partition between the three subsystems in these simulations. We find that $m_{\text{LHCII},G}$ and $m_{\text{PSII},G}$ (black bars) are greater than f_G (lower dotted line), while $m_{\text{PSI},G}$, $m_{\text{ATPase},G}$, and $m_{\text{cytb6f},G}$ are less than f_G; these trends are expected because the LHCII stacking interaction drives PSII and LHCII to localize to the grana subsystem, which then drives steric crowding of the other particles from the grana. These driving forces alone are sufficient to yield partitionings of LHCII and PSII that compare quite favorably to the values of $m_{\text{PSII},i}$ and $m_{\text{LHCII},i}$ measured by Kirchhoff's group (compare bar boundaries to red markers in Figure 4.2a). This agreement further validates our choice of $\epsilon_{\text{L-L}} = 4\ k_{\text{B}}T$ in Chapter 3, and suggests that LHCII stacking is a primary driving force for LHCII and PSII lateral heterogeneity.

On the other hand, the simulated and experimental values of $m_{\alpha,i}$ do not match as well for PSI and ATPase. The simulated value of $m_{\text{PSI},G}$ is much lower than Kirchhoff's value for control membranes, but only slightly higher than typical values in the literature such as

parameter	control	high-light
A	667 nm × 157 nm = 0.105μm^2	667 nm × 157 nm = 0.105μm^2
f_G	0.60	0.48
f_M	0.20	0.35
f_S	0.20	0.17
N_{PSII}	110	110
N_{LHCII}	635	635
N_{PSI}	90	90
N_{ATPase}	25	25
N_{cytb6f}	45	45
ρ	0.705	0.705

Table 4.3: Actual simulation input parameters. All numbers are for a single membrane layer, and based on C_2S_2 PSII, monomeric PSI, dimeric cyt b_6f, and trimeric free LHCII.

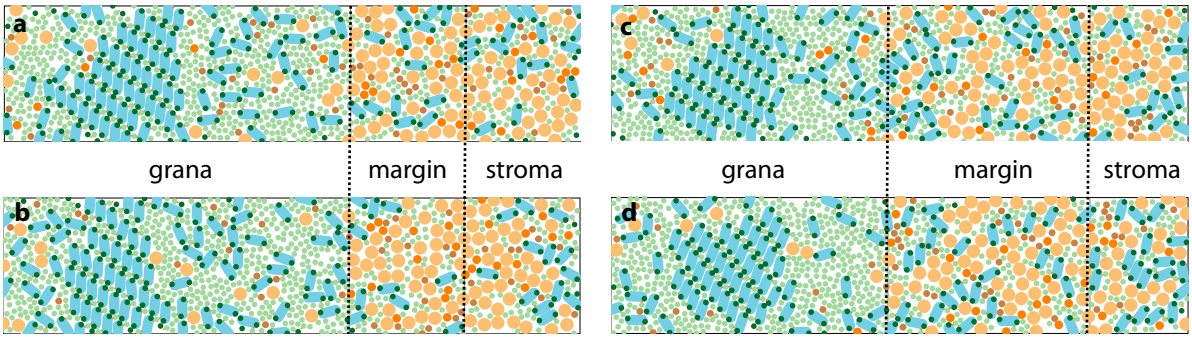
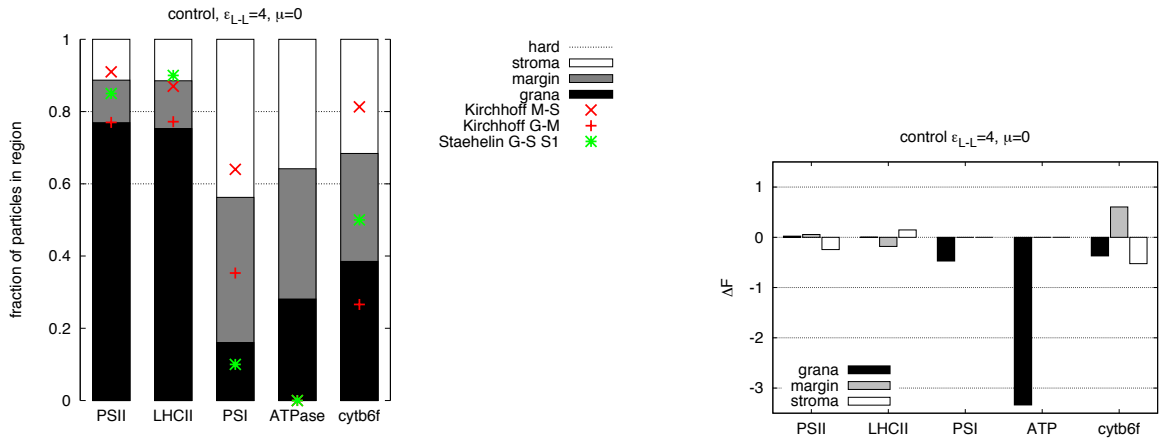


Figure 4.1: Particle configurations taken from the top (lumen-side-up) layer of simulations at the parameters in Table 4.3. (a) and (b) are "control", (c) and (d) are "high-light;" (a) and (c) have $\mu_{\alpha,i} = 0$ for all α, i; (b) and (d) have $\mu_{\mathrm{PSI},G} = 0.5$, $\mu_{\mathrm{ATPase},G} = 3$, and $\mu_{\mathrm{cytb6f},G} = -0.5$ $k_{\mathrm{B}}T$, and with all other $\mu_{\alpha,i} = 0$. Green discs are LHCII particles, blue rods are PSII particles, large tan discs are PSI particles, medium-sized orange discs are ATP synthase particles, and small brown discs are cyt b_6f particles.



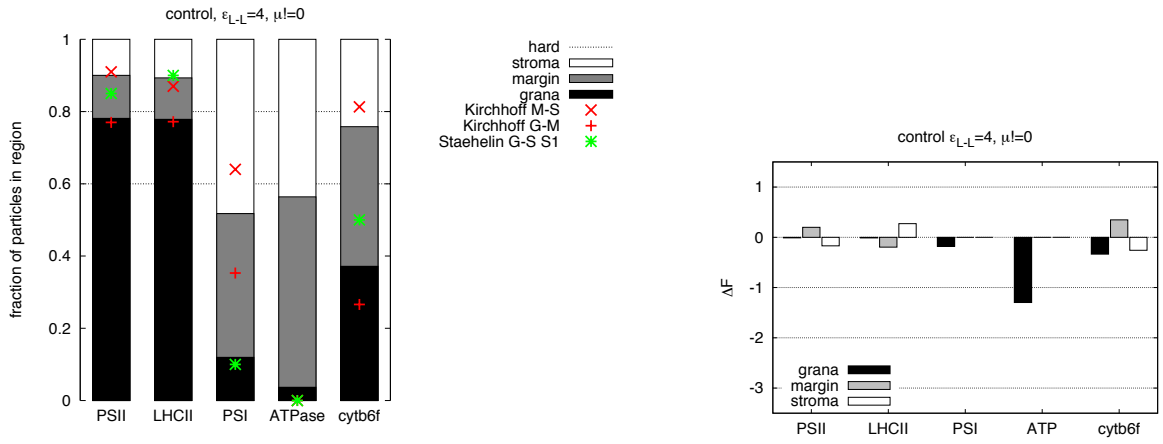
(a) (b)

Figure 4.2: Comparison of particle partitioning in "control" thylakoids between simulation and experiment. (a) Mole fractions $m_{\alpha,i}$ of each particle species in each subsystem. Bars are simulation, markers are experiment (red markers from Kirchhoff, green markers from [165]), dotted lines are the expected mole fractions in a purely hard simulation (i.e., based on the system geometry alone). (b) Free energy difference $\Delta F_{\alpha,i}$ for each particle species in each subsystem, such that $\Delta F_{\alpha,i} > 0$ when the particle experiences an excess repulsion from the subsystem in simulations, and $\Delta F_{\alpha,i} < 0$ when the particle experiences an excess attraction.

[165] (green markers in Figure 4.2a). Because the mole fraction of PSI in grana observed by Kirchhoff is significantly higher than the consensus in the literature, we choose to focus on Staehelin's target value of $m_{\mathrm{PSI},G} = 0.1$ instead. Kirchhoff did not measure $m_{\mathrm{ATPase},i}$ so we focus on Staehelin's value for ATPase as well, i.e., $m_{\mathrm{ATPase},G} \approx 0$, which is much lower than the simulated value. This suggests that PSI and ATPase experience repulsions from the grana in addition to crowding in the plane of the membrane, consistent with the standard theory of lateral heterogeneity due to steric exclusion of lumenal protrusions.

In addition, we find that cyt b_6f is distributed roughly equally between the three subsystems. This distribution agrees qualitatively with published reports [165, 48] but does not agree quantitatively with Kirchhoff's measurements; $m_{\mathrm{cytb6f},M}$ is lower in simulation than in experiment.

To estimate the magnitude of the effective potentials in vivo that are not accounted for by our model, we constructed free energy differences ΔF as the difference in the logs of the probabilities of particle species α appearing in subsystem i in simulation vs. in experiment, i.e., $\Delta F_{\alpha,i} = -\log\left(m_{\alpha,i,\mathrm{sim}}/m_{\alpha,i,\mathrm{expt}}\right)$. Figure 4.2b shows that PSI and ATPase particles experience an excess attraction to the grana of 0.5 and 3.3 $k_{\mathrm{B}}T$, respectively, in simulations of control membranes, while cyt b_6f particles experience an excess repulsion from the margin



(a) (b)

Figure 4.3: Comparison of particle partitioning in "control" thylakoids between simulation and experiment, with $\mu_{\mathrm{PSI},G} = 0.5$, $\mu_{\mathrm{ATPase},G} = 3$, and $\mu_{\mathrm{cytb6f},G} = -0.5\ k_{\mathrm{B}}T$, and with all other $\mu_{\alpha,i} = 0$. Rest of key as in Figure 4.2.

of 0.6 $k_{\mathrm{B}}T$ and an excess attraction to the stroma of 0.5 $k_{\mathrm{B}}T$. All other excess interactions are less than 0.5 $k_{\mathrm{B}}T$ in absolute value; given the number of assumptions we have made in constructing the simulations and interpreting the experiments, we consider $|\Delta F_{\alpha,i}| < 0.5\ k_{\mathrm{B}}T$ to indicate quite good agreement.

Next, we incorporated these effective interactions by subjecting each particle species α to a generic spatially-varying external potential

$$u_\alpha(\mathbf{r}_i) = \begin{cases} \mu_{\alpha,G} & \mathbf{r}_i \in G \\ \mu_{\alpha,M} & \mathbf{r}_i \in M \\ \mu_{\alpha,S} & \mathbf{r}_i \in S. \end{cases} \tag{4.5}$$

These potentials can be thought of as accounting in a very coarse-grained, mean-field way for any chemical details that vary between regions and that we do not explicitly include in the model, such as lipid-mediated effects, lumen-side interactions, curvature effects, or, in particular, steric repulsions of stromal protrusions from grana stacks. Figures Figure 4.1b and 4.3 show the results of simulations with $\mu_{\mathrm{PSI},G} = 0.5$, $\mu_{\mathrm{ATPase},G} = 3$, and $\mu_{\mathrm{cytb6f},G} = -0.5\ k_{\mathrm{B}}T$, and with all other $\mu_{\alpha,i} = 0$. As expected, the inclusion of these external potentials on PSI, ATPase, and cyt b_6f particles decreased the magnitude of the unaccounted-for interactions of those species. Importantly, the redistribution of these particles did not dramatically affect the distribution of PSII and LHCII particles; $|\Delta F_{\mathrm{PSII},i}|$ and $|\Delta F_{\mathrm{LHCII},i}|$ remained less than 0.5 $k_{\mathrm{B}}T$ for all subsystems i.

In summary, we find that an extension of the model in Chapter 2 that includes simple

coarse-grained representations of PSI, ATP synthase, and cyt b_6f complexes, and adds a spatially-varying component to the LHCII stacking interaction, is sufficient to capture much of the lateral heterogeneity seen in thylakoid membranes in experiment. Thus, it appears that LHCII stacking and crowding in the grana are primary driving forces for lateral heterogeneity. The fit of simulated composition gradients can be further tuned to match experimental measurements by imposing spatially-varying external potentials on PSI, ATP synthase, and cyt b_6f particles. The nature of the effective repulsion of PSI and ATP synthase from the grana is easy to rationalize based on sterics, but the physical basis of the apparent slight preference of cyt b_6f for the margin is more mysterious. Further experimental and theoretical work will be necessary to confirm these initial observations and refine our predictions for the nature and magnitude of the driving forces for lateral heterogeneity in vivo.

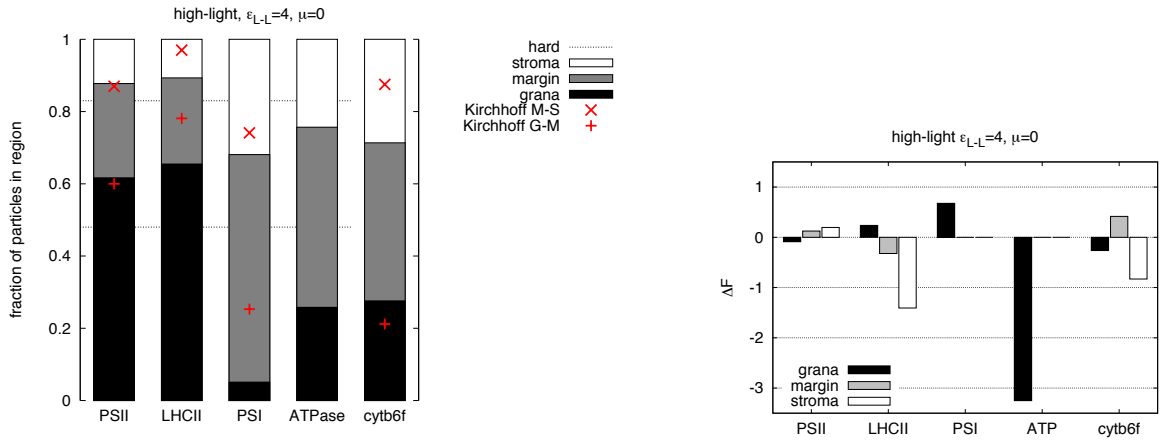
4.2 PSII transport during repair is coupled to changes in grana size

In Kirchhoff's "high-light" experiments from the previous section, PSII photodamage was triggered, causing many PSII supercomplexes to be phosphorylated and disassembled and the PSII cores to migrate to unstacked regions of the thylakoid membrane. In the "control" condition, the most likely driving force for the localization of non-photodamaged PSII to the grana is stroma-side stacking attractions, either between LHCII complexes embedded in PSII supercomplexes or between PSII cores directly. Thus, two hypotheses for PSII core redistribution include a weakening of the LHCII-mediated driving force, due to supercomplex disassembly, or a weakening of the PSII-mediated driving force, due to PSII core phosphorylation. (Both of these hypotheses are often discussed in terms of changes in PSII core diffusion coefficient, but a change in diffusion coefficient does not a priori change the steady-state probabilities of particle partitioning; here we focus on equilibrium phenomena.)

Rather than directly testing those two hypotheses, we instead determined the extent of the driving forces for PSII redistribution in "high-light" membranes that are unaccounted for by the model developed in the previous section. Figures 4.1cd, 4.4, and 4.5 show the results of simulations that are analogous to those of Figures 4.1ab, 4.2, and 4.3, respectively, but with f_G, f_M, and f_S selected to match Kirchhoff's "high-light" scenario.

Surprisingly, we find excellent agreement between simulation and experiment for the partitioning of PSII between thylakoid subsystems without any further adjustments to the model. This agreement suggests that the driving force for PSII transport out of the grana during repair is primarily entropic, and is equivalent to a reduction in the relative size of the stacked region of the membrane ($f_G = 0.60$ in "control" vs $f_G = 0.48$ in "high light"). This then shifts the question to, what is the driving force for shrinkage of the grana upon exposure to high light? PSII core phosphorylation may play an energetic role in this process; a simulation ensemble in which the relative areas of the subsystems can dynamically fluctuate would be an appropriate framework for investigating this hypothesis.

All other trends in particle partitionings in Figures 4.4 and 4.5 are analogous to those



(a) (b)

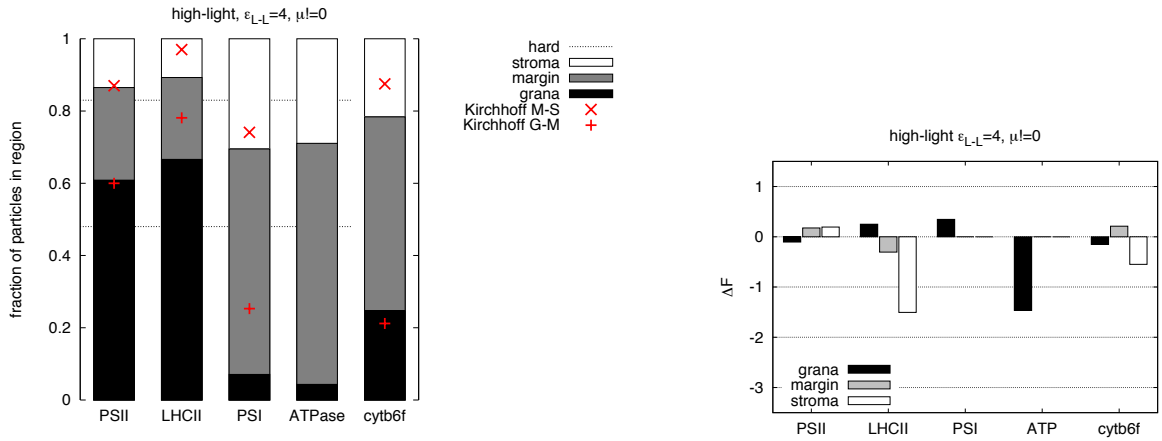
Figure 4.4: Comparison of particle partitioning in "high-light" thylakoids between simulation and experiment, with $\mu_{\alpha,i} = 0$. Rest of key as in Figure 4.2.

in Figures 4.2 and 4.3. The only notable discrepancy is in the extent of LHCII partitioning to stroma lamellae, but we hesitate to speculate on this because the experimental value of $m_{\mathrm{LHCII},S}$ is surprisingly small.

4.3 Extent of lateral heterogeneity depends on LHCII stacking

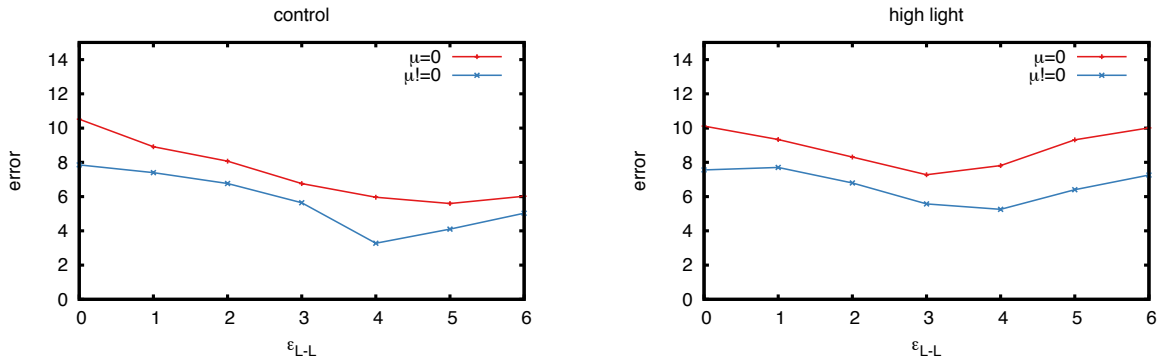
To further investigate the role of the LHCII stacking interaction in our model of thylakoid membranes, we repeated the above simulations with values of the interaction strength $\epsilon_{\mathrm{L\text{-}L}}$ between 0 and 6 $k_{\mathrm{B}}T$. We evaluated the goodness of fit to experiment by constructing an error function as the sum of the excess effective potentials as shown in Figures 4.2–4.5, i.e., error $= \sum_i \sum_\alpha |\Delta F_{\alpha,i}|$. For $\mu_{\alpha,i} = 0$, this error was minimized at $\epsilon_{\mathrm{L\text{-}L}} = 5\ k_{\mathrm{B}}T$ in "control" simulations and at $\epsilon_{\mathrm{L\text{-}L}} = 3\ k_{\mathrm{B}}T$ in "high-light" simulations (red lines in Figure 4.6). The locations of these minima further validate $\epsilon_{\mathrm{L\text{-}L}} = 4\ k_{\mathrm{B}}T$ as a reasonable approximate strength for the phenomenological LHCII stacking interaction in vivo, while also suggesting that deviations from this value are likely to occur in different experimental conditions. When nonzero external potentials are imposed (blue lines), $\epsilon_{\mathrm{L\text{-}L}} = 4\ k_{\mathrm{B}}T$ minimizes the error for both "control" and "high-light" scenarios. This shift in the minima is not surprising because the magnitudes of the external potentials were selected based on the $\epsilon_{\mathrm{L\text{-}L}} = 4\ k_{\mathrm{B}}T$ simulations.

Thus far, this Chapter has focused on composition gradients between regions of the thylakoid membrane. A gradient in the total packing fraction is also expected between



(a) (b)

Figure 4.5: Comparison of particle partitioning in "high-light" thylakoids between simulation and experiment, with $\mu_{\mathrm{PSI},G} = 0.5$, $\mu_{\mathrm{ATPase},G} = 3$, and $\mu_{\mathrm{cytb6f},G} = -0.5\ k_{\mathrm{B}}T$, and with all other $\mu_{\alpha,i} = 0$. Rest of key as in Figure 4.2.



(a) (b)

Figure 4.6: Error $= \sum_i \sum_\alpha |\Delta F_{\alpha,i}|$ between simulation and experiment as a function of LHCII stacking for (a) "control" and (b) "high-light" scenarios. Red: $\mu_{\alpha,i} = 0$ for all α, i. Blue: $\mu_{\mathrm{PSI},G} = 0.5$, $\mu_{\mathrm{ATPase},G} = 3$, and $\mu_{\mathrm{cytb6f},G} = -0.5\ k_{\mathrm{B}}T$, and all other $\mu_{\alpha,i} = 0$.

the grana and stroma lamellae regions: typical packing fractions for the grana are in the range ≈ 0.6–0.8 [133, 107, 70, 110], while the stroma is expected to be much less densely packed with proteins [3]. Figure 4.7 shows the packing fractions in each simulation subsystem in simulations under a range of Hamiltonians. Surprisingly, we find very little difference in packing fractions between the stacked and unstacked regions of the simulation box; all average packing fractions are within ± 0.03 of the overall packing fraction of 0.705. Even more surprisingly, we find that the packing fraction of the grana is not monotonic as a function of $\epsilon_{\text{L-L}}$, and is in fact less dense that the margin and stroma regions for moderate values of $\epsilon_{\text{L-L}}$ in the presence of the external potentials (right panels of Figure 4.7). Although the total packing fraction of the grana is nonmonotonic, the packing fractions of each particle species are monotonic and show the expected trends: the concentrations of LHCII and PSII increase in the grana as $\epsilon_{\text{L-L}}$ increases, while the concentrations of the other components decrease.

This dramatic discrepancy between simulation and experiment could be due to several reasons. One possibility is that conventional wisdom is incorrect, and there is in fact very little difference in protein packing fraction between stacked and unstacked membranes in vivo. Another possibility is that the simulations were conducted with parameters that are not representative of the membrane in vivo. Despite the care we took in deriving the f_i and N_α parameters that define these canonical ensemble simulations from experimental measurements, and despite the good agreement between simulation and experiment in the rest of this dissertation for $\epsilon_{\text{L-L}} = 4\ k_{\text{B}}T$, we hesitate to make a strong prediction that the protein density in unstacked membranes is higher than previously thought. Further work is needed to resolve this discrepancy.

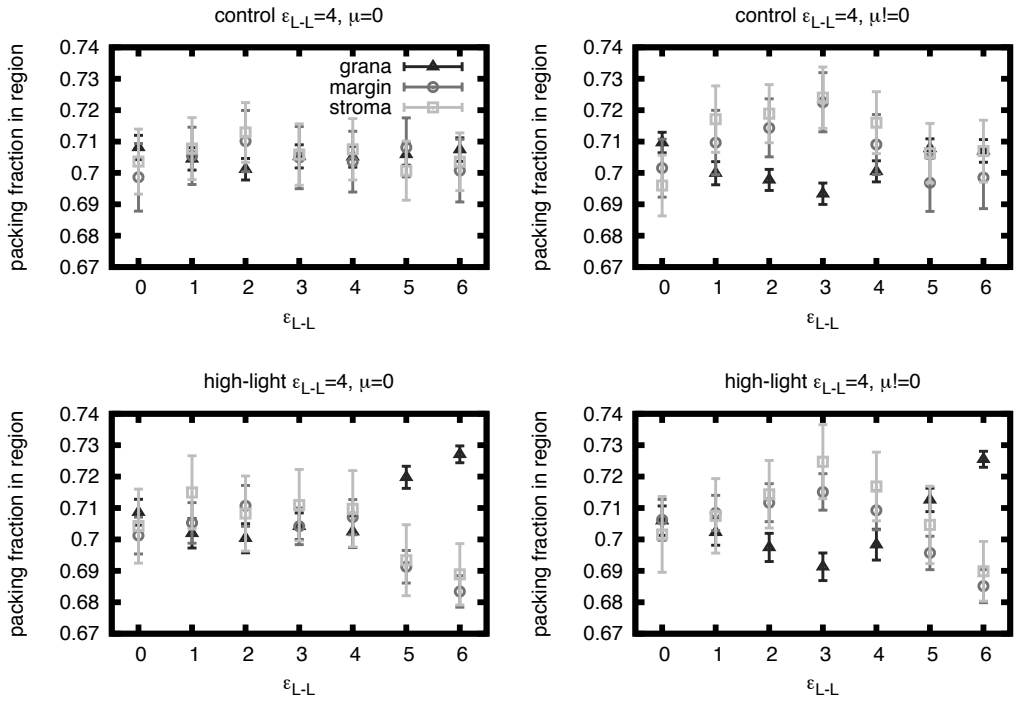


Figure 4.7: Dependence of protein particle packing fraction in regions of simulated thylakoids on LHCII stacking. Top and bottom are "control" and "high-light" scenario simulations, respectively; left and right are $\mu_{\alpha,i} = 0$ for all α, i and ($\mu_{\mathrm{PSI},G} = 0.5$, $\mu_{\mathrm{ATPase},G} = 3$, $\mu_{\mathrm{cytb6f},G} = -0.5\ k_{\mathrm{B}}T$), respectively. Black triangles, medium gray circles, and light gray squares are for grana, margin, and stroma subsystems, respectively. Error bars are standard deviations.

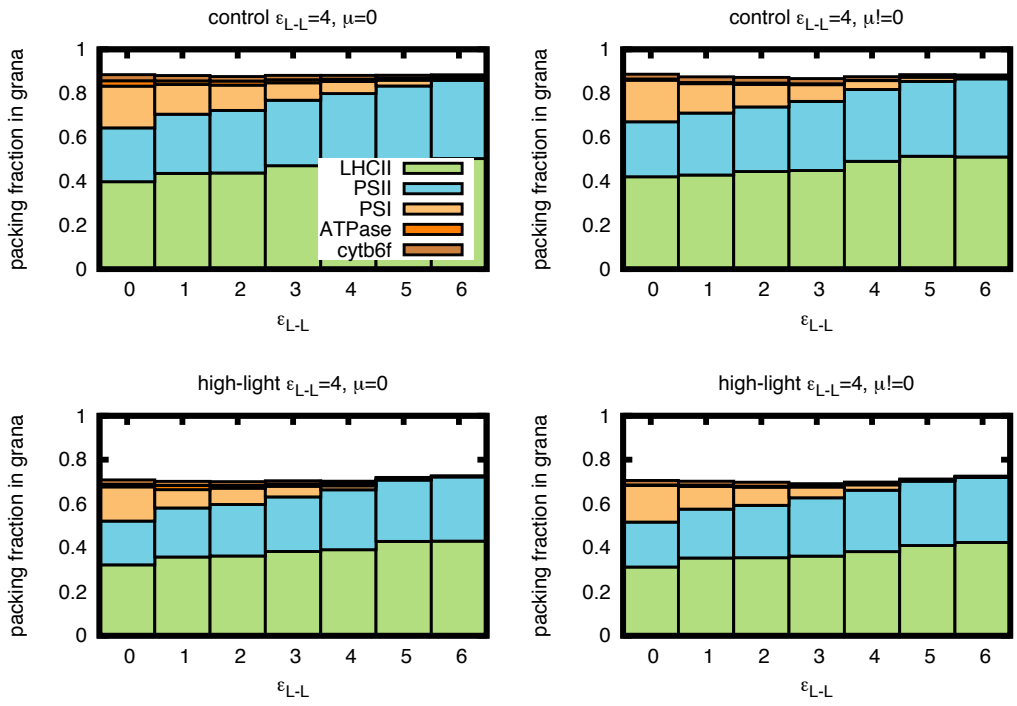


Figure 4.8: Dependence of protein particle localization to the grana region of simulated thylakoids on LHCII stacking. Top and bottom are "control" and "high-light" scenario simulations, respectively; left and right are $\mu_{\alpha,i} = 0$ for all α, i and ($\mu_{\mathrm{PSI},G} = 0.5$, $\mu_{\mathrm{ATPase},G} = 3$, $\mu_{\mathrm{cytb6f},G} = -0.5\ k_{\mathrm{B}}T$), respectively. Average packing fractions due to LHCII, PSII, PSI, ATP synthase, and cyt b_6f particles are green, blue, tan, orange, and brown, respectively, to match the particle colors in Figure 4.1.

4.4 LHCII aggregation can be induced by a square-well attraction

As a final vignette, we return to the grana and consider the question of LHCII aggregation. There is evidence that LHCII complexes colocalize or "aggregate" during qE (reviewed in Chapter 1 and, e.g., [149]). In particular, Johnson et al. [91] found that LHCII and PSII clustering in the grana of intact *Arabidopsis* thylakoids is correlated to the degree of NPQ, and is related to the presence of different xanthophyll pigments zeaxanthin and violaxanthin, which bind to the periphery of LHCII [120]. The combination of energetic and entropic forces that drive this behavior are not known, nor is it known whether it indicates an underlying LHCII fluid-crystal phase transition that could amend the PSII–LHCII phase diagram presented in Chapter 3 [154]; for convenience, I will use the term aggregation nevertheless. Many of the proposed physicochemical processes associated with qE could favor LHCII aggregation: increased protonation or conversion of bound violaxanthin to zeaxanthin [149] could, either directly or indirectly (e.g., by inducing a conformational change of LHCII), increase energetic attractions among LHCII complexes to favor LHCII clustering, or could decrease energetic attractions between LHCII and PSII complexes to favor LHCII phase separation from PSII. On the other hand, a qE-associated conformational change of LHCII that decreases its volume [178] would decrease the entropic driving force for aggregation.

To investigate the effect of a non-specific energetic driving force for LHCII aggregation, we extended the model in Chapter 2 to include an in-plane square well interaction between LHCII particles

$$u_{\mathrm{sw,L-L}}(r) = \begin{cases} 0 & r > \lambda\sigma_{\mathrm{L}} \\ -\epsilon_{\mathrm{sw}} & \lambda\sigma_{\mathrm{L}} \geq r > \sigma_{\mathrm{L}} \\ \infty & r \leq \sigma_{\mathrm{L}} \end{cases} \tag{4.6}$$

where r is the radial distance between the centers of LHCII particles in the same membrane plane. Simulations with interaction strengths $\epsilon_{\mathrm{sw}} = 0.5, 1.0\ k_{\mathrm{B}}T$ were compared to simulations with this interaction turned off ($\epsilon_{\mathrm{sw}} = 0.0$). We used an interaction range of $\lambda = 1.15$, corresponding to an edge-to-edge distance of 0.975 nm. This interaction range is comparable to that of the square well PSII–LHCII interaction in Chapter 2, and is in a region of parameter space where a pure 2d square well system has neither a solid-solid transition [29] nor a fluid-fluid transition [141]. Canonical ensemble simulations were performed at $\phi = 2.7, 6$ (to match [30] and [176], respectively); with $\rho = 0.70, 0.75, 0.80$; and with all other parameters as in Chapter 2 ($\epsilon_{\mathrm{L-L}} = 4, \epsilon_{\mathrm{SL-P}} = \epsilon_{ML-P} = 2$). A representative particle configuration from these simulations is shown in Figure 4.9.

Unsurprisingly, the addition of a generic isotropic attraction between LHCII particles leads to LHCII aggregation and ordering. Figure 4.10 shows two standard measures of particle ordering in 2d, the radial distribution function $g(r)$ (left) and the probability distribution function of the local hexatic bond-orientational order parameter $\psi_{6,i} = 1/N_i \sum_j \exp\left[6\mathrm{i}\theta_{ij}\right]$ where θ_{ij} is the angular polar coordinate of the vector from particle i to particle j [135] (middle). As expected from Chapter 3, liquid-like structure is present in both distributions

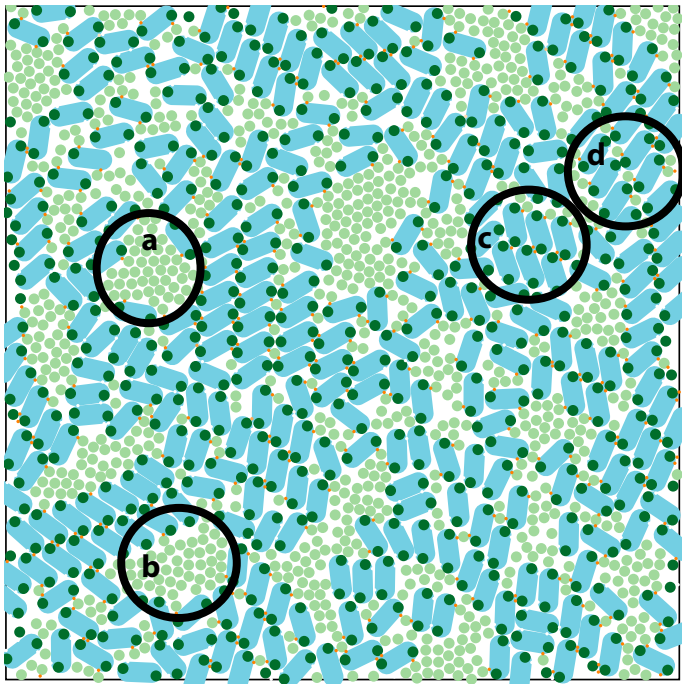


Figure 4.9: Particle configuration taken from the top (lumen-side-up) layer of a simulation with $\epsilon_{sw} = 1.0, \rho = 0.75, \phi = 2.7$ in a 400-nm square box. (a) highlights a region of ordered LHCII; (b) highlights a region of disordered LHCII; (c) highlights a C_2S_2 PSII crystal; (d) highlights a C_2S_2M PSII crystal.

at $(\phi = 6.0, \epsilon_{sw} = 0, \rho = 0.70)$ (top): $g(r)$ is weakly structured, and the distribution of ψ_6 has a large peak near 0 and a shoulder near 1. Increasing the total protein packing fraction to $\rho = 0.75$ increases the structure of the LHCII fluid slightly, due to entropic driving forces. Increasing the LHCII square-well interaction strength to $\epsilon_{sw} = 0.5$ has a similar structuring effect, but due instead to energetic driving forces. Further increases to $\rho = 0.80$ or $\epsilon_{sw} = 1.0$ lead to dramatic increases in structure, consistent with a fluid-crystal transition. Although the precise location of the LHCII fluid-crystal transition in the context of this complex fluid would be difficult to predict a priori (and indeed I make no attempt to do so), all of these qualitative trends are exactly in line with what one would expect.

An interaction strength of $\epsilon_{sw} = 1.0$ between LHCII particles also increases the ordering of PSII particles (Figures 4.10, right). In-plane energetic attractions between LHCII particles increase the line tension between LHCII-rich and PSII-rich domains, lowering the nucleation barrier to formation and/or the thermodynamic stability of PSII crystals within the PSII-rich domains. These trends are present but weaker at $\phi = 2.7$ (Figure 4.11). This may be because the effective LHCII density is lower; e.g., when $(\rho = 0.75, \phi = 2.7)$, LHCII particles are present at an effective packing fraction of 0.42 in the space unoccupied by PSII particles (without accounting for the inaccessible space due to packing effects), but at an effective packing fraction of 0.55 when $(\rho = 0.75, \phi = 6.0)$. Anecdotal evidence suggests that the

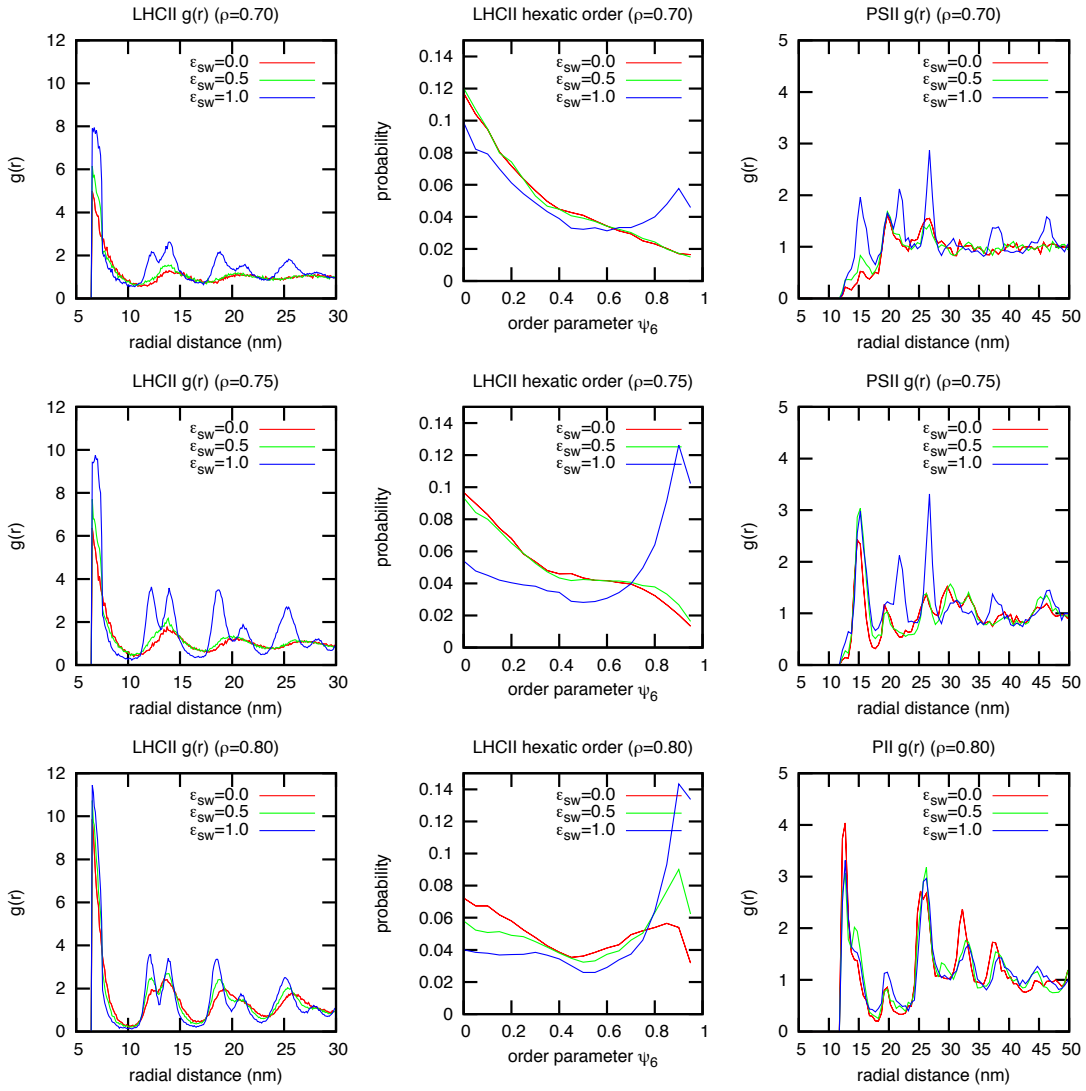


Figure 4.10: LHCII and PSII ordering induced by a square-well attraction at $\phi = 6.0$. Top, middle, and bottom rows have $\rho = 0.70, 0.75, 0.80$, respectively. Left, middle, and right columns plot LHCII $g(r)$, LHCII ϕ_6, and PSII $g(r)$, respectively.

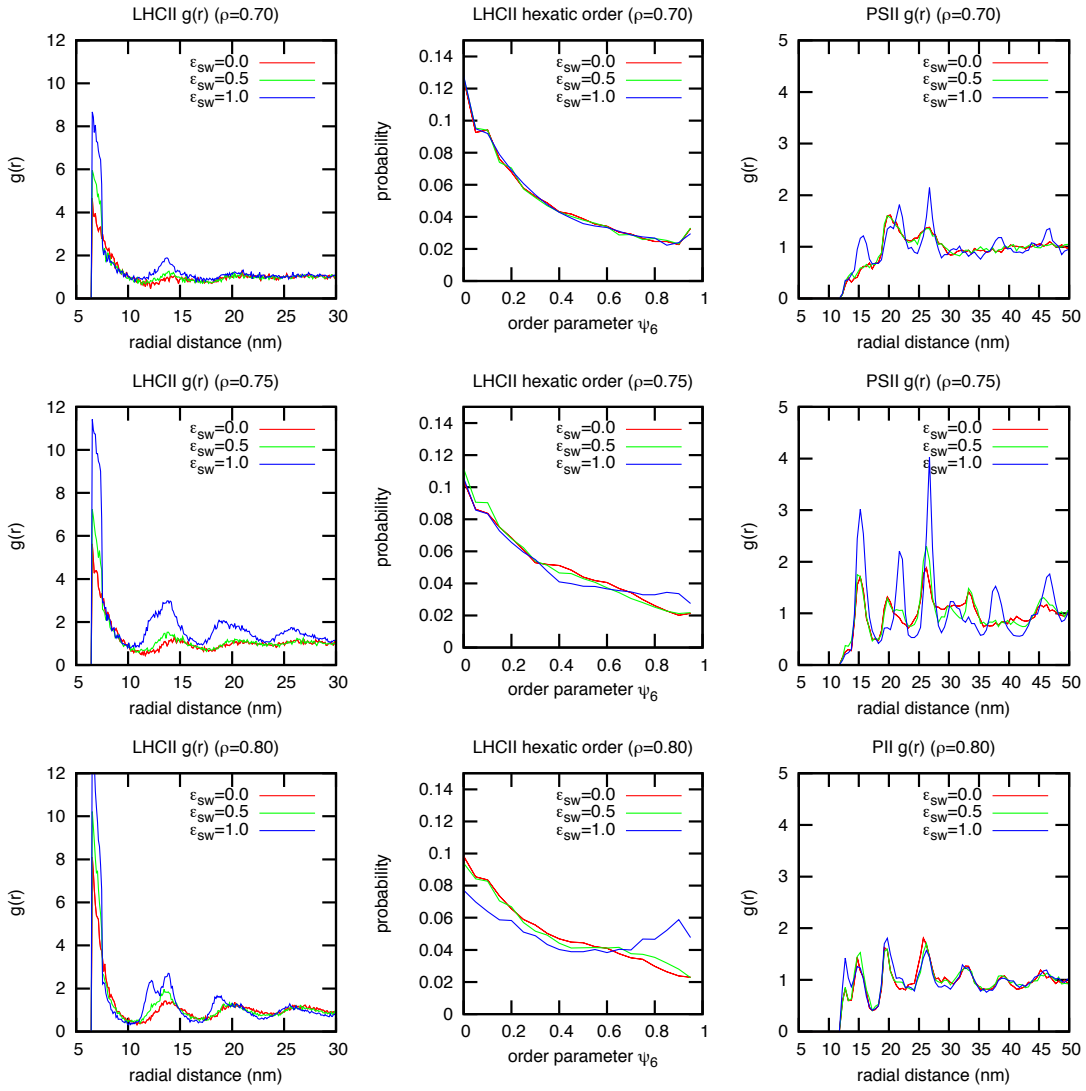


Figure 4.11: LHCII and PSII ordering induced by a square-well attraction at $\phi = 2.7$. Top, middle, and bottom rows have $\rho = 0.70, 0.75, 0.80$, respectively. Left, middle, and right columns plot LHCII $g(r)$, LHCII ϕ_6, and PSII $g(r)$, respectively.

LHCII interaction favors a mixture of C_2S_2 and C_2S_2M crystals (e.g., Figure 4.9).

How do these ordering phenomena compare to the LHCII aggregation reported in experiments? Johnson et al. use a nonstandard measure of particle clustering, namely the number of particles of the same species α within a given radial cutoff distance $r_{c,\alpha}$ [91]. Specifically, they select $r_{c,\mathrm{L}} = 25\mathrm{nm} \approx 3.8\sigma_{\mathrm{L}}$ and $r_{c,\mathrm{P}} = 50\mathrm{nm} \approx 4.2\mathrm{D_L} \approx 3.4\mathrm{L_P}$. They find that the distributions of these quantities are visually well fit by Gaussians, with means and variances that are positively correlated to the amount of NPQ (i.e., more NPQ means more clustering). We calculated the same quantities from the simulated data; Figure 4.12 shows the means of these distributions along with the means of the experimental distributions [91].

The experimental and simulated data in Figure 4.12 do not agree well. The simulated data with $\phi = 2.7$ spans roughly the same range of average PSII clustering as the experimental data but has greater LHCII clustering, while the simulated data with $\phi = 6.0$ spans roughly the same range of LHCII clustering but has greater PSII clustering. Because the experimental data lie on a diagonal between the two sets of simulated data, it is likely that the experimental data could be better fit by simulated membranes at an intermediate ϕ between 2.7 and 6.0, which would be within the range of other experiments and would be unsurprising. Yet a rough interpolation suggests that adjusting ϕ alone would result in simulated data with significantly more clustering of both LHCII and PSII than seen in experiment; a drastic reduction of total density to $\rho \ll 0.70$ would be needed to bring simulations into quantitative agreement with experiment. Such a low packing fraction would be outside the range of other experiments and would be quite surprising.

Because the disagreement between experiment and simulation is larger than can be accounted for by fitting via adjustments to the coarse-grained LHCII aggregation potential alone, we hesitate to draw conclusions about the realism of this model for LHCII aggregation or to make further predictions based on it. Accurate simultaneous measurements of ϕ, ρ, and any clustering order parameter as a function of NPQ could better constrain future modeling and simulation efforts, leading to a better understanding of the structural and thermodynamic basis for LHCII aggregation during NPQ.

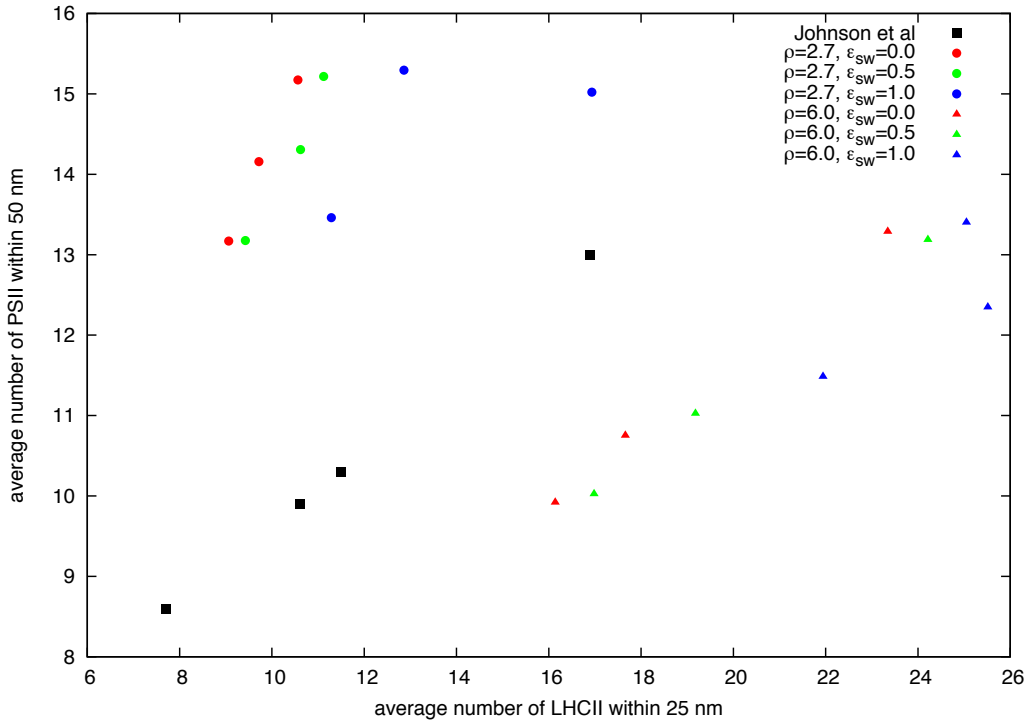


Figure 4.12: Average number of nearby particles in simulation and experiment. "Johnson et al." data taken from Table 2 of [91].

Chapter 5

A statistical learning pipeline can classify local order of PSII in atomic force micrographs of grana membranes

In this Chapter, we develop and apply a statistical learning pipeline for the analysis of PSII organization in grana membranes at fluid-crystal phase coexistence (Chapter 3). By fitting a Gaussian mixture model to our training data set of PSII crystalline unit cells (comprised of data from the literature and of Bravais lattices fit to nearest-neighbor shells of particles in AFM images), we obtain two important results: an intuitive yet statistically grounded taxonomy of PSII unit cell classes and a tool for classifying any unit cell into one of these classes. We apply this classifier to a larger data set of particle positions extracted from AFM images of isolated grana from wild-type and *soq1* *Arabidopsis* plants, and find that the *soq1* mutation has a structural phenotype that may support its functional phenotype in NPQ [31]. This work was a close collaboration between myself and Bibiana Onoa with help from Matt Brooks and Patricia Grob, and of course the guidance of our advisers (Phill Geissler, Carlos Bustamante, Kris Niyogi, and Eva Nogales, respectively).

5.1 Background

Because many grana membranes may be at crystal-fluid phase coexistence in vivo (Chapter 3, [154]), a common step in the analysis of nano-resolution EM and AFM images is separating particles with crystalline local environments (henceforth "crystalline particles") from those with disordered local environments (henceforth "disordered particles"). Yet previous studies have not published an algorithmic basis for distinguishing between crystalline and disordered PSII complexes [150, 190, 112, 46, 66, 96, 43, 102, 110]. Robust algorithmic identification of spatial motifs like crystalline unit cells is a ubiquitous task in biophysics and nanoscience,

yet different fields tend to take different approaches. In the electron microscopy (EM) literature, local motifs are typically identified by cross-correlation of pixel intensity and class averaging [72], however this approach does not perform well on atomic force microscopy (AFM) micrographs. In contrast, local motifs in configurations taken from molecular simulations are often characterized using shape-matching methods on Fourier or Zernike rotation-invariant feature sets [98, 99], which can take point particles as input but require inherently high-dimensional data sets to classify motifs that are not well described by a simple n-fold symmetry.

In this Chapter, we develop an analysis pipeline for detecting and classifying local patterns in configurations of 2d coordinates using an easy-to-understand feature set: the (a, b, θ) parameters of a unit cell in a Bravais lattice. Because our method acts on the spatial coordinates of a set of particles, rather than on image data, it can be applied to coordinates extracted from images taken with EM, AFM, or any other imaging modality, or from computer simulations. This generality puts disparate experimental techniques on equal footing, promoting straightforward aggregation of and comparison between large data sets. We applied this method to a high-resolution AFM data set of *Arabidopsis* wild-type and *soq1* [31] grana membranes collected before and after exposure to excess light (see Appendix A for data collection methods), and found that it was robust enough to be consistent with previous reports of total crystallinity in wild type, while powerful enough to quantitatively discriminate between coexisting PSII crystals. Through detailed analysis of the local organization of PSII supercomplexes in ordered and disordered phases, we found evidence that interactions among light-harvesting antenna complexes are weakened in the absence of SOQ1, inducing protein rearrangements that favor larger separations between PSII complexes in the majority (disordered) phase and reshaping the PSII crystallization landscape. Because the quenching that is suppressed by SOQ1 is thought to occur in antennae, the structural effects that we detect could play a role in its NPQ function. The features we observe are distinct from known protein rearrangements associated with NPQ, providing further support for a role of SOQ1 in a novel NPQ pathway.

To aid in future clustering efforts and to promote the use of standardized methods for analyzing PSII organization, we developed Picolo (Point-Intensity Classification of Local Order) [153]. Picolo is an open-source Python package that supports a complete pipeline for analysis of local spatial organization in sets of $x - y$ coordinates, particularly for sets of PSII coordinates picked from micrographs of grana membranes. Picolo includes all the algorithms used in this Chapter, as well as standard algorithms for rotation-invariant Fourier and Zernike features, support vector machine classifiers, and radial distribution functions.

5.2 Extraction of Bravais lattice features

After identifying PSII particle positions in grana discs (described in Appendix **??**), we characterized the local environment of each PSII particle by its unit cell, using an algorithm in Picolo that optimizes a Bravais lattice to the coordinates of a particle's nearest-neighbor

shell. Unit cells are convenient for our analysis because, having only three features (vector lengths a and b, and angle θ), they are inherently lower-dimensional than the typical feature sets used in shape-matching [98, 99]. Additionally, unit cells are commonly published in the thylakoid electron microscopy literature, allowing us to include many previously published unit cells in the clustering stage of the analysis.

Specifically, for particle i at position $\mathbf{r}_i$, a local unit cell $\Theta_i = (a_i, b_i, \theta_i)$ was extracted by the following algorithm:

1. Determine the set A_i of neighboring particles j around particle i, $A_i = \{\mathbf{r}_j \,|\, r_{\min} < r_{ij} < r_{\max}\}$, where the cutoffs $r_{\min} = 14$ nm and $r_{\max} = 33$ nm are chosen to select the first PSII coordination shell based on typical PSII $g(r)$ data. Let n_i be the number of particles in A_i.

2. Find the Bravais lattice B_i centered at $\mathbf{r}_i$ spanned by the real-space lattice vectors $(\mathbf{u}_i, \mathbf{v}_i)$, $B_i = \{\mathbf{r}_k \,|\, \mathbf{r}_k = n_{u_k}\mathbf{u}_i + n_{v_k}\mathbf{v}_i \;\forall\; n_u, n_v \in \mathbb{Z}\}$, that minimizes the penalty function $\pi(i)$

$$\pi(i) = \frac{1}{n_i} \sum_{\mathbf{r}_j \in A_i} \min_{\mathbf{r}_k \in B_i} \epsilon(r_{jk}) \tag{5.1}$$

   using the radial error function $\epsilon(r)$

$$\epsilon(r) = \begin{cases} r^2/r_B^2 & \text{if } r < r_B \\ 1 & \text{otherwise} \end{cases} \tag{5.2}$$

   and a cutoff distance $r_B = 7$ nm. COBYLA [146] was used to minimize $\pi(i)$ subject to the constraints $r_{\min} < |\mathbf{u}_i| < r_{\max}$ and $r_{\min} < |\mathbf{v}_i| < r_{\max}$.

3. If $n_i < 4$ or $\pi(i) > \pi_{\min} = 0.2$, then Θ_i does not exist. Otherwise, Θ_i exists; continue.

4. We can write any point $\mathbf{r}_k$ in B_i in polar coordinates, $\mathbf{r}_k = n_{u_k}\mathbf{u}_i + n_{v_k}\mathbf{v}_i = (r_k, \theta_k)$. For any pair of distinct points $\mathbf{r}_k, \mathbf{r}_l \in B_i$, define the unit cell $\Theta_{kl} = (a_{kl}, b_{kl}, \theta_{kl})$ by $a_{kl} = r_k$, $b_{kl} = r_l$, and $\theta_{kl} = \theta_l - \theta_k \in [0, 2\pi)$. Let C_i be the set of smallest unit cells, $C_i = \{\Theta_{kl} | n_{u_k}, n_{v_k}, n_{u_l}, n_{v_l} \in \{-1, 0, 1\}\}$. Then we select Θ_i via

$$\Theta_i = \arg\min_{\Theta_{kl} \in C_i} |\theta_{kl} - \theta^*| \tag{5.3}$$

   subject to the constraints $a_{kl} < b_{kl} + r_e$ and $\theta_{\min} < \theta_{kl} < \theta_{\max}$. The angular parameters $\theta_{\min} = 50°$, $\theta_{\max} = 120°$, and $\theta^* = 75°$ were chosen to favor unit cells in the range of the unit cells in [110]; $r_e = 1$ nm is a small error term on the scale of the pixel size that allows the angular constraints to be satisfied consistently when $a_{kl} \approx b_{kl}$.

This scheme robustly yields unit cells that agree with unit cells obtained from traditional Fourier transform methods on large ordered regions, but does not rely on long-range periodicity for its success. Our choices of error function and minimization algorithm for this fitting

procedure lead to fast, numerically stable code that accurately identifies unit cells when the coordinates of the neighboring particles are highly ordered. When positional noise is above an upper threshold, i.e., when the neighboring particles are not well fit by any Bravais lattice, the particle is assigned as disordered. This can lead to false positives (assigning disordered particles as crystalline) if the threshold on the error function is too high, or false negatives (assigning crystalline particles as disordered) if the threshold is too low. To be conservative, we have selected a relatively low threshold, which means that we are erring on the side of underestimating the total crystal content of the membranes.

5.3 Selecting a GMM clustering model

A key step in our analysis pipeline is selecting a taxonomy of unit cell classes into which the unit cells of individual particles are classified. The PSII crystalline array literature contains more than 100 unit cells, each comprising a single C_2S_2, C_2S_2M, or $C_2S_2M_2$ particle with a different placement and orientation [25, 110, 190, 43, 45, 130, 96, 102, 110]. Although qualitative distinctions have been drawn between sets of $C_2S_2M_2$ unit cells [48, 110] to our knowledge, no authors have previously presented an objective, quantitative method for separating the unit cells for each particle type into structurally distinct classes.

In the language of statistical learning [74], constructing this taxonomy is an unsupervised classification or clustering task, and we approached it using the Gaussian mixture model clustering method. A Gaussian mixture model (GMM) is a probabilistic model for a data set that is derived by fitting several multivariate Gaussian distributions to the data, typically using the expectation-maximization algorithm [50]. Each Gaussian component in the mixture is characterized by a vector of means μ that describes the center of the cluster, a covariance matrix σ^2 that describes the extent of the cluster along each dimension of the data, and a scalar mixture weight that describes the relative probability of that cluster within the data set; the only other fitting parameter is the number of components k. In contrast to other popular clustering methods such as hierarchical clustering, a GMM can be completely described using this small set of easy-to-interpret parameters. Moreover, a GMM is readily amenable to conversion to a probabilistic classifier [168], allowing us to transition from the original unsupervised classification problem to the target supervised classification problem. Ultimately, fitting a GMM to our training data set yielded two important results: an intuitive yet statistically grounded taxonomy of PSII unit cell classes, and a tool for classifying any unit cell into one of these classes.

Specifically, we constructed a training data set consisting of 3 features (a_i, b_i, θ_i) for each of 101 unit cells from published PSII crystalline arrays [25, 110, 190, 43, 45, 96, 109] and 303 unit cells extracted from particles in 31 high-resolution (150 nm $\times$ 150 nm) AFM images (Appendix A). The scikit-learn Python package [145] was used to fit diagonal Gaussian mixture models to this data set.

The GMM method alone does not yield information about which value of k produces the most meaningful clustering; additional model selection criteria are necessary if k is

not known a priori. To select the number k of components in the Gaussian mixture, 1000 bootstrap-resampled data sets were generated and the best value of k in the range 1–10 was selected for each data set via the Bayes Information Criterion (BIC) [74]; Figure 5.1 shows that $k = 6$ was most frequently selected as the best number of clusters. Yet overall, $k = 6$ was selected in only 27% of the bootstrap resamplings, while $k = 5$, 7, and 10 were each selected in 15–20% of resamplings. While this model selection process guided us to present results with $k = 6$ in this work, it also suggests that one should use caution when drawing conclusions about the "ground truth" number and identity of classes of PSII unit cells in this data set. For all three features in the $k = 6$ model, the mean μ and standard deviation σ of each Gaussian mixture component of the non-resampled model fell within the bootstrapped 95% confidence interval (Table 5.1).

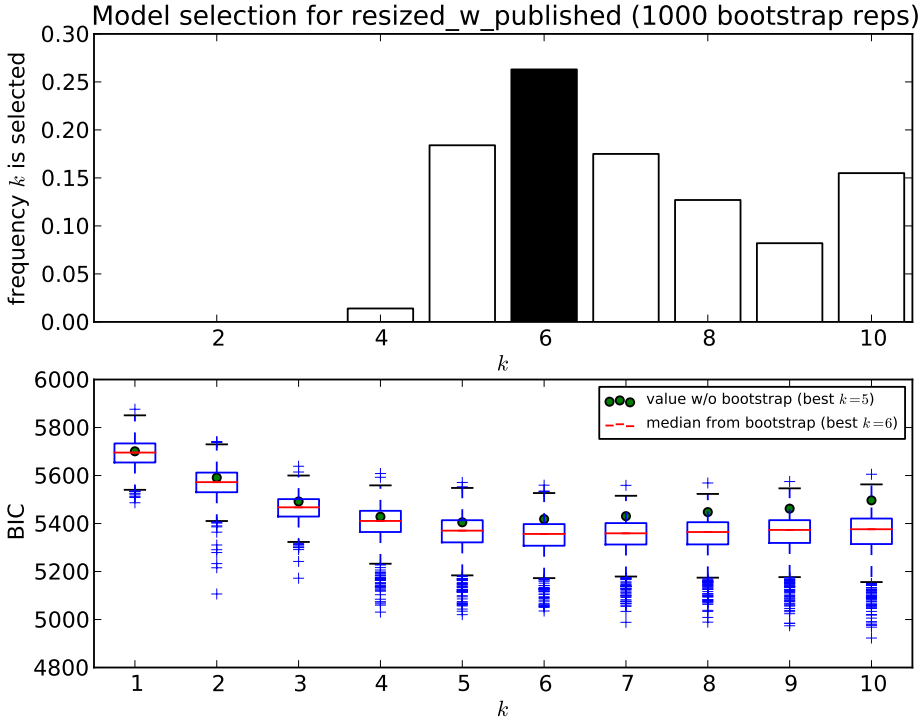


Figure 5.1: Bootstrap and BIC analysis for GMM model selection. Top: Frequency with which each value of k had the minimum value of BIC out of 1000 bootstrap resamplings. Bottom: BIC values for each value of k, with (red and blue boxplot) and without (green) bootstrap resampling.

Figure 5.2 illustrates the six classes of unit cells that arose from cluster analysis of the training data set. Classes (a) through (f) were ordered by decreasing mixture weight, such that class (a) had the most members (98 particles) and class (f) had the fewest (33 particles). Class (a) is prevalent in the $C_2S_2M_2$ data set of [110], allowing us to assign it as being

Table 5.1: Estimation of GMM parameters μ and σ^2 for the $k = 6$ model in Figure 5.2. The first line of each pair of rows is the fit to the training data set, and the second is the mean and 95% confidence interval of fits to 1000 bootstrap resampled data sets.

parameter	(a)	(b)	(c)	(d)	(e)	(f)
mixture weight	0.29	0.24	0.18	0.11	0.12	0.05
μ_a (nm)	20.8	21.9	20.3	18.9	19.7	22.0
μ_a CI (nm)	20.9 ± 0.7	21.7 ± 0.8	20.3 ± 1.5	18.9 ± 0.4	19.7 ± 0.4	21.8 ± 0.9
μ_b (nm)	25.6	23.1	24.0	23.0	28.2	26.0
μ_b CI (nm)	25.5 ± 0.6	23.2 ± 0.8	24.4 ± 2.1	22.7 ± 1.1	28.1 ± 0.6	26.1 ± 1.3
μ_θ (degrees)	71.6	74.9	76.5	83.2	68.9	63.4
μ_θ CI (degrees)	71.6 ± 1.9	74.6 ± 3.5	75.6 ± 11.2	83.9 ± 2.5	68.9 ± 2.3	63.5 ± 8.1
σ_a (nm)	1.27	0.74	0.60	1.01	0.69	1.18
σ_a CI (nm)	0.96 ± 0.49	0.81 ± 0.35	0.82 ± 0.45	0.95 ± 0.33	0.69 ± 0.21	1.10 ± 0.54
σ_b (nm)	1.01	0.78	0.93	1.74	0.41	1.53
σ_b CI (nm)	1.00 ± 0.37	0.77 ± 0.25	1.00 ± 0.49	1.53 ± 0.52	0.40 ± 0.35	1.35 ± 0.54
σ_θ (degrees)	2.50	3.03	3.21	3.85	1.86	5.86
σ_θ CI (degrees)	2.34 ± 1.25	2.60 ± 1.27	2.91 ± 1.60	3.33 ± 1.72	1.88 ± 0.67	5.86 ± 3.09

composed of $C_2S_2M_2$ particles. Classes (b) and (c) had slightly smaller b values than class (a) and thus smaller average areas; these classes may be composed of $C_2S_2M_2$ and/or slightly smaller supercomplexes (e.g., C_2S_2M). Class (d) contained the C_2S_2 unit cells [25, 45, 130], had the smallest average area, and was reminiscent of other rectangular C_2S_2 unit cells [112, 46]. Class (e) was distinguished by a larger b value, which manifests as a larger spacing between rows and the largest average area, and was almost entirely composed of $C_2S_2M_2$ unit cells like those in Kouřil et al.'s "normal light" and "low light" conditions. Class (f) contained the $C_2S_2M_2$ unit cells from [110], as well as some outliers with unusually acute angles ($\theta < 65°$) that may be artifacts of the unit cell determination algorithm.

5.4 Detection and classification of local order in grana membranes

For classification, a two-step pipeline was used. First, a probabilistic classifier was constructed from the $k = 6$ Gaussian mixture model using the maximum likelihood decision rule; a rejection cutoff on the class probability of $p_c = 0.005$ was used to classify outliers into an additional "disordered" class (also occupied by particles for which no unit cell was found). (Multiclass linear support vector machines were also investigated as possible classifiers, but this approach was rejected because outlier detection was not possible without moving to a less human-interpretable class boundary using a kernel such as the radial basis function.)

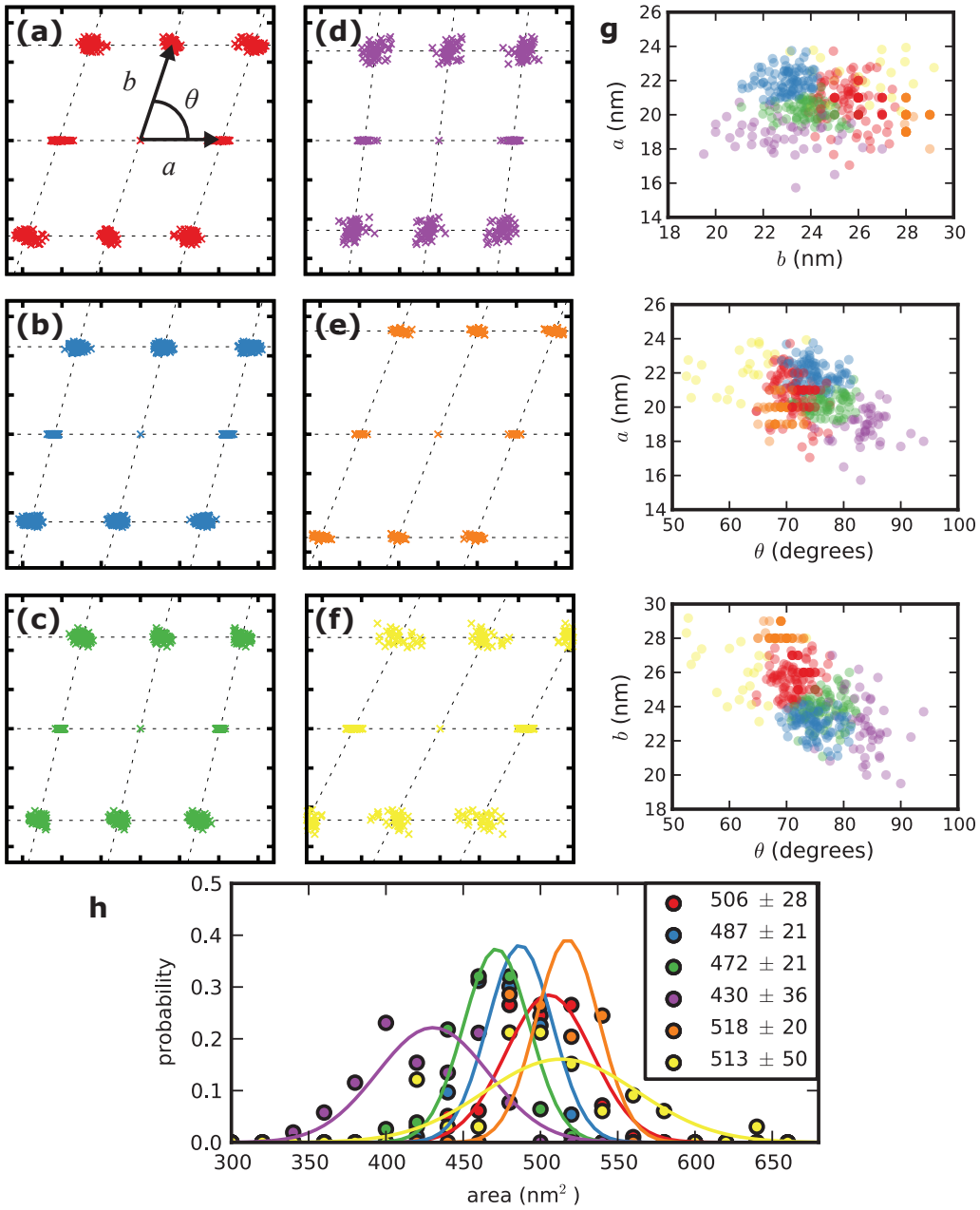


Figure 5.2: Results of cluster analysis of unit cells in the training data set. (a)-(f): Fitted Bravais lattices for each particle in the training set, grouped by Gaussian mixture model (GMM) component. The a vector of each unit cell lattice is aligned with the horizontal axis. Tick marks are spaced 10 nm apart. Panel (a) illustrates the unit cell conventions used throughout this work. g: Two-dimensional views of the three-dimensional (a, b, θ) data points in the training set, colored by GMM component. h: Histograms of unit cell area (area $= ab \sin\theta$), grouped by GMM component. Values in legend are means $\pm$ s.d. in square nanometers; lines are normal distributions with these means and standard deviations.

Second, a spatial k-nearest neighbor majority-rule filter, where k is the number of particles within a radius r_{max} of particle i, was used on the categorical output of the classifier to reduce the noise.

We tested this newly developed methodology on larger, lower-resolution AFM micrographs (500 nm × 500 nm). Figure 5.3 shows particle classifications for several representative AFM images. Examination of the classification results reveals a mixture of contiguous disordered regions, contiguous regions of a single crystal type, and regions with small pockets of several crystal types. By eye, the classification results had a very low false positive rate (i.e., few particles that appear to have disordered local environments were assigned to any crystal class) but a higher false negative rate (i.e., the algorithm did not give a crystal label to all particles that a human might assign as crystalline). In the absence of "true" labels, a more thorough analysis of prediction accuracy or false negative rate is not possible.

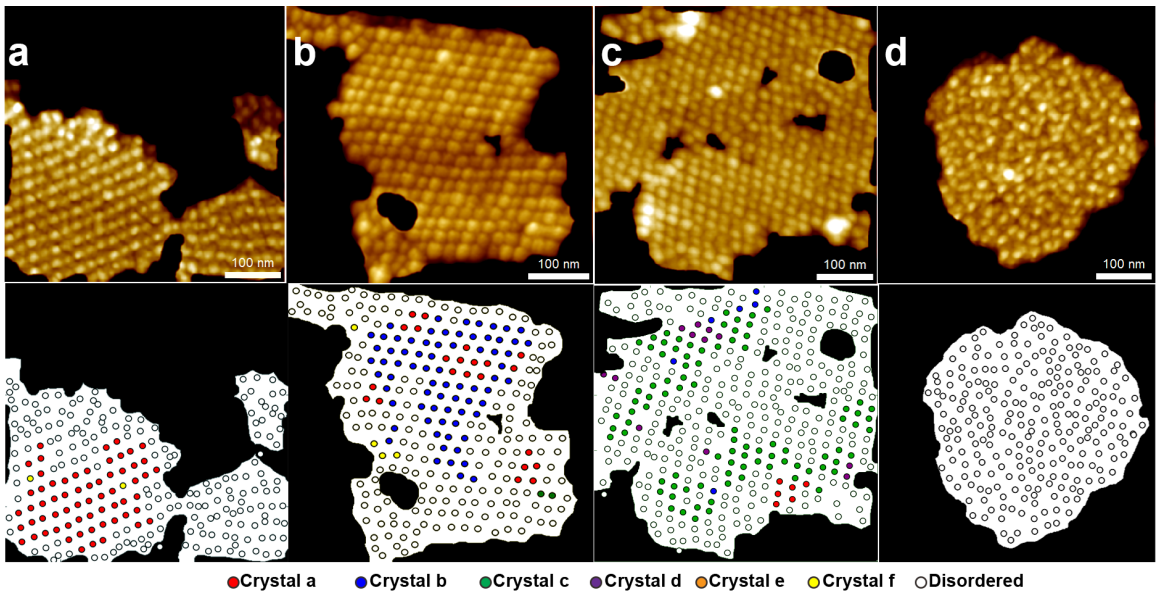


Figure 5.3: Classification of crystalline particles in low-resolution AFM images. Representative masked AFM micrographs (top) and their respective reconstructed images showing results of particle classification (bottom) for membranes enriched with crystal classes (a), (b), and (c), respectively, and crystal-free grana disc (panel d).

While several previous studies have reported the average net crystallinity of grana membranes, either by particle or by area, these authors do not report their crystal detection methods, nor do they provide crystallinity distribution functions that could be compared to predictions from phase diagrams like the one presented in Chapter 3 [154]. To measure the net crystallinity in each low-resolution AFM image, we divided the number of particles assigned to any crystal class by the total number of particles in the image. Histograms of this quantity (Figure 5.4a) show that PSII membranes from plants acclimated to low light have a higher crystal content than membranes from high-light-treated membranes. The effect

of high light was more drastic for wild-type plants than for *soq1* plants: high light treatment decreased the average net crystallinity from 9.2% to 3.5% in wild-type membranes, but only from 8.1% to 5.3% in *soq1* membranes. It is important to note, however, that the average crystallinity does not capture the full probability distribution: at least one membrane patch with >25% crystallinity was observed in every experimental condition, while the majority of images have <5% crystallinity (Figure 5.4a).

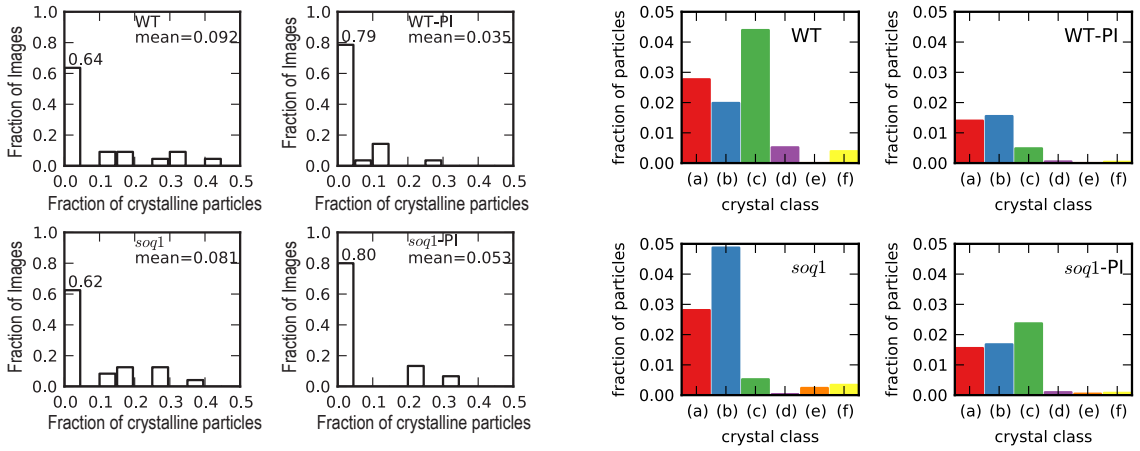
In control light conditions, our results indicate that, regardless of the genetic background, a small amount of crystalline arrays is present (8–10% of total particles). Under comparable experimental conditions, our result for WT membranes is consistent with a recent EM study [110]. Specifically, our observed average crystallinity by area is similar to their fraction of micrographs with any crystal. Although we detect crystals in more micrographs than [110], many of our crystals are small, and the big crystals dominate the average. Also, the average crystal's unit cells reported by EM are slightly larger than those presented here. This could be due to the presence of crystals in the EM study that are absent from our samples, or could be due to differences in experimental and data analysis methods.

We found that even relatively short exposure (90 minutes) to high light triggers crystal dissolution in both WT and *soq1* membranes. Our results are in line with observations reported for fully high-light-acclimated membranes, in which the crystalline regions were severely reduced [110]. Crystallinity in wild-type membranes appears to be more sensitive to photoinhibition than mutant membranes; average net crystallinity decreased by almost a factor of 3 in the former, and only decreased by a factor of 1.5 in the latter.

The frequency of different crystal types was also found to depend on the light treatment and the *soq1* mutation. Figure 5.4b shows the distribution of particles between crystal classes; the remaining particles were disordered. As expected from the clustering model, classes (a), (b), and (c) together accounted for the majority of crystalline particles in each experimental condition, with extremely few particles assigned to classes (d), (e), and (f).

5.5 Application: structural phenotype of the *soq1* mutation

The recently-characterized SOQ1 protein is implicated in a potentially novel NPQ pathway [31]. Because this pathway displays slow relaxation kinetics after high light exposure, Brooks et al. speculated that the *soq1* mutation may shift the position of grana membranes within the fluid-crystal coexistence region. Figure 5.4a does not support this hypothesis because there is not a dramatic difference in the overall crystallinity between wild-type and *soq1* membranes in this data set. Nevertheless, the differences in the distribution of particles between different crystal classes, particularly classes (b) and (c), in different experimental conditions (Figure 5.4b) are intriguing. In control light conditions, crystal (c) prevails in WT membranes and crystal (b) in *soq1*. Upon high-light treatment, crystal (c) is drastically reduced in WT-PI but is enhanced in *soq1*-PI. Crystal (b), on the other hand, is reduced in both WT and *soq1* membranes, the latter being more severely affected. Therefore, although SOQ1 does



(a) Histograms of net crystallinity. Numbers over the left-most bar in each histogram indicate the fraction of images with <5% crystalline particles. The mean fractions are weighted averages, weighted by the total number of particles in each image.

(b) Distribution of crystalline particles between crystal classes. Color scheme and class names are as in Figures 5.2 and 5.3. The y axis indicates the fraction of the total number of particles analyzed; the remaining fraction of particles is disordered.

Figure 5.4: Classification of crystallinity by particle in AFM images of WT (upper left), high-light-treated WT (WT-PI, upper right), *soq1* (lower left), and high-light-treated *soq1* (*soq1*-PI, lower right) membranes.

not have more than a minor effect on the stability of the PSII-LHCII fluid phase relative to any crystalline phase, it appears to reshape the phase diagram within the crystalline phase, including exerting a stabilizing effect on class (c) relative to class (b) in control light conditions and destabilizing class (c) after prolonged illumination.

It is difficult to know whether these differences are physiologically significant based on this structural analysis alone; atomic-resolution structures of each of the crystals would be necessary to determine the functional consequences of this modulation of the crystallization landscape. Based on the nanoscale AFM data, we propose internal structures for the dominant PSII-LHCII crystal classes (a), (b), and (c) as depicted in Figure 5.5. As expected, class (a) was very well fit by the model in Figure 3a of [110], which features end-to-end arrangements of CP26 (cyan) and of CP24 (green) subunits on adjacent supercomplexes, with separations of several nanometers within each pair of minor complexes (white arrows). The best-fit models for class (c) were similar to class (a), but with smaller separations between the peripheral antenna of adjacent supercomplexes. In contrast, the Bravais lattice of class (b) was best fit by a model with face-to-face arrangements of CP26 subunits, and with close end-to-end contacts between CP24 subunits. Rotations of the PSII axis by $\pm 6^\circ$ were possible within crystal (a) without creating steric clashes, while the smaller unit cells of crystals (b) and (c) led to smaller ranges of rotational freedom ($\pm 2^\circ$ and $\pm 0.5^\circ$, respectively). We note that each of these lattices could also be fit by tiling molecular models of smaller supercomplexes, e.g., C_2S_2M. If in vivo crystals lack supercomplex subunits, these models should be refined.

These possible structures suggest an intriguing hypothesis for the structural modulations that occur in the absence of SOQ1. Crystals (a) and (c) may be able to locally interconvert via slight rotations or translations while maintaining similar antenna contacts. On the other hand, the rotational restriction and distinct minor antenna contacts of crystal (b), which dominates in *soq1* control-light membranes, indicates that it may be stabilized by different energetic driving forces and that a cooperative transition would be necessary to convert between crystal (b) and the other crystals. Slight changes to the relative organization of chlorophyll dipoles can have dramatic effects on fluorescence and energy transfer efficiency [177, 5], so it is conceivable that the distinct CP26 and CP24 contacts in crystal class (b) could have a functional effect.

To investigate the possibility that absence of SOQ1 also affects inter-antenna interactions in the fluid phase, we measured the particle number density (Figure 5.6) and radial nearest-neighbor distance distribution function (Figure 5.7) within the fluid phase. Other EM and AFM studies of grana membranes have reported these metrics, but have not separated out the contributions from fluid and crystalline phases; this is a strength of our approach. In disordered regions, the density difference between WT and *soq1* membranes are very pronounced (Figure 5.6 center). Another significant difference was found between *soq1* membranes before and after photoinhibitory light treatment: the particle density in disordered regions significantly decreased upon light treatment, consistent with a net flow of PSII from the grana to the stroma lamellae during the PSII damage-and-repair cycle [66] and in agreement with the finding that *soq1* plants are not deficient in PSII repair [31]. In general, the crystalline particle densities were higher than the disordered densities, and were tightly clustered around

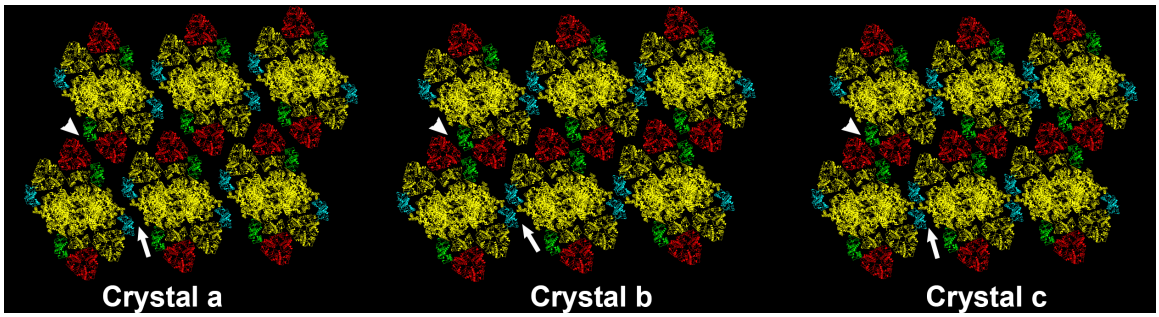


Figure 5.5: Proposed molecular models of crystal classes (a), (b), and (c). These structures were generated by fitting a molecular model of the PSII $C_2S_2M_2$ supercomplex [35] into the average unit cell of each class (Table 5.1), then refined by determining the range of PSII orientations that did not yield steric collision. Reaction center core dimer (C_2), S-LHCII trimers, and minor antenna CP29 are represented in yellow; M-LHCII, CP24, and CP26 antenna monomers are depicted in red, green, and cyan respectively. Inter-supercomplex M-CP24 interactions are indicated by white arrowheads, and CP26 interactions are indicated by white arrows.

2000–2100 particles/μm^{2}for all conditions (Figure 5.6 right). Note that 2050 particles/μm^2 corresponds to 488 nm^2/particle, approximately the size of a $C_2S_2M_2$ unit cell (Figure 5.2h), as expected.

While a unit cell completely describes the internal organization of a crystal, the particle density is insufficient to describe the internal characteristics of a disordered particle configuration. Therefore, we calculated the nearest-neighbor distribution function (NNDF) for particles in micrographs with no crystalline particles (Figure 5.7). The NNDF for WT was consistent with previously published distributions from Arabidopsis and spinach plants grown under similar conditions [23, 67, 107, 103, 110]. NNDFs from all conditions had a dominant intermediate-distance peak (≈18 nm), with flanking peaks centered at shorter (≈14 nm) and longer (≈22 nm) distances, similar to previous reports [23, 103]. The structure of the NNDF was similar for disordered particles in WT, WT-PI, and *soq1*-PI membranes. Interestingly, the short-distance shoulder was markedly larger in the *soq1* NNDF compared to the other conditions, and the long-distance shoulder was smaller.

These findings point to a role for SOQ1 in affecting PSII density in the grana by disrupting protein-protein interactions that favor smaller PSII separations. One candidate driving force that could be affected by SOQ1 in the absence of photoinhibitory light is attractive interactions between minor antenna complexes of neighboring PSII complexes. We favor this hypothesis because different contacts between minor antenna complexes are present in the crystals in *soq1* membranes (Figure 5.5), because SOQ1 is thought to affect an antenna-associated component of NPQ [31], and because arguments about entropic driving forces are not consistent with a higher PSII density in *soq1* grana. The reducing function of the lumen-localized thioredoxin-like domain of SOQ1 [31] could be involved, directly or indirectly, in a biochemical modification that causes the modulation in antenna interactions.

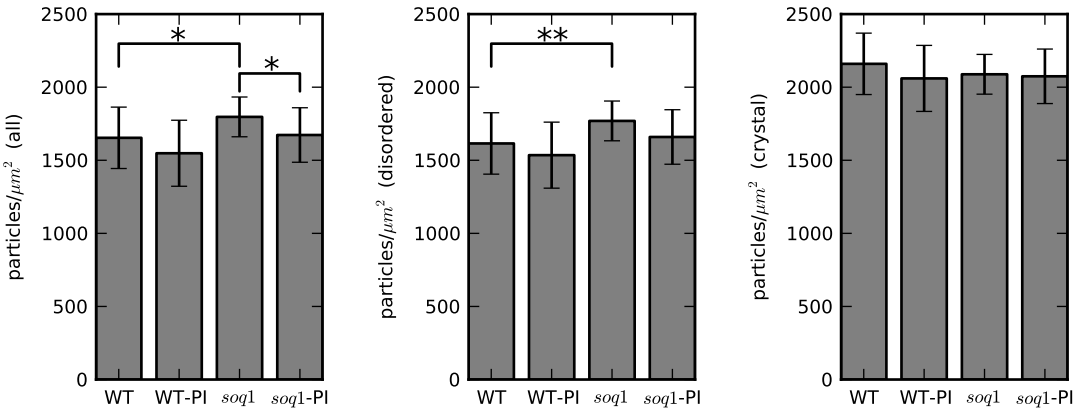


Figure 5.6: Particle number density in all (left), disordered (center), and crystalline (right) regions of grana membranes. Bars and error bars are means and SEM, respectively, from each experimental condition ($N = 5$–28 micrographs). All p-values are from Welch's t-test; * indicates $p < 0.05$, and ** $p < 0.01$.

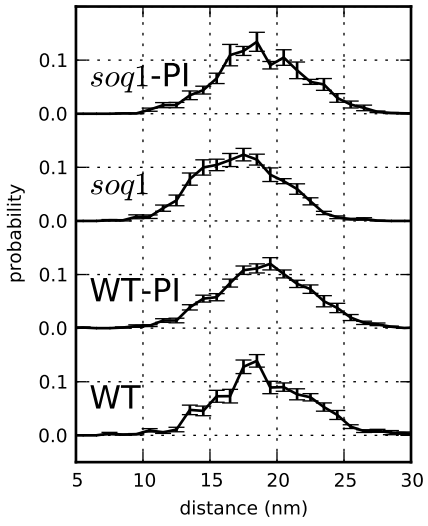


Figure 5.7: Nearest-neighbor distance distributions of disordered particles in WT, WT-PI, *soq1*, and *soq1*-PI membranes, from bottom to top; error bars are SEM ($N = 5$–28 micrographs).

In the future, possible structural roles of SOQ1, or of other grana proteins, could be further investigated by continued collection of AFM or EM data sets. PSII particles picked from these images could be classified using the GMM classifier presented here, or clustering and model building could be repeated on the expanded data sets. While the data set used to build our model uses much of the data at our disposal, it is still relatively small and incomplete; we anticipate that clustering models trained on expanded data sets may yield different results.

Chapter 6

Conclusions

6.1 Implications

Plant photosynthetic membranes are remarkable examples of active materials in which emergent phenomena underlie biological functions that are essential to life on Earth, providing a rich source of opportunities for fruitful applications of soft matter physics to open biological questions. This dissertation has explored the driving forces for protein organization in photosynthetic membranes from the perspective of statistical thermodynamics, and demonstrates that modeling and experiment have great potential to play complementary roles in future studies of nanoscale processes in plant photosynthetic membranes.

Specifically, we have shown in Chapters 2–4 that simulations of relatively simple, biologically-motivated models of protein complexes in stacked thylakoid membranes can capture a broad range of known structural motifs, including extended PSII arrays, composition gradients between thylakoid subsystems ("lateral heterogeneity"), and LHCII aggregates. By studying PSII arrays within the theoretical framework of thermodynamic phase transitions, we found that the arrays are evidence of a true crystalline phase, which coexists with a fluid phase in the experimentally-relevant region of our predicted phase diagram. Our results suggested the existence of a driving force for crystallization in vivo, providing a mechanism for collective rearrangements of the photosynthetic light harvesting and energy conversion apparatus. Two key components of this driving force are stroma-side stacking interactions between LHCII particles in adjacent membrane layers, and the high protein packing fraction in grana membranes; we found that these factors also drive lateral heterogeneity.

Although the topics we studied with simulation are long-standing questions in the photosynthesis literature, future experimental work will be needed to test our predictions. For instance, the PSII-LHCII phase diagram in Figure 3.10 can be tested by measuring the partitioning of PSII complexes between crystalline and fluid phases at many points in (ϕ, ρ) parameter space using EM or AFM, and comparing the crystallinity to the predicted tie lines through the coexistence region. This approach would be greatly aided by a quantitative method of assigning PSII particles in nanoscale microscopy images as crystalline or disordered that standardizes the qualitative manual assignments used in the past. Towards this end, in

Chapter 5 we developed a statistical learning pipeline that can classify the nearest-neighbor shell of individual PSII particles within a taxonomy of PSII crystalline unit cells. Applying this classifier to PSII particle positions picked from AFM images of grana membranes isolated from wild-type and *soq1* mutant *Arabidopsis* plants, we found evidence that the absence of the SOQ1 protein modulates interactions between peripheral antenna complexes; this structural phenotype could be associated with the functional phenotype of SOQ1 in suppressing an antenna-associated component of NPQ.

Other recent advances in experimental techniques for thylakoid biophysics could also aid in the parameterization and verification of physical and statistical models of thylakoid protein nano- and mesoscale structure. Reconstituted proteoliposomes like the one demonstrated in [187] allow greater control over the identity and quantity of each component of the system, creating experimental model systems that are more directly comparable to simulation models. In addition, AFM methods including high-speed AFM [36] and force-spectroscopy AFM [119] have been used to determine protein interaction free energies and diffusion coefficients in bacterial photosynthetic membranes; similar studies of the plant photosynthetic apparatus could inform and test coarse-grained models.

6.2 Future directions

This dissertation opens up several new avenues of interdisciplinary research in the molecular biophysics of photosynthesis. Here I outline a few specific projects (in no particular order) that one could pursue as extensions of this work.

Membrane morphology Similar approaches to that of Chapters 2–4 could be used to understand the morphogenesis of the plant thylakoid membrane. There is experimental evidence that crowding in the aqueous stroma induces an entropic attraction between membrane layers in grana stacks [100]. This hypothesis could be tested by constructing a triangulated [65] or patchy-particle [144] membrane with a thylakoid-like topology, equilibrating the membrane in the presence of varying densities of hard spheres in the stroma-like space, and measuring grana cohesion. In addition, the influence of energetic forces that have been proposed to play important roles in grana formation [38, 154] could be investigated by introducing appropriate coarse-grained potentials between fluctuating membrane sites. Independent control over each driving force has not been achieved in experiment, but could be straight-forward in simulation, allowing computational studies to advance our understanding of the biophysics of thylakoid morphology.

Energy transfer Ensembles of particle configurations taken from coarse-grained simulations like those in Chapters 2–4 could be used as the foundation for future studies of energy transfer in photosynthetic membranes. Models have been developed that yield energy transfer rate matrices within PSII supercomplexes, including a recent study by the Fleming group [19], but they have not been extended to study energy transfer on longer length scales, such as

between PSII supercomplexes in a crystal, within pools of peripheral antenna, in the presence of qE quenching sites, between stacked membrane layers, or in heterogeneous environments with all of these features. Observables that could depend on the mesoscale structure of the grana in interesting ways include exciton diffusion path lengths and timescales, light harvesting quantum efficiency (i.e., the fraction of absorbed photons that lead to productive photochemistry), and the existence of spatial regions that are energetically disconnected from each other (e.g., LHCII aggregates that are disconnected from any PSII reaction center). Each of these observables are calculable given a set of chlorophyll coordinates extracted from a coarse-grained particle-based simulation like those of Chapters 3 and 4, a rate matrix for energy transfer within pigment-protein complexes like that of [19], a model for short-range energy transfer between complexes, and standard network data analysis methods. A collaboration between myself and the authors of [19] is underway to pursue this strategy.

PSII supercomplex assembly PSII repair requires the complete disassembly of the PSII supercomplex [13], and qE [23] and state transitions [87] may involve at least partial supercomplex disassembly as well. The model of the C_2S_2 PSII supercomplex proposed in Chapter 2 is insufficient to capture phenomena in which supercomplex disassembly is a key feature. Instead, a more detailed model would be appropriate for the study of these questions, with a larger number of particle types (at least five for the PSII core monomer, LHCII trimer, CP24, CP26, and CP29; possibly also PsbS, cyt b_6f, Lhcb123 monomers, etc.) that obey potentials that favor the supercomplex assembly pathway suggested by Figure 1.3. Due to the large number of particle types, the highly anisotropic self-assembly target, the possible importance of membrane-mediated interactions, and the lack of experimental results to guide intuition, a phenomenological model like that of Chapter 2 is likely to be underconstrained and difficult to parameterize, with or without the use of path sampling techniques. The best approach here may be to attempt a systematic coarse-graining by calculating potentials of mean force between pairs of PSII supercomplex components in atomistic membrane-pigment-protein simulations.

Coupling through the lumen The lumenal space within grana stacks is crowded [104]. Steric repulsions between the extrinsic portions of PSII complexes are likely to create lumen-side couplings between adjacent membrane layers; other types of protein-protein interactions could be important as well. Together with the stroma-side coupling explored in Chapter 3, this may lead to cooperative phase transitions that propagate throughout the full vertical extent of the grana stack. This hypothesis could be investigated by extending the model of Chapter 2 by providing PSII particles with a volume-excluding interaction site on the lumenal side (e.g., a coarse-grained version of the steric model in Figure 3 of Ref. [104]), and simulating a PSII-LHCII mixture in a taller grana stack geometry. It would be difficult but not impossible to test predictions generated by this analysis using electron tomography [45]; an interesting alternative could be to predict material properties such as stiffness that are observable by AFM [147].

State transitions as phase transitions State transitions (qT) involve the transport of phosphorylated LHCII from PSII-rich membranes in "state 1" to PSI-rich membranes in "state 2," leading to a smoothing of composition gradients between membrane regions (reviewed in Chapter 1 and, e.g., [6, 128]). Because a large fraction of LHCII complexes participate in this process, it may be subject to interesting collective effects due to a demixing transition. For instance, the fraction of LHCII that localize to stroma lamellae at equilibrium could have a nonlinear dependence on the fraction of LHCII that are phosphorylated. To the best of our knowledge, this data set has not been collected experimentally, in part because the notion of an equilibrium or steady-state is not well defined for this biological context; state transitions occur on intermediate time scales (minutes to tens of minutes) and are triggered in situations that also trigger processes that occur on faster (qE, seconds to minutes) and slower (protein translation, tens of minutes to hours) time scales.

Several different hypotheses have been put forward for the physicochemical driving force created by LHCII phosphorylation that leads to state transitions [7]. (1) LHCII phosphorylation could disrupt attractive LHCII stacking interactions in the grana, either due to the additional negative charge or to protein conformational changes, weakening the driving force for localization of phosphorylated LHCII to grana. (2) LHCII phosphorylation could disrupt lateral attractions between LHCII and PSII, favoring dissociation of LHCII from PSII and subsequent demixing [87]; dissociation is less pronounced in plants than in algae [186]. (3) Because LHCII-PSI co-complexes can be isolated from membranes and are more common in state 2 [111, 63], phosphorylation may increase the binding affinity of LHCII to PSI. Each of these hypotheses could be tested in simulations by modifying the appropriate pair potentials for LHCII particles that are tagged as "phosphorylated." Preliminary work to test hypothesis (1) using a model similar to that of Chapter 4 has not shown any evidence of cooperativity, but this model did not allow for realistic membrane destacking transitions. A model that includes explicit membrane bending fluctuations, or an effective potential that mimics bending fluctuations, would be useful for further investigations.

Other crystals Chapter 3 focuses on a specific C_2S_2M PSII crystal, but Chapter 5 makes it clear that other unit cells are also prevalent in grana membranes in vivo and should be included in a complete PSII-LHCII phase diagram. When $\epsilon_{SL-P} \neq 0$, a C_2S_2 crystal is favored by the $(C_2S_2)_2$ interaction highlighted in Figure 3.1 and is at least metastable. A straightforward extension of Chapter 3 would be to conduct similar simulations with $\epsilon_{SL-P} \neq 0$ to add a C_2S_2 crystal region to the phase diagram in Figure 3.10. We have not observed metastability of the other unit cell classes in Chapter 5 using the model of Chapter 2; a more detailed model, for instance at the level described in the "PSII supercomplex assembly" paragraph above, may be necessary to stabilize these other crystals.

Mobility It is very appealing to use coarse-grained models in which desired thermodynamic properties have been demonstrated to explain or predict the kinetics of protein mobility that accompany perturbative phenomena like qE, state transitions, and PSII photoinhibition and repair. The biggest challenge with projects of this nature is in posing the right questions,

namely ones where simulation could lead to nontrivial, falsifiable, experimentally-accessible predictions. On the simulation side, one must select a dynamical scheme that yields sufficiently realistic statistics for particle displacements over time in a dense system at phase coexistence and can access experimentally relevant time scales (at least seconds), and use biological intuition when constructing the time-varying part of the Hamiltonian. On the experimental side, thylakoid protein mobility is typically probed using native chlorophyll fluorescence, traditionally by FRAP [103, 66, 91, 79, 106] and recently by FCS [86], yielding a signal that is difficult to interpret rigorously because it convolutes a multicomponent, multiphase system in a confined geometry. Advances in high-speed AFM by the Bustamante and Scheuring groups [36] or in near-field cathololuminescence microscopy by the Ginsberg group [95], or further use of single-particle tracking methods [40], could yield signatures of thylakoid protein mobility that are more easily compared to those measured by simulations.

6.3 Broader impacts

Photosynthesis is a venerable topic of study in part because it spans so many length and time scales and can be approached by scientists from so many backgrounds. Indeed, as someone who could claim to be a computational biophysical chemist, the interdisciplinary nature of basic research in photosynthesis was a large part of its original appeal to me and has been a continuing source of intellectual stimulation. More broadly, however, photosynthesis is also foundational to our economy, society, and global environment. Although the work in this dissertation is many steps removed from crops that could fix more carbon, feed more people, yield more biofuels, or better adapt to climate change, or from commercially-viable synthetic systems that capture more solar energy or convert it into more useful forms, these larger societal goals have been motivating to me as well. It would be an honor if this work contributes to the future development of design principles for light-harvesting and energy conversion systems that can more efficiently function across the wide range of environmental conditions that the world offers.

Bibliography

[1] J. P. Abrahams, A. G. W. Leslie, R. Lutter, and J. E. Walker. Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature*, 370:621–628, 1994.

[2] Z. Adam, D. Charuvi, O. Tsabari, R. R. Knopf, and Z. Reich. Biogenesis of thylakoid networks in angiosperms: knowns and unknowns. *Plant Molecular Biology*, 76(3-5):221–234, July 2011.

[3] P.-A. Albertsson. A quantitative model of the domain structure of the photosynthetic membrane. *Trends in Plant Science*, 6(8):349–58, Aug. 2001.

[4] B. J. Alder and T. E. Wainwright. Phase transition for a hard sphere system. *Journal of Chemical Physics*, 27(5):1208, Nov. 1957.

[5] S. Allakhverdiev, F. Müh, and T. Renger. Refined structure-based simulation of plant light-harvesting complex II: Linear optical spectra of trimers and aggregates. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1817(8):1446–1460, 2012.

[6] J. F. Allen. Protein phosphorylation in regulation of photosynthesis. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1098(3):275–335, 1992.

[7] J. F. Allen and J. Forsberg. Molecular recognition in thylakoid structure and function. *Trends in Plant Science*, 6(7):317–326, 2001.

[8] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, USA, 1989.

[9] T. K. Antal, I. B. Kovalenko, A. B. Rubin, and E. Tyystjärvi. Photosynthesis-related quantities for education and modeling. *Photosynthesis Research*, Oct. 2013.

[10] U. Armbruster, M. Labs, M. Pribil, S. Viola, W. Xu, M. Scharfenberg, A. P. Hertle, U. Rojahn, P. E. Jensen, F. Rappaport, P. Joliot, P. Dörmann, G. Wanner, and D. Leister. Arabidopsis CURVATURE THYLAKOID1 proteins modify thylakoid architecture by inducing membrane curvature. *Plant Cell*, 25(7):2661–78, July 2013.

[11] S. Asakura and F. Oosawa. Interaction between particles suspended in solutions of macromolecules. *Journal of Polymer Science*, 33(126):183–192, Dec. 1958.

[12] J. R. Austin and L. A. Staehelin. Three-dimensional architecture of grana and stroma thylakoids of higher plants as determined by electron tomography. *Plant Physiology*, 155(4):1601–1611, 2011.

[13] E. Baena-González and E.-M. Aro. Biogenesis, assembly and turnover of photosystem II units. *Philosophical Transactions of the Royal Society B, Biological Sciences*, 357(1426):1451–9; discussion 1459–60, Oct. 2002.

[14] J. Barber. Influence of surface charges on thylakoid structure and function. *Annual Review of Plant Physiology*, 33(1):261–295, June 1982.

[15] T. Barros and W. Kühlbrandt. Crystallisation, structure, and function of plant light-harvesting complex II. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1787(6):753–772, 2009.

[16] R. Bassi, F. Rigoni, R. Barbato, and G. M. Giacometti. Light-harvesting chlorophyll a/b proteins (LHCII) populations in phosphorylated membranes. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 936(1):29–38, 1988.

[17] M. A. Bates and D. Frenkel. Phase behavior of two-dimensional hard rod fluids. *Journal of Chemical Physics*, 112(22):10034, 2000.

[18] S. Bellafiore, F. Barneche, G. Peltier, and J.-D. Rochaix. State transitions and light adaptation require chloroplast thylakoid protein kinase STN7. *Nature*, 433(7028):892–5, Feb. 2005.

[19] D. I. G. Bennett, K. Amarnath, and G. R. Fleming. A structure-based model of energy transfer reveals the principles of light harvesting in photosystem II supercomplexes. *Journal of the American Chemical Society*, 135(24):9164–73, June 2013.

[20] E. Bergantino, A. Segalla, A. Brunetta, E. Teardo, F. Rigoni, G. M. Giacometti, and I. Szabò. Light- and pH-dependent structural changes in the PsbS subunit of photosystem II. *Proceedings of the National Academy of Sciences*, 100(25):15265–70, Dec. 2003.

[21] L. Berthier and W. Kob. The Monte Carlo dynamics of a binary Lennard-Jones glass-forming mixture. *Journal of Physics: Condensed Matter*, 19(20):205130, May 2007.

[22] D. A. Berthold, G. T. Babcock, and C. F. Yocum. A highly resolved, oxygen-evolving photosystem II preparation from spinach thylakoid membranes. *FEBS Letters*, 134(2):231–234, 1981.

[23] N. Betterle, M. Ballottari, S. Zorzan, S. de Bianchi, S. Cazzaniga, L. Dall'osto, T. Morosinotto, and R. Bassi. Light-induced dissociation of an antenna hetero-oligomer is needed for non-photochemical quenching induction. *Journal of Biological Chemistry*, 284(22):15255–15266, May 2009.

[24] E. J. Boekema, P. E. Jensen, E. Schlodder, J. F. L. van Breemen, H. van Roon, H. V. Scheller, and J. P. Dekker. Green plant photosystem I binds light-harvesting complex I on one side of the complex. *Biochemistry*, 40(4):1029–1036, Jan. 2001.

[25] E. J. Boekema, J. F. L. van Breemen, H. van Roon, and J. P. Dekker. Arrangement of photosystem II supercomplexes in crystalline macrodomains within the thylakoid membrane of green plant chloroplasts. *Journal of Molecular Biology*, 301(5):1123–33, Sept. 2000.

[26] E. J. Boekema, H. van Roon, F. Calkoen, R. Bassi, and J. P. Dekker. Multiple types of association of photosystem II and its light-harvesting antenna in partially solubilized photosystem II membranes. *Biochemistry*, 38(8):2233–9, Feb. 1999.

[27] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler. Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annual Review of Physical*

Chemistry, 53:291–318, Jan. 2002.

[28] P. G. Bolhuis and D. Frenkel. Tracing the phase boundaries of hard spherocylinders. *Journal of Chemical Physics*, 106(2):666, 1997.

[29] P. G. Bolhuis, M. Hagen, and D. Frenkel. Isostructural solid-solid transition in crystalline systems with short-ranged interaction. *Physical Review E*, 50(6):4880–4890, Dec. 1994.

[30] K. Broess, G. Trinkunas, C. D. V. D. W.-d. Wit, J. P. Dekker, A. V. Hoek, and H. van Amerongen. Excitation Energy Transfer and Charge Separation in Photosystem II Membranes Revisited. *Biophysical Journal*, 91(10):3776–3786, 2006.

[31] M. D. Brooks, E. J. Sylak-glassman, G. R. Fleming, and K. K. Niyogi. A thioredoxin-like/β-propeller protein maintains the efficiency of light harvesting in Arabidopsis. *Proceedings of the National Academy of Sciences*, 110(29):E2733–E2740, 2013.

[32] F. L. H. Brown. Elastic modeling of biomembranes and lipid bilayers. *Annual Review of Physical Chemistry*, 59:685–712, Jan. 2008.

[33] V. Brumfeld, D. Charuvi, R. Nevo, S. G. Chuartzman, O. Tsabari, I. Ohad, E. Shimoni, and Z. Reich. A note on three-dimensional models of higher-plant thylakoid networks. *Plant Cell*, 20(10):2546–9, Oct. 2008.

[34] A. Busch and M. Hippler. The structure and function of eukaryotic photosystem I. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1807(8):864–877, 2011.

[35] S. Caffarri, R. Kouřil, S. Kereïche, E. J. Boekema, and R. Croce. Functional architecture of higher plant photosystem II supercomplexes. *EMBO Journal*, 28:2052–3063, 2009.

[36] I. Casuso, P. Sens, F. Rico, and S. Scheuring. Experimental evidence for membrane-mediated protein-protein interaction. *Biophysical Journal*, 99(7):L47–9, Oct. 2010.

[37] D. Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, USA, 1987.

[38] W. S. Chow, E.-H. Kim, P. Horton, and 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(12):1081–1090, Dec. 2005.

[39] S. G. Chuartzman, R. Nevo, E. Shimoni, D. Charuvi, V. Kiss, I. Ohad, V. Brumfeld, and Z. Reich. Thylakoid membrane remodeling during state transitions in Arabidopsis. *Plant Cell*, 20(4):1029–39, Apr. 2008.

[40] E. Consoli, R. Croce, D. D. Dunlap, and L. Finzi. Diffusion of light-harvesting complex II in the thylakoid membranes. *EMBO Reports*, 6(8):782–786, Aug. 2005.

[41] R. Croce and H. van Amerongen. Light-harvesting and structural organization of Photosystem II: from individual complexes to thylakoid membrane. *Journal of Photochemistry and Photobiology. B, Biology*, 104(1-2):142–53, 2011.

[42] P. F. Damasceno, M. Engel, and S. C. Glotzer. Predictive self-assembly of polyhedra into complex structures. *Science*, 337(6093):453–457, July 2012.

[43] J. T. Damkjæ r, S. Kereïche, M. P. Johnson, L. Kovács, A. Z. Kiss, E. J. Boekema, A. V. Ruban, P. Horton, and S. Jansson. The photosystem II light-harvesting protein Lhcb3 affects the macrostructure of photosystem II and the rate of state transitions in Arabidopsis. *Plant Cell*, 21(10):3245–3256, 2009.

[44] B. Daum and W. Kühlbrandt. Electron tomography of plant thylakoid membranes. *Journal of Experimental Botany*, 62(7):2393–2402, Mar. 2011.

[45] B. Daum, D. Nicastro, J. R. Austin, J. R. McIntosh, and W. Kühlbrandt. Arrangement of photosystem II and ATP synthase in chloroplast membranes of spinach and pea. *Plant Cell*, 22(4):1299–312, Apr. 2010.

[46] S. de Bianchi, L. Dall'Osto, G. Tognon, T. Morosinotto, and R. Bassi. Minor antenna proteins CP24 and CP26 affect the interactions between photosystem II subunits and the electron transport rate in grana membranes of Arabidopsis. *Plant Cell*, 20(4):1012–1028, 2008.

[47] F. J.-M. de Meyer, A. Benjamini, J. M. Rodgers, Y. Misteli, and B. Smit. Molecular simulation of the DMPC-cholesterol phase diagram. *Journal of Physical Chemistry B*, 114(32):10451–61, Aug. 2010.

[48] J. P. Dekker and E. J. Boekema. Supramolecular organization of thylakoid membrane proteins in green plants. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1706(1-2):12–39, Jan. 2005.

[49] R. Delosme, J. Olive, and F.-A. Wollman. Changes in light energy distribution upon state transitions: an in vivo photoacoustic study of the wild type and photosynthesis mutants from Chlamydomonas reinhardtii. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1273(2):150–158, 1996.

[50] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39(1):1–38, 1977.

[51] L. Dietzel, K. Bräutigam, S. Steiner, K. Schüffler, B. Lepetit, B. Grimm, M. A. Schöttler, and T. Pfannschmidt. Photosystem II supercomplex remodeling serves as an entry mechanism for state transitions in Arabidopsis. *Plant Cell*, 23(8):2964–2977, 2011.

[52] F. Drepper, I. Carlberg, B. Andersson, and W. Haehnel. Lateral diffusion of an integral membrane protein: Monte Carlo analysis of the migration of phosphorylated light-harvesting complex II in the thylakoid membrane. *Biochemistry*, 32(44):11915–11922, Nov. 1993.

[53] D. J. Earl and M. W. Deem. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910, Nov. 2005.

[54] S. Eberhard, G. Finazzi, and F.-A. Wollman. The dynamics of photosynthesis. *Annual Review of Genetics*, 42:463–515, Jan. 2008.

[55] E. B. El Mendoub, J.-F. Wax, and N. Jakse. Evolution of the liquid-vapor coexistence of the hard-core Yukawa fluid as a function of the interaction range. *Journal of Chemical Physics*, 132(16):164503, Apr. 2010.

[56] D. L. Ermak and J. A. McCammon. Brownian dynamics with hydrodynamic interactions. *Journal of Chemical Physics*, 69(4):1352, Aug. 1978.

[57] K. N. Ferreira, T. M. Iverson, K. Maghlaoui, J. Barber, and S. Iwata. Architecture of the photosynthetic oxygen-evolving center. *Science*, 303(5665):1831–1838, 2004.

[58] L. Foret. A simple mechanism of raft formation in two-component fluid membranes. *Europhysics Letters*, 71(3):508–514, 2005.

[59] D. Frenkel. Simulations: The dark side. *European Physical Journal Plus*, 128(1):10,

Jan. 2013.

[60] D. Frenkel and B. Smit. *Understanding Molecular Simulation, Second Edition: From Algorithms to Applications (Computational Science)*. Academic Press, 2001.

[61] R. N. Frese, J. C. Pàmies, J. D. Olsen, S. Bahatyrova, C. D. van der Weij-de Wit, T. J. Aartsma, C. Otto, C. N. Hunter, D. Frenkel, and R. van Grondelle. Protein shape and crowding drive domain formation and curvature in biological membranes. *Biophysical Journal*, 94(2):640–647, Jan. 2008.

[62] R. Fristedt, A. Willig, P. Granath, M. Crèvecoeur, J.-D. Rochaix, and A. V. Vener. Phosphorylation of photosystem II controls functional macroscopic folding of photosynthetic membranes in Arabidopsis. *Plant Cell*, 21(12):3950–64, Dec. 2009.

[63] P. Galka, S. Santabarbara, T. T. H. Khuong, H. Degand, P. Morsomme, R. C. Jennings, E. J. Boekema, and S. Caffarri. Functional analyses of the plant photosystem I-light-harvesting complex II supercomplex reveal that light-harvesting complex II loosely bound to photosystem II is a very efficient antenna for photosystem I in state II. *Plant Cell*, 24(7):2963–78, July 2012.

[64] G. Garab and C. A. Mannella. Reply: On three-dimensional models of higher-plant thylakoid networks: Elements of consensus, controversies, and future experiments. *Plant Cell*, 20(10):2549–2551, Oct. 2008.

[65] G. Gompper and D. M. Kroll. Network models of fluid, hexatic and polymerized membranes. *Journal of Physics: Condensed Matter*, 9(42):8795–8834, Oct. 1997.

[66] T. K. Goral, M. P. Johnson, A. P. R. Brain, H. Kirchhoff, A. V. Ruban, and C. W. Mullineaux. Visualizing the mobility and distribution of chlorophyll proteins in higher plant thylakoid membranes: effects of photoinhibition and protein phosphorylation. *Plant Journal*, 62(6):948–959, 2010.

[67] T. K. Goral, M. P. Johnson, C. D. P. Duffy, A. P. R. Brain, A. V. Ruban, and C. W. Mullineaux. Light-harvesting antenna composition controls the macrostructure and dynamics of thylakoid membranes in Arabidopsis. *Plant Journal*, 69(2):289–301, 2012.

[68] A. Grundmeier and H. Dau. Structural models of the manganese complex of photosystem II and mechanistic implications. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1817(1):88–105, 2012.

[69] W. I. Gruszecki, W. Grudzinski, M. Gospodarek, M. Patyra, and W. Maksymiec. Xanthophyll-induced aggregation of LHCII as a switch between light-harvesting and energy dissipation systems. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1757(11):1504–1511, 2006.

[70] S. Haferkamp, W. Haase, A. A. Pascal, H. van Amerongen, and H. Kirchhoff. Efficient light harvesting by photosystem II requires an optimized protein packing density in grana thylakoids. *Journal of Biological Chemistry*, 285(22):17020–17028, 2010.

[71] B. Hankamer, J. Barber, and E. J. Boekema. Structure and membrane organization of photosystem II in green plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, 48(1):641–671, 1997.

[72] B. Hankamer, J. Barber, and J. Nield. Structural analysis of the photosystem II core/antenna holocomplex by electron microscopy. In T. J. Wydrzynski, K. Satoh,

and J. A. Freeman, editors, *Photosystem II: The light-driven water:plastoquinone oxidoreductase*, volume 22 of *Advances in Photosynthesis and Respiration*, chapter Structural, pages 403–424. Springer-Verlag, Berlin/Heidelberg, 2005.

[73] T. A. Harroun, W. T. Heller, T. M. Weiss, L. Yang, and H. W. Huang. Experimental evidence for hydrophobic matching and membrane-mediated interactions in lipid bilayers containing gramicidin. *Biophysical Journal*, 76(2):937–945, Mar. 1999.

[74] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)*. Springer, 2011.

[75] T. K. Haxton and S. Whitelam. Design rules for the self-assembly of a protein crystal. *Soft Matter*, 8(13):3558, Mar. 2012.

[76] L. O. Hedges and S. Whitelam. Limit of validity of Ostwald's rule of stages in a statistical mechanical model of crystallization. *Journal of Chemical Physics*, 135(16):164902, Oct. 2011.

[77] W. Helfrich. Steric interaction of fluid membranes in multilayer systems. *Zeitschrift Naturforschung Teil A*, 33:305, 1978.

[78] J. Henzie, M. Grünwald, A. Widmer-Cooper, P. L. Geissler, and P. Yang. Self-assembly of uniform polyhedral silver nanocrystals into densest packings and exotic superlattices. *Nature Materials*, 11(2):131–137, Feb. 2012.

[79] M. Herbstová, S. Tietz, C. Kinzel, M. V. Turkina, and H. Kirchhoff. Architectural switch in plant photosynthetic membranes induced by light stress. *Proceedings of the National Academy of Sciences*, 109(49):20130–20135, Dec. 2012.

[80] J. K. Holm, Z. Várkonyi, L. Kovács, D. Posselt, and G. Garab. Thermo-optically induced reorganizations in the main light harvesting antenna of plants. II. Indications for the role of LHCII-only macrodomains in thylakoids. *Photosynthesis Research*, 86(1-2):275–82, Nov. 2005.

[81] A. R. Holzwarth, Y. Miloslavina, M. Nilkens, and P. Jahns. Identification of two quenching sites active in the regulation of photosynthetic light-harvesting studied by time-resolved fluorescence. *Chemical Physics Letters*, 483(4):262–267, 2009.

[82] J. Hopcroft and R. Tarjan. Algorithm 447: efficient algorithms for graph manipulation. *Comm ACM*, 16(6):372–378, 1973.

[83] P. Horton, M. P. Johnson, M. L. Perez-Bueno, A. Z. Kiss, and A. V. Ruban. Photosynthetic acclimation: Does the dynamic structure and macro-organisation of photosystem II in higher plant grana membranes regulate light harvesting states? *FEBS Journal*, 275(6):1069–1079, 2008.

[84] P. Horton, M. Wentworth, and A. V. Ruban. Control of the light harvesting function of chloroplast membranes: The LHCII-aggregation model for non-photochemical quenching. *FEBS Letters*, 579(20):4201–4206, 2005.

[85] A. Huerta, D. Henderson, and A. Trokhymchuk. Freezing of two-dimensional hard disks. *Physical Review E*, 74:61106, 2006.

[86] M. Iwai, C.-G. Pack, Y. Takenaka, Y. Sako, and A. Nakano. Photosystem II antenna phosphorylation-dependent protein diffusion determined by fluorescence correlation

spectroscopy. *Scientific Reports*, 3:2833, Jan. 2013.

[87] M. Iwai, Y. Takahashi, and J. Minagawa. Molecular remodeling of photosystem II during state transitions in Chlamydomonas reinhardtii. *Plant Cell*, 20(8):2177–2189, 2008.

[88] M. Iwai, M. Yokono, N. Inada, and J. Minagawa. Live-cell imaging of photosystem II antenna dissociation during state transitions. *Proceedings of the National Academy of Sciences*, 107(5):2337–2342, 2010.

[89] S. Izvekov and G. A. Voth. Solvent-free lipid bilayer model using multiscale coarse-graining. *Journal of Physical Chemistry B*, 113(13):4443–4455, Apr. 2009.

[90] S. Jansson, H. Stefánsson, U. Nyström, P. Gustafsson, and P.-A. Albertsson. Antenna protein composition of PSI and PSII in thylakoid sub-domains. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1320(3):297–309, 1997.

[91] M. P. Johnson, T. K. Goral, C. D. P. Duffy, A. P. R. Brain, C. W. Mullineaux, and 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(4):1468–79, Apr. 2011.

[92] P. Joliot and A. Joliot. Cyclic electron flow in C3 plants. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1757(5):362–368, 2006.

[93] E. Joly. Hypothesis: could the signalling function of membrane microdomains involve a localized transition of lipids from liquid to solid state? *BMC Cell Biology*, 5(3), 2004.

[94] N. Kamiya and J.-R. Shen. Crystal structure of oxygen-evolving photosystem II from Thermosynechococcus vulcanus at 3.7-Å resolution. *Proceedings of the National Academy of Sciences*, 100(1):98–103, 2003.

[95] D. M. Kaz, C. G. Bischak, C. L. Hetherington, H. H. Howard, X. Marti, J. D. Clarkson, C. Adamo, D. G. Schlom, R. Ramesh, S. Aloni, D. F. Ogletree, and N. S. Ginsberg. Bright Cathodoluminescent Thin Films for Scanning Nano-Optical Excitation and Imaging. *ACS Nano*, 7(11):10397–10404, Oct. 2013.

[96] S. Kereïche, A. Z. Kiss, R. Kouřil, E. J. Boekema, and P. Horton. The PsbS protein controls the macro-organisation of photosystem II complexes in the grana membranes of higher plant chloroplasts. *FEBS Letters*, 584(4):759–764, 2010.

[97] J. Kern, R. Alonso-Mori, J. Hellmich, R. Tran, J. Hattne, H. Laksmono, C. Glöckner, N. Echols, R. G. Sierra, J. Sellberg, B. Lassalle-Kaiser, R. J. Gildea, P. Glatzel, R. W. Grosse-Kunstieve, M. J. Latimer, T. A. McQueen, D. DiFiore, A. R. Fry, M. Messerschmidt, A. Miahanahri, D. W. Schager, M. M. Seibert, D. Sokaras, T.-C. Weng, P. H. Zwart, W. E. White, P. D. Adams, M. J. Bogan, S. Boutet, G. J. Williams, J. Messinger, N. K. Sauter, A. Zouni, U. Bergmann, J. Yano, and V. K. Yachnadra. Room temperature femtosecond x-ray diffraction of photosystem II microcrystals. *Proceedings of the National Academy of Sciences*, 109(25):9721–9726, 2012.

[98] A. S. Keys, C. R. Iacovella, and S. C. Glotzer. Characterizing complex particle morphologies through shape matching: Descriptors, applications, and algorithms. *Journal of Computational Physics*, 230(17):6438–6463, July 2011.

[99] A. S. Keys, C. R. Iacovella, and S. C. Glotzer. Characterizing structure through

shape matching and applications to self-assembly. *Annual Review of Condensed Matter Physics*, 2(1):263–285, Mar. 2011.

[100] E.-H. Kim, W. S. Chow, P. Horton, and J. M. Anderson. Entropy-assisted stacking of thylakoid membranes. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1708(2):187–95, June 2005.

[101] H. Kirchhoff. Molecular crowding and order in photosynthetic membranes. *Trends in Plant Science*, 13(5):201–207, 2008.

[102] H. Kirchhoff, W. Haase, S. Wegner, R. Danielsson, R. Ackermann, and P.-A. Albertsson. Low-light-induced formation of semicrystalline photosystem II arrays in higher plant chloroplasts. *Biochemistry*, 46(39):11169–76, Oct. 2007.

[103] H. Kirchhoff, S. Haferkamp, J. F. Allen, D. B. a. Epstein, and C. W. Mullineaux. Protein diffusion and macromolecular crowding in thylakoid membranes. *Plant Physiology*, 146(4):1571–1578, Apr. 2008.

[104] H. Kirchhoff, C. Hall, M. Wood, M. Herbstová, O. Tsabari, R. Nevo, D. Charuvi, E. Shimoni, and Z. Reich. Dynamic control of protein diffusion within the granal thylakoid lumen. *Proceedings of the National Academy of Sciences*, pages 1104–1109, Nov. 2011.

[105] H. Kirchhoff, U. Mukherjee, and H.-J. Galla. Molecular architecture of the thylakoid membrane: lipid diffusion space for plastoquinone. *Biochemistry*, 41:4872–4882, 2002.

[106] H. Kirchhoff, R. M. Sharpe, M. Herbstová, R. Yarbrough, and G. E. Edwards. Differential mobility of pigment-protein complexes in granal and agranal thylakoid membranes of C3 and C4 plants. *Plant Physiology*, 161(1):497–507, Jan. 2013.

[107] H. Kirchhoff, I. G. Tremmel, W. Haase, and U. Kubitscheck. Supramolecular photosystem II organization in grana thylakoid membranes: evidence for a structured arrangement. *Biochemistry*, 43(28):9204–9213, July 2004.

[108] A. Z. Kiss, A. V. Ruban, and P. Horton. The PsbS protein controls the organization of the photosystem II antenna in higher plant thylakoid membranes. *Journal of Biological Chemistry*, 283(7):3972–8, Feb. 2008.

[109] R. Kouřil, J. P. Dekker, and E. J. Boekema. Supramolecular organization of photosystem II in green plants. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1817(1):2–11, 2012.

[110] R. Kouřil, E. Wientjes, J. B. Bultema, R. Croce, and E. J. Boekema. High-light vs. low-light: Effect of light acclimation on photosystem II composition and organization in Arabidopsis thaliana. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1827(3):411–419, 2013.

[111] R. Kouřil, A. Zygadlo, A. A. Arteni, C. D. de Wit, J. P. Dekker, P. E. Jensen, H. V. Scheller, and E. J. Boekema. Structural characterization of a complex of photosystem I and light-harvesting complex II of Arabidopsis thaliana. *Biochemistry*, 44(33):10935–40, Aug. 2005.

[112] L. Kovács, J. T. Damkjæ r, S. Kereïche, C. Ilioaia, A. V. Ruban, E. J. Boekema, S. Jansson, and P. Horton. Lack of the light-harvesting complex CP24 affects the structure and function of the grana membranes of higher plant chloroplasts. *Plant Cell*,

18(11):3106–20, Nov. 2006.

[113] G. Kurisu, H. Zhang, J. L. Smith, and W. A. Cramer. Structure of the cytochrome b6f complex of oxygenic photosynthesis: tuning the cavity. *Science*, 302(5647):1009–1014, 2003.

[114] S. Lemeille and J.-D. Rochaix. State transitions at the crossroad of thylakoid signalling pathways. *Photosynthesis Research*, 106(1-2):33–46, Nov. 2010.

[115] J. E. Lennard-Jones. On the determination of molecular fields. II. From the equation of state of a gas. *Proceedings of the Royal Society of London. Series A*, 106(738):463–477, Oct. 1924.

[116] X.-P. Li, O. Björkman, C. Shih, A. R. Grossman, M. Rosenquist, S. Jansson, and K. K. Niyogi. A pigment-binding protein essential for regulation of photosynthetic light harvesting. *Nature*, 403(6768):391–5, Jan. 2000.

[117] X.-P. Li, P. Muller-Moule, A. M. Gilmore, and K. K. Niyogi. PsbS-dependent enhancement of feedback de-excitation protects photosystem II from photoinhibition. *Proceedings of the National Academy of Sciences*, 99(23):15222–7, Nov. 2002.

[118] R. Lipowsky and S. Leibler. Unbinding transitions of interacting membranes. *Physical Review Letters*, 56(23):2541–2544, June 1986.

[119] L.-N. Liu, K. Duquesne, F. Oesterhelt, J. N. Sturgis, and S. Scheuring. Forces guiding assembly of light-harvesting complex 2 in native membranes. *Proceedings of the National Academy of Sciences of the United States of America*, 108(23):9455–9459, June 2011.

[120] Z. Liu, H. Yan, K. Wang, T. Kuang, J. Zhang, L. Gui, X. An, and W. Chang. Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution. *Nature*, 428(6980):287–92, Mar. 2004.

[121] S. J. Ludtke, P. R. Baldwin, and W. Chiu. EMAN: semiautomated software for high-resolution single-particle reconstructions. *Journal of Structural Biology*, 128:82–97, 1999.

[122] C. Lunde, P. E. Jensen, A. Haldrup, J. Knoetzel, and H. V. Scheller. The PSI-H subunit of photosystem I is essential for state transitions in plant photosynthesis. *Nature*, 408(6812):613–5, Nov. 2000.

[123] D. Marenduzzo, K. Finan, and P. R. Cook. The depletion attraction: an underappreciated force driving cellular organization. *Journal of Cell Biology*, 175(5):681–686, Dec. 2006.

[124] H. T. McMahon and J. L. Gallop. Membrane curvature and mechanisms of dynamic cell membrane remodelling. *Nature*, 438(7068):590–596, Dec. 2005.

[125] M. Mehta and D. A. Kofke. Coexistence diagrams of mixtures by molecular simulation. *Chemical Engineering Sciences*, 49(16):2633–2645, 1994.

[126] N. Meilhac and N. Destainville. Clusters of proteins in biomembranes: insights into the roles of interaction potential shapes and of protein diversity. *Journal of Physical Chemistry B*, 115(22):7190–7199, 2011.

[127] C. Mellwig and B. Böttcher. A unique resting position of the ATP-synthase from chloroplasts. *Journal of Biological Chemistry*, 278(20):18544–9, May 2003.

[128] J. Minagawa. State transitions—The molecular remodeling of photosynthetic super-

complexes that controls energy flow in the chloroplast. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1807(8):897–905, 2011.

[129] A. G. Moreira and R. R. Netz. Binding of similarly charged plates with counterions only. *Physical Review Letters*, 87(7):78301, July 2001.

[130] T. Morosinotto, R. Bassi, S. Frigerio, G. Finazzi, E. P. Morris, and J. Barber. Biochemical and structural analyses of a higher plant photosystem II supercomplex of a photosystem I-less mutant of barley. *FEBS Journal*, 273(20):4616–4630, 2006.

[131] T. Morosinotto, A. Segalla, G. M. Giacometti, and R. Bassi. Purification of structurally intact grana from plants thylakoids membranes. *Journal of Bioenergetics and Biomembranes*, 42(1):37–45, Feb. 2010.

[132] J. E. Mullet and C. J. Arntzen. Simulation of grana stacking in a model membrane system. Mediation by a purified light-harvesting pigment-protein complex from chloroplasts. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 589(1):100–117, 1980.

[133] D. J. Murphy. The molecular organisation of the photosynthetic membranes of higher plants. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 864(1):33–94, 1986.

[134] L. Mustárdy, K. Buttle, G. Steinbach, and G. Garab. The three-dimensional network of the thylakoid membranes in plants: quasihelical model of the granum-stroma assembly. *Plant Cell*, 20(10):2552–7, Oct. 2008.

[135] D. Nelson and B. Halperin. Dislocation-mediated melting in two dimensions. *Physical Review B*, 19(5):2457–2484, Mar. 1979.

[136] D. Nečas and P. Klapetek. Gwyddion: an open-source software for SPM data analysis. *Central European Journal of Physics*, 10:181–188, 2011.

[137] R. Nevo, D. Charuvi, O. Tsabari, and Z. Reich. Composition, architecture and dynamics of the photosynthetic apparatus in higher plants. *The Plant Journal*, 70(1):157–176, Apr. 2012.

[138] J. Nield and J. Barber. Refinement of the structural model for the photosystem II supercomplex of higher plants. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1757(5–6):353–361, 2006.

[139] J. Nield, E. V. Orlova, E. P. Morris, B. Gowen, M. van Heel, and J. Barber. 3D map of the plant photosystem II supercomplex obtained by cryoelectron microscopy and single particle analysis. *Nature Structural Biology*, 7(1):44–7, Jan. 2000.

[140] G. Odriozola. Revisiting the phase diagram of hard ellipsoids. *Journal of Chemical Physics*, 136(13):134505, Apr. 2012.

[141] D. L. Pagan, A. Shiryayev, T. P. Connor, and J. D. Gunton. Simple model of membrane proteins including solvent. *Journal of Chemical Physics*, 124(18):184904, May 2006.

[142] D. J. Paolillo. The three-dimensional arrangement of intergranal lamellae in chloroplasts. *Journal of Cell Science*, 6(1):243–55, Jan. 1970.

[143] R. B. Park and J. Biggins. Quantasome: size and composition. *Science*, 144(3621):1009–1011, 1964.

[144] A. Pasqua, L. Maibaum, G. Oster, D. A. Fletcher, and P. L. Geissler. Large-scale simulations of fluctuating biological membranes. *Journal of Chemical Physics*, 132(15):154107, Apr. 2010.

[145] F. Pedregosa, R. Weiss, and M. Brucher. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[146] M. J. D. Powell. A direct search optimization method that models the objective and constraint functions by linear interpolation. In S. Gomez and J.-P. Hennart, editors, *Advances in Optimization and Numerical Analysis*, pages 51–67. Springer Netherlands, Dordrecht, 1994.

[147] F. Rico, L. Picas, A. Colom, N. Buzhynskyy, and S. Scheuring. The mechanics of membrane proteins is a signature of biological function. *Soft Matter*, 9(32):7866, July 2013.

[148] A. V. Ruban, F. Calkoen, S. L. S. Kwa, R. van Grondelle, P. Horton, and J. P. Dekker. Characterisation of LHCII in the aggregated state by linear and circular dichroism spectroscopy. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1321(1):61–70, 1997.

[149] A. V. Ruban, M. P. Johnson, and C. D. P. Duffy. The photoprotective molecular switch in the photosystem II antenna. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1817(1):167–181, 2012.

[150] A. V. Ruban, M. Wentworth, A. E. Yakushevska, J. Andersson, P. J. Lee, W. Keegstra, J. P. Dekker, E. J. Boekema, S. Jansson, and P. Horton. Plants lacking the main light-harvesting complex retain photosystem II macro-organization. *Nature*, 421(6923):648–52, Feb. 2003.

[151] S. Safran. *Statistical Thermodynamics Of Surfaces, Interfaces, And Membranes (Frontiers in Physics)*. Westview Press, 2003.

[152] U. Schmidt, G. Guigas, and M. Weiss. Cluster formation of transmembrane proteins due to hydrophobic mismatching. *Physical Review Letters*, 101(12):128104, Sept. 2008.

[153] A. R. Schneider. *Picolo: Point-Intensity Classification Of Local Order*, 2013.

[154] A. R. Schneider and P. L. Geissler. Coexistence of fluid and crystalline phases of proteins in photosynthetic membranes. *Biophysical Journal*, 105(5):1161–1170, Sept. 2013.

[155] G. D. Scholes, G. R. Fleming, A. Olaya-Castro, and R. van Grondelle. Lessons from nature about solar light harvesting. *Nature Chemistry*, 3(10):763–774, 2011.

[156] E. Selstam, J. Schelin, W. P. Williams, and A. P. R. Brain. Structural organisation of prolamellar bodies (PLB) isolated from Zea mays. Parallel TEM, SAXS and absorption spectra measurements on samples subjected to freeze-thaw, reduced pH and high-salt perturbation. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1768(9):2235–45, Sept. 2007.

[157] G. A. Semenova. Particle regularity on thylakoid fracture faces is influenced by storage conditions. *Canadian Journal of Botany*, 73(10):1676–1682, 1995.

[158] L.-X. Shi, M. Hall, C. Funk, and W. P. Schröder. Photosystem II, a growing complex: Updates on newly discovered components and low molecular mass proteins. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1817(1):13–25, 2012.

[159] E. Shimoni, O. Rav-Hon, I. Ohad, V. Brumfeld, and Z. Reich. Three-dimensional organization of higher-plant chloroplast thylakoid membranes revealed by electron

tomography. *Plant Cell*, 17(9):2580–6, Sept. 2005.

[160] M. R. Shirts and J. D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *Journal of Chemical Physics*, 129(12):124105, Sept. 2008.

[161] I. Simidjiev, S. Stoylova, H. Amenitsch, T. Javorfi, L. Mustárdy, P. Laggner, A. Holzenburg, and G. Garab. Self-assembly of large, ordered lamellae from non-bilayer lipids and integral membrane proteins in vitro. *Proceedings of the National Academy of Sciences*, 97(4):1473–6, Feb. 2000.

[162] F. Smallenburg, L. Filion, M. Marechal, and M. Dijkstra. Vacancy-stabilized crystalline order in hard cubes. *Proceedings of the National Academy of Sciences*, 109(44):17886–17890, Oct. 2012.

[163] J. C. Stachowiak, E. M. Schmid, C. J. Ryan, H. S. Ann, D. Y. Sasaki, M. B. Sherman, P. L. Geissler, D. A. Fletcher, and C. C. Hayden. Membrane bending by protein-protein crowding. *Nature Cell Biology*, 14(9):944–949, Sept. 2012.

[164] L. A. Staehelin. Chloroplast structure and supramolecular organization of photosynthetic membranes. In L. A. Staehelin and C. J. Arntzen, editors, *Photosynthesis III. Photosynthetic membranes and light harvesting systems.*, pages 1–84. Springer-Verlag, 1986.

[165] L. A. Staehelin and G. W. M. V. D. Staay. Structure, composition, functional organization and dynamic properties of thylakoid membranes. In D. R. Ort and C. F. Yocum, editors, *Oxygenic Photosynthesis: The Light Reactions*, pages 11–30. Kluwer Academic Publishers, The Netherlands, 1996.

[166] J. Standfuss, A. C. Terwisscha van Scheltinga, M. Lamborghini, and W. Kühlbrandt. Mechanisms of photoprotection and nonphotochemical quenching in pea light-harvesting complex at 2.5 A resolution. *Journal of Biological Chemistry*, 24(5):919–28, Mar. 2005.

[167] K. Sznee, J. P. Dekker, R. T. Dame, H. van Roon, G. J. L. Wuite, and R. N. Frese. Jumping mode atomic force microscopy on grana membranes from spinach. *Journal of Biological Chemistry*, 286(45):39164–39171, Sept. 2011.

[168] D. M. J. Tax and R. P. W. Duin. Growing a multi-class classifier with a reject option. *Pattern Recognition Letters*, 29(10):1565–1570, July 2008.

[169] A. Tian and T. Baumgart. Sorting of lipids and proteins in membrane curvature gradients. *Biophysical Journal*, 96(7):2676–2688, Apr. 2009.

[170] M. Tikkanen and E.-M. Aro. Thylakoid protein phosphorylation in dynamic regulation of photosystem II in higher plants. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1817(1):232–238, 2012.

[171] M. Tikkanen, M. Nurmi, S. Kangasjärvi, and E.-M. Aro. Core protein phosphorylation facilitates the repair of photodamaged photosystem II at high light. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1777(11):1432–1437, 2008.

[172] R. Tokutsu, N. Kato, K. H. Bui, T. Ishikawa, and J. Minagawa. Revisiting the supramolecular organization of photosystem II in Chlamydomonas reinhardtii. *Journal of Biological Chemistry*, 287(37):31574–31581, 2012.

[173] I. G. Tremmel, H. Kirchhoff, E. Weis, and G. D. Farquhar. Dependence of plastoquinol diffusion on the shape, size, and density of integral thylakoid proteins. *Biochimica et*

Biophysica Acta (BBA) - Bioenergetics, 1607(2-3):97–109, Dec. 2003.

[174] I. G. Tremmel, E. Weis, and G. D. Farquhar. The influence of protein-protein interactions on the organization of proteins within thylakoid membranes. *Biophysical Journal*, 88(4):2650–2660, Apr. 2005.

[175] Y. Umena, K. Kawakami, J.-R. Shen, and N. Kamiya. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9âĂĽÅ. *Nature*, 473(7345):55–60, Apr. 2011.

[176] B. van Oort, M. Alberts, S. de Bianchi, L. Dall'Osto, R. Bassi, G. Trinkunas, R. Croce, and H. van Amerongen. Effect of Antenna-Depletion in Photosystem II on Excitation Energy Transfer in Arabidopsis thaliana. *Biophysical Journal*, 98(5):922–931, 2010.

[177] B. van Oort, A. Maréchal, A. V. Ruban, B. Robert, A. A. Pascal, N. C. A. de Ruijter, R. van Grondelle, and H. van Amerongen. Different crystal morphologies lead to slightly different conformations of light-harvesting complex II as monitored by variations of the intrinsic fluorescence lifetime. *Physical Chemistry Chemical Physics*, 13(27):12614–22, July 2011.

[178] B. van Oort, A. van Hoek, A. V. Ruban, and H. van Amerongen. Equilibrium between quenched and nonquenched conformations of the major plant light-harvesting complex studied with high-pressure time-resolved fluorescence. *Journal of Physical Chemistry B*, 111(26):7631–7, July 2007.

[179] I. Vass. Molecular mechanisms of photodamage in the Photosystem II complex. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1817(1):209–217, 2012.

[180] J. Veerman, M. D. McConnell, S. Vasilév, F. Mamedov, S. Styring, and D. Bruce. Functional heterogeneity of photosystem II in domain specific regions of the thylakoid membrane of Spinach (Spinacia Oleracea L). *Biochemistry*, 46(11):3443–3453, 2007.

[181] A. V. Vener. Environmentally modulated phosphorylation and dynamics of proteins in photosynthetic membranes. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1767(6):449–57, June 2007.

[182] J. D. Weeks, D. Chandler, and H. C. Andersen. Role of repulsive forces in determining the equilibrium structure of simple liquids. *Journal of Chemical Physics*, 54(12):5237, June 1971.

[183] B. West, F. L. H. Brown, and F. Schmid. Membrane-protein interactions in a generic coarse-grained model for lipid bilayers. *Biophysical Journal*, 96(1):101–115, 2009.

[184] S. Whitelam, E. H. Feng, M. F. Hagan, and P. L. Geissler. The role of collective motion in examples of coarsening and self-assembly. *Soft Matter*, 5(6):1251, Mar. 2009.

[185] S. Whitelam and P. L. Geissler. Avoiding unphysical kinetic traps in Monte Carlo simulations of strongly attractive particles. *Journal of Chemical Physics*, 127(15):154101, Oct. 2007.

[186] E. Wientjes, B. Drop, R. Kouřil, E. J. Boekema, and R. Croce. During state 1 to state 2 transition in Arabidopsis thaliana the Photosystem II supercomplex gets phosphorylated but does not disassemble. *Journal of Biological Chemistry*, pages M113.511691–, Oct. 2013.

[187] L. Wilk, M. Grunwald, P.-N. Liao, P. J. Walla, and W. Kühlbrandt. Direct interaction

of the major light-harvesting complex II and PsbS in nonphotochemical quenching. *Proceedings of the National Academy of Sciences*, 110(14):5452–5456, Mar. 2013.

[188] W. P. Williams. The physical properties of thylakoid membrane lipids and their relation to photosynthesis. In S. Paul-André and M. Norio, editors, *Lipids in Photosynthesis: Structure, Function and Genetics*, volume 6 of *Advances in Photosynthesis and Respiration*, pages 103–118. Kluwer Academic Publishers, Dordrecht, 2004.

[189] A. E. Yakushevska, P. E. Jensen, W. Keegstra, H. van Roon, H. V. Scheller, E. J. Boekema, and J. P. Dekker. Supermolecular organization of photosystem II and its associated light-harvesting antenna in Arabidopsis thaliana. *European Journal of Biochemistry*, 268(23):6020–6028, 2001.

[190] A. E. Yakushevska, W. Keegstra, E. J. Boekema, J. P. Dekker, J. Andersson, S. Jansson, A. V. Ruban, and P. Horton. The structure of photosystem II in Arabidopsis: Localization of the CP26 and CP29 antenna complexes. *Biochemistry*, 42(3):608–613, 2003.

[191] J. Zaks, K. Amarnath, D. M. Kramer, K. K. Niyogi, and G. R. Fleming. A kinetic model of rapidly reversible nonphotochemical quenching. *Proceedings of the National Academy of Sciences*, 109(39):15757–62, Sept. 2012.

[192] J. Zaks, K. Amarnath, E. J. Sylak-Glassman, and G. R. Fleming. Models and measurements of energy-dependent quenching. *Photosynthesis Research*, 116(2-3):389–409, Oct. 2013.

[193] A. Zouni, H.-T. Witt, J. Kern, P. Fromme, N. Krauss, W. Saenger, and P. Orth. Crystal structure of photosystem II from Synechococcus elongatus at 3.8 Å resolution. *Nature*, 409:739–743, 2001.

Appendix A

Experimental materials and methods

The work described in this Appendix was a collaboration supported by the SISGR grant. Matt Brooks prepared the samples, Bibiana Onoa performed the AFM data acquisition and analysis, and Patricia Grob assisted with data analysis.

A.1 Sample preparation

Arabidopsis thaliana plants were grown at 150–200 μmol photons m^{-2} s^{-1} of light (10 h/day) at 21.5 °C for 7–10 weeks. For high-light treated samples, leaves were exposed at 1,200 μmol photons m^{-2} s^{-1} for 90 minutes before isolating thylakoids. Grana membranes were isolated as described [35] with the following modifications. Leaves were homogenized in a commercial blender by 8–10 half second pulses and filtered through a 41 μm nylon membrane using a light hand-generated vacuum. Thylakoid membranes were adjusted to a chlorophyll concentration of 2.5 mg/mL and 3/16 volumes of 7.6% (w/v) n-dodecyl-α maltoside (α-DM), 15 mM NaCl, 5 mM $MgCl_2$ was added and incubated with gentle rocking for 20 minutes at 4 °C in the dark. The detergent α-DM was used rather than Triton X-100, as this treatment yielded larger and more homogeneous membranes in agreement with previous reports [131]. The final sample was frozen immediately in liquid nitrogen as 5 μL aliquots for microscopic inspection.

A.2 AFM data acquisition

Four different types of samples were imaged: WT low-light-acclimated (WT) and photoinhibited (WT-PI) grana membranes; and soq-1 mutant low-light-acclimated (*soq1*) and photoinhibited (*soq1*-PI) membranes. Grana aliquots were thawed and diluted 1:15 in deposition buffer (10 mM tris-HCl pH 7.5, 150 mM KCl and 25 mM $MgCl_2$) on freshly cleaved ultra-clean mica (Grade V, Ted Pella, Inc.), and incubated at room temperature for 1–3 hours. The mica discs were then rinsed with purified 18.2 MΩ deionized water ten times and dried under a gentle N_2 gas flow, perpendicular to the mica surface. AFM measurements were

performed with a Multimode AFM Nanoscope V (Bruker corporation, Santa Barbara, CA, USA) equipped with a vertical engagement E-scanner (15-micron × 15-micron × 2.5-micron vertical range). The samples were imaged in tapping mode; the probes were excited on resonance with free amplitudes of 2–15 nm and phases of 20–30 degrees. The image amplitude (set point As) and free amplitude (A0) ratio (As/A0) was kept at 0.8. Images were acquired using silicon cantilevers (Nanosensors), with a high-resonance frequency in the range of 280–350 kHz and a spring constant of 46 N/m. Images (512 × 512 pixels) were captured in the trace direction; the scan angle was maintained at 0 degrees. All samples were imaged at room temperature in air, at a relative humidity of 30%. More than eight different grana patches were scanned per each type of membrane. Bi-layered patches were fully mapped at scans of 500 × 500 nm. Higher resolution of 150 × 150 nm images were also recorded.

A.3 Image processing and particle picking

Raw AFM images were flattened and leveled using Gwyddion 2.3 [136]. To determine the height of the membranes and/or particles, bare mica was set at zero nm. Single particles were picked from the 500 and 150 nm images. For simplicity, the particle picking was restricted to bi-layer areas by masking out multi-layer and bare mica areas. Single particles were selected interactively using Boxer from EMAN [121], with a window size of 20 × 20 pixels. The center of mass of each particle was extracted using SPIDER. Truncated particles were manually excluded from the images.

A.4 Membrane characterization

AFM showed a heterogeneous mixture of grana membranes patches, as shown in Figure A.1. The patches varied in shape and size (from 150 nm to a few microns in diameter) indicating a general trend of grana fusion. Most of these patches were double membranes based on their average height of 11 nm, which we refer to as grana discs (Figure A.1). Some membrane patches also had higher-order stacking of additional membranes distributed randomly throughout the patch (Figure A.1a). The residual upper layers corresponded to partially disrupted grana discs, because their heights are multiples of 11 nm (Figure A.1a inset) in agreement with previous reports [103]. Our analyses were limited to the double membrane grana discs.

Each grana disc was densely packed with particles, with approximate diameters of 15–25 nm and protrusions above the lipid bilayer of 2–4 nm. High-resolution images revealed internal structure within the particles (Figure A.1b, phase inset). For some particles, two prominent peaks were easily detected with a peak-to-peak separation of approximately 7–9 nm (Figure A.1b cross-section profile inset). This separation is expected between the two extrinsic oxygen-evolving complexes (OECs) in the dimeric PSII. Thus, we assign these particles as PSII reaction centers protruding from the grana luminal face in good agreement with other groups [101, 167].

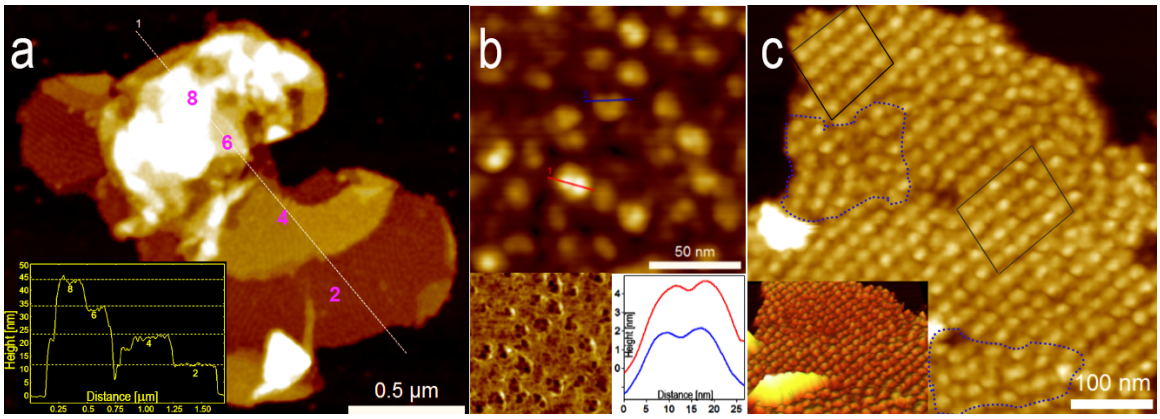


Figure A.1: AFM characterization of grana membranes. (a) Typical fused grana membranes displaying various levels of thylakoid stacking. White line is a cross-section site with a height profile represented in the inset. As many as eight different membrane layers form four stacked grana discs. z scale 0–60 nm. (b) A high resolution image of a grana disc exposes individual PSII supercomplexes protruding from the membrane. AFM Phase contrast image resolves internal structure in each supercomplex (left inset). The cross-section profile on individual complexes (red and blue lines) distinguishes two prominent peaks per complex separated by 9 nm (right inset). (c) Grana membranes can be composed of semicrystalline arrays of complexes (marked by black polygons) adjacent to disordered regions (marked by blue dotted lines). Inset shows a 3d representation to facilitate visual array detection. z scale 0–20 nm.