Biophysical Journal, Volume *119*

Supplemental Information

Modeling the Role of LHCII-LHCII, PSII-LHCII, and PSI-LHCII Interactions in State Transitions

William H.J. Wood and Matthew P. Johnson

Appendix 1. A lattice-based algorithm for calculating spatial overlap of particles of arbitrary geometry

The particles were regarded as "hard" objects, meaning that translations and rotations that resulted in any two particles occupying the same lattice site were forbidden. To calculate whether particle overlap had occured during the movement of a given particle, we developed the following algorithm (Supplementary Fig. 2).

Particle geometries, as determined from PDB structures, were represented by a matrix P_i of dimension $2 \times n$ where n is the number of lattice sites occupied by the particle and the subscript i refers to the i^{th} particle. The columns of P_i contain the coordinates of lattice sites occupied by the particle.

We define the matrix M as the concatenation of all P_i along the column axis, and is therefore a matrix which contains, in its columns, the coordinates of the lattice sites occupied by all m particles

$$M(2 \times N) = [P_1, P_2, \ldots, P_m] = \begin{bmatrix} x_1 & x_2 & \ldots & x_N \\ y_1 & y_2 & \ldots & y_N \end{bmatrix} \tag{1}$$

where N is the total number of occupied lattice sites for all particles.

During a given iteration of the simulation, the i^{th} particle may be translated and rotated (ie. the particle may be moved). Now, the coordinates of P_i have changed but the the other particles have not. So in order to assess whether the new P_i is overlapping with any other particle, we can ascertain whether there is any overlap with $M|P_i$, the matrix containing the lattice sites occupied by all particles except P_i.

Theorem 1. We define the function $N_{unique}(M)$ as the number of unique columns in M. P_i spatially overlaps at least one other particle if and only if

$$N_{unique}(M) < N_{unique}(P_i) + N_{unique}(M|P_i) \tag{2}$$

where $M|P_i$ is the columnwise concatenation of all P_j such that $j \neq i$.

Intuitively, if the total number of occupied lattice sites is less than the number of occupied by the particle of interest plus the number occupied by all other particles, and assuming there is no overlap of other particles, then the particle of interest must share a lattice site with at least one other particle (Supplementary Fig. 2).

Proof. As $N_{unique}(M)$ is equivalent to the number of elements in the set of columns of M, or $N_{unique}(M) = |Col(M)|$, and because M by definition (1) contains the columns of P_i Let x be an column in M. We have

$$x \in \{Col(M)\} \implies$$

$$x \in \{Col(P_i)\} \, or \, x \in \{Col(M|P_i)\} \, or \, x \in \{Col(P_i)\} \cap \{Col(P_i)\} \tag{3}$$

and so

$$N_{unique}(M) = N_{unique}(P_i) + N_{unique}(M|P_i) - N_{unique}(\{Col(P_i)\} \cap \{Col(P_i)\}) \tag{4}$$

Hence if

$$N_{unique}(M) < N_{unique}(P_i) + N_{unique}(M|P_i)$$

this implies

$$N_{unique}(\{Col(P_i)\} \cap \{Col(P_i)\}) > 0$$

and so there exists at least one column (and therefore at least one lattice site) common to both P_i and $M|P_i$.

Optimisation of the particle overlap calculation.
We used existing Python libraries in the calculation of $N_{unique}(M)$ but found that we could significantly reduce the time taken to compute $N_{unique}(M)$ if lattice sites where given a unique index, effectively making the lattice geometry 1-D.
In order to achieve this we define a suitable function

$$f : M(2 \times N) \rightarrow M(1 \times N) \tag{5}$$

such that the uniqueness of the columns of M is preserved. For 2 dimensions we find a suitable f in the form

$$f(M) = CM \tag{6}$$

where C is a 1×2 matrix of the form

$$C = [c_1\, 1] \tag{7}$$

and $c_1 > x$ for all x elements in M. for example, let

$$M = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

Since no elements of M are greater than or equal to 10, we may choose $c_1 = 10$

$$C = [10\, 1]$$

then

$$f(M) = CM = [10\, 1] \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} = [12\ 34\ 56]$$

In this study we used $c_1 = 10^4$.

Theorem 2. Column i of $f(M) = CM$ is unique in CM if column i in M is unique in M.

Proof. Consider column $\begin{bmatrix} x \\ y \end{bmatrix}$ of M. From (6) and (7), we have

$$f(\begin{bmatrix} x \\ y \end{bmatrix}) = \begin{bmatrix} c_1 x + y \end{bmatrix} \tag{8}$$

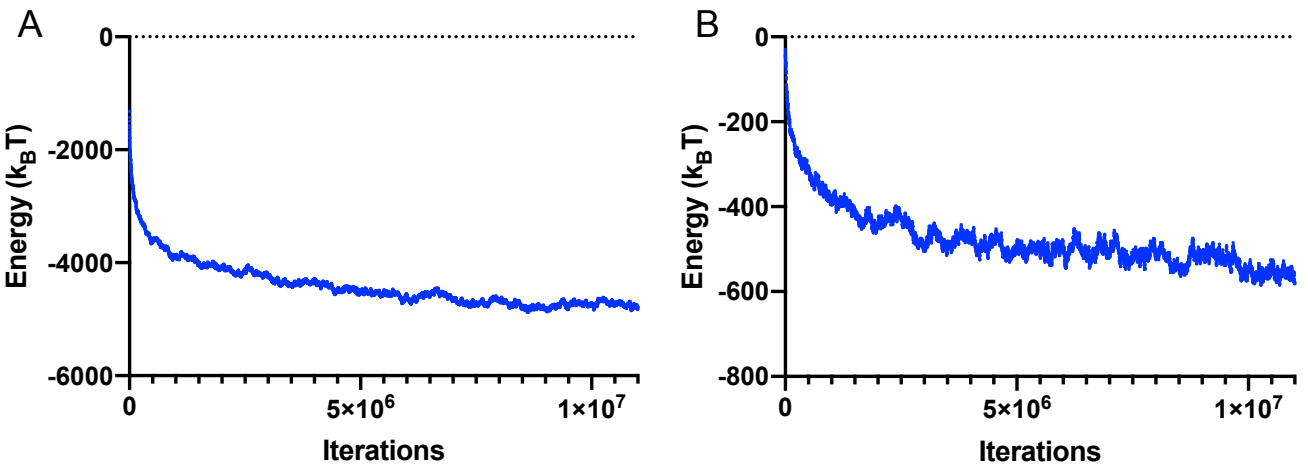
where $c_1 > x, y \; \forall x, y$. Let $x = 0$. then

$$f(\begin{bmatrix} x \\ y \end{bmatrix}) = \begin{bmatrix} y \end{bmatrix} \tag{9}$$

It is clear in this case $f(\begin{bmatrix} x \\ y \end{bmatrix})$ is uniquely defined for all y.
Now for $x > 0$

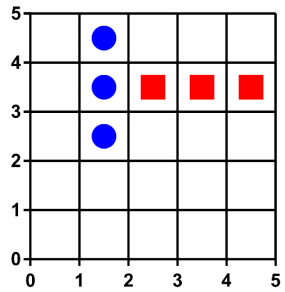
$$f(\begin{bmatrix} x \\ y \end{bmatrix}) = \begin{bmatrix} c_1 x + y \end{bmatrix} < \begin{bmatrix} c_1 x + c_1 \end{bmatrix} = f(\begin{bmatrix} x+1 \\ y \end{bmatrix}) \tag{10}$$

for any y. Note that the $<$ sign in (10) is an elementwise less than relation on the single element in the matrices. In summary, any change in x and/or y result in a unique $f(\begin{bmatrix} x \\ y \end{bmatrix})$ if $c_1 > x, y$.
By induction, the uniqueness of $f(\begin{bmatrix} 0 \\ 0 \end{bmatrix}) = \begin{bmatrix} 0 \end{bmatrix}$ implies $f(\begin{bmatrix} x \\ y \end{bmatrix})$ is unique for all x, y.



Supplementary Figure 1. Equilibration of model prior to Monte Carlo sampling. Samples were taken after 10^7 iterations to ensure the system had reached equilibrium. This corresponds to 4.5 and 5.6 time constants (Tau) of the energy decay with the number of iterations for **A**, State I and **B**, State II respectively.

A

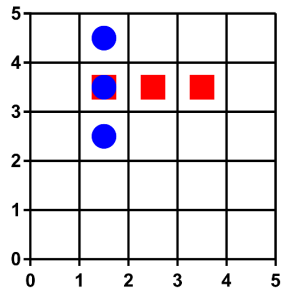


$$M = \begin{bmatrix} 1 & 1 & 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 3 & 3 & 3 \end{bmatrix}$$

$$Number\ of\ unique\ columns\ in\ M = Number\ of\ columns\ in\ M$$

No overlap

B

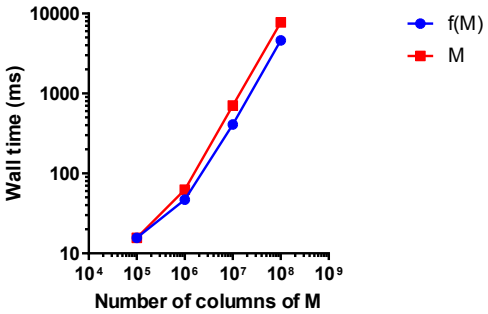


$$M = \begin{bmatrix} 1 & \boxed{1} & 1 & \boxed{1} & 2 & 3 \\ 2 & \boxed{3} & 4 & \boxed{3} & 3 & 3 \end{bmatrix}$$

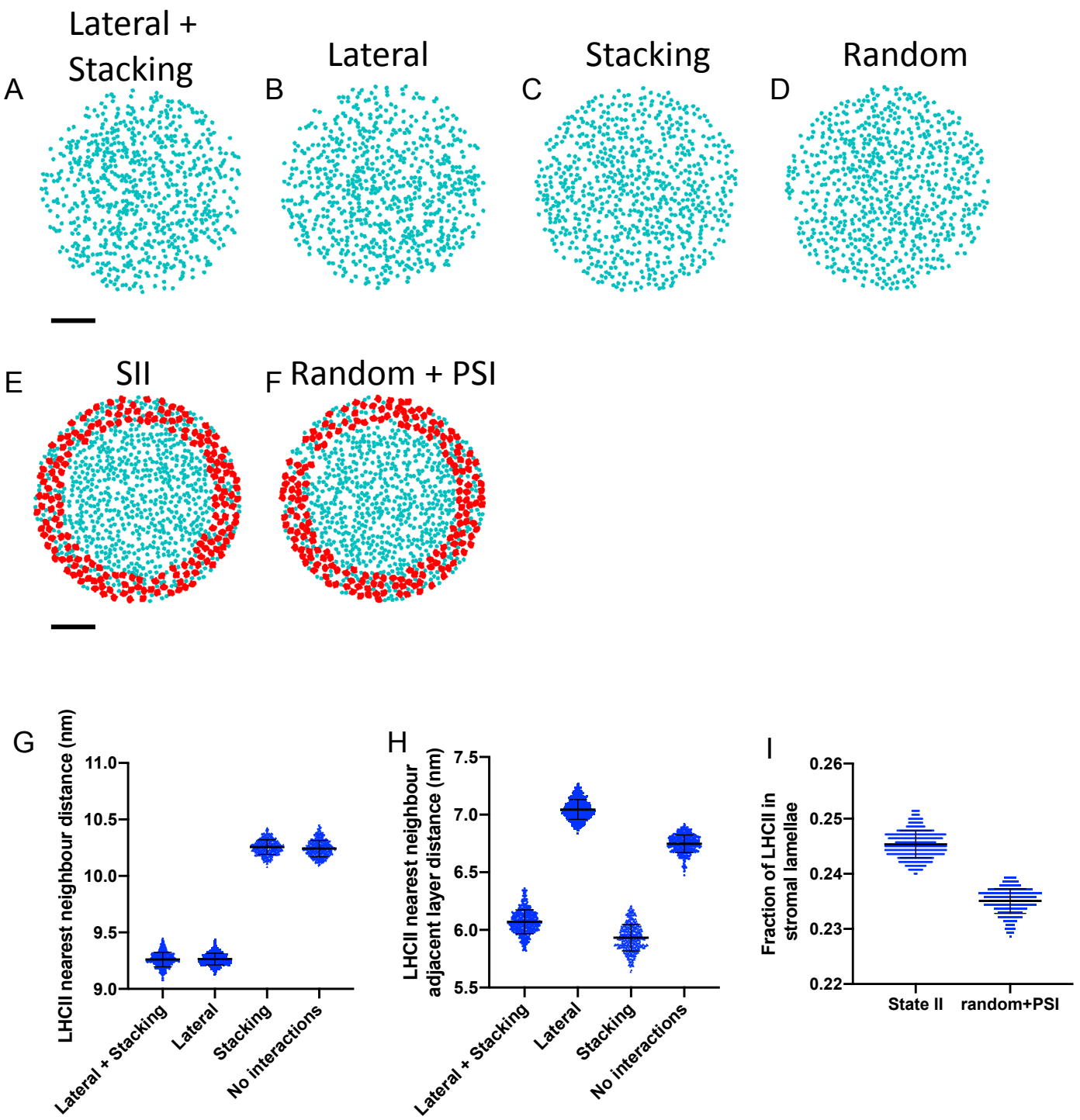
$$Number\ of\ unique\ columns\ in\ M < Number\ of\ columns\ in\ M$$

Overlap

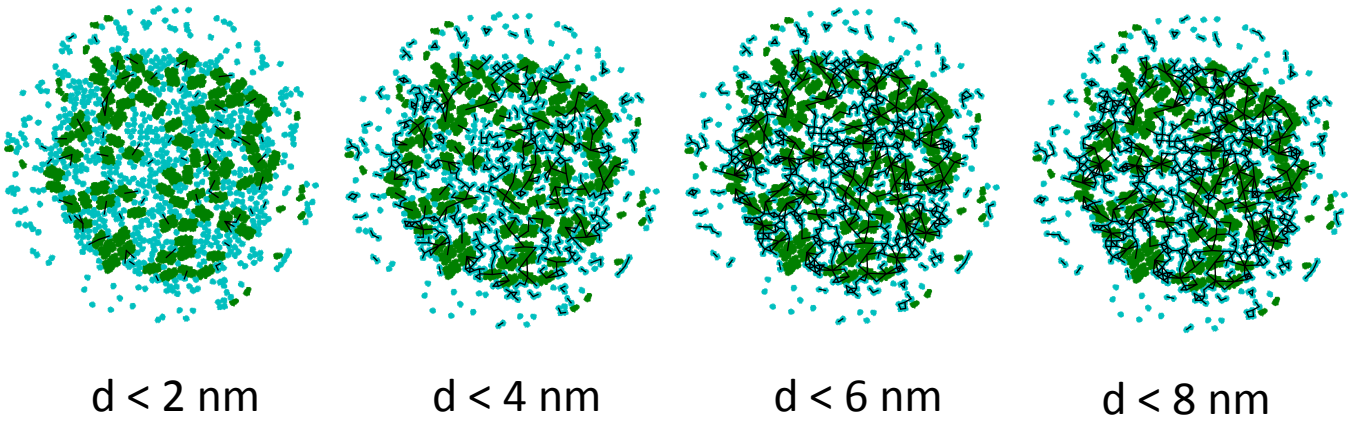
C



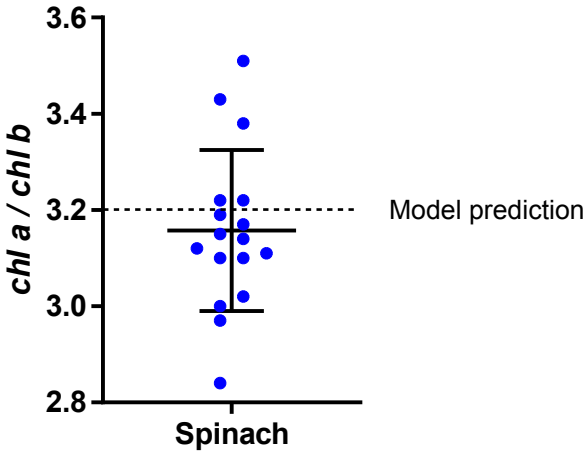
Supplementary Figure 2. Overlap algorithm for particles of arbitrary geometry. **A,** Particle one (blue circles) does not overlap particle two (red squares). In this case, the number of unique columns in the M matrix (which contains the lattice sites occupied by all particles) is equal to the total number of columns in M. **B**, Here, particle one and particle two share a lattice site (1,3) and so they overlap. As a consequence the number of unique columns of M is less than the total number of columns of M. **C,** The scalability of the overlap algorithm applied to M (red) and the transform f(M) which transforms M into a 1-D vector (blue).



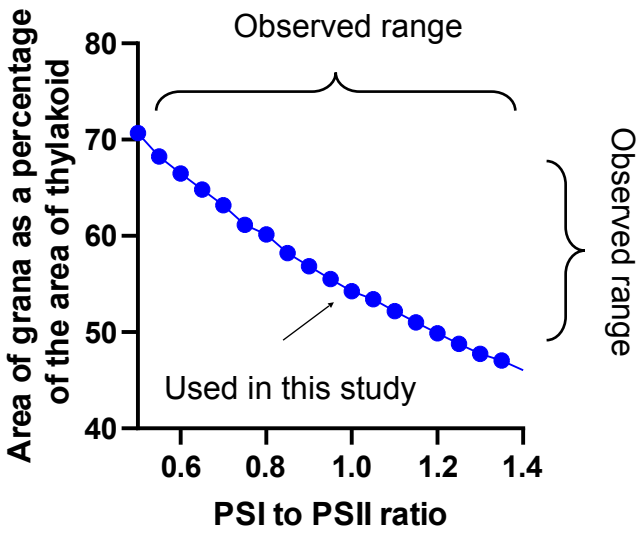
Supplementary Figure 3. Test simulations with minimal components. **A-D,** Images from LHCII only simulations with lateral and stacking interactions (a), Lateral interactions only (b), Stacking interactions only (c), and no interactions (random, d). **E,F,** images from LHCII and PSI only simulations with PSI-LHCII interactions (State II, e) or not interactions (f). **G,** Nearest neighbour distances for LHCII particles in the same layer from simulations shown in a-d. **H,** Nearest neighbour distances for LHCII particles in adjacent layers from simulations shown in a-d. **I,** Fraction of LHCII in the stromal lamellae from simulations in e,f.



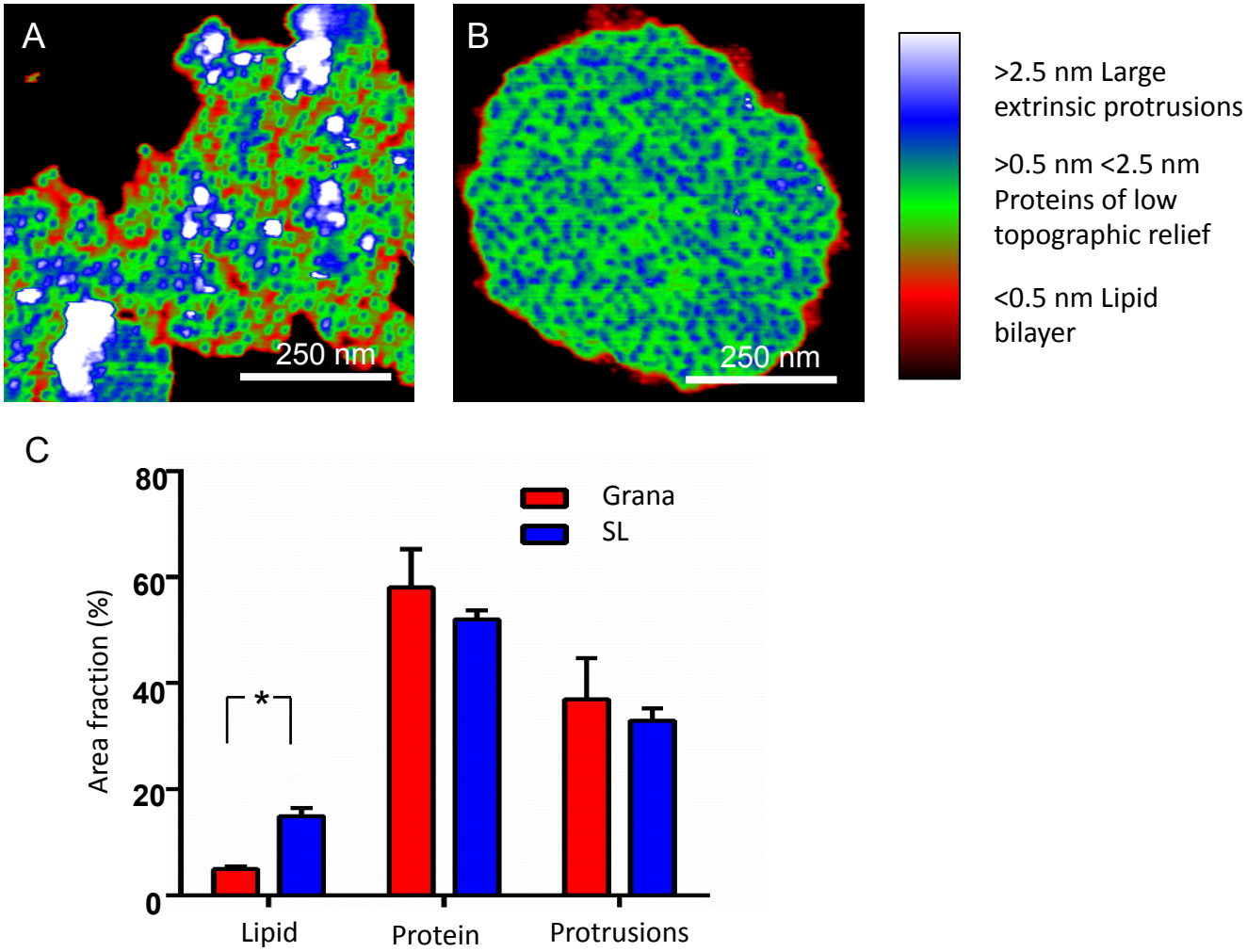
Supplementary Figure 4. Construction of Chlorophyll networks. Edges (black lines) between any two protein complexes (LHCII and PSII) are assigned if any chlorophyll on one complex is less than a certain distance called the distance threshold (d) from any chlorophyll on the other complex. As the distance threshold is increased, the chlorophyll network becomes more connected. We investigated the structure of the networks arising from simulations containing lateral and/or stacking interactions over a range of distance thresholds (2-8 nm shown above).



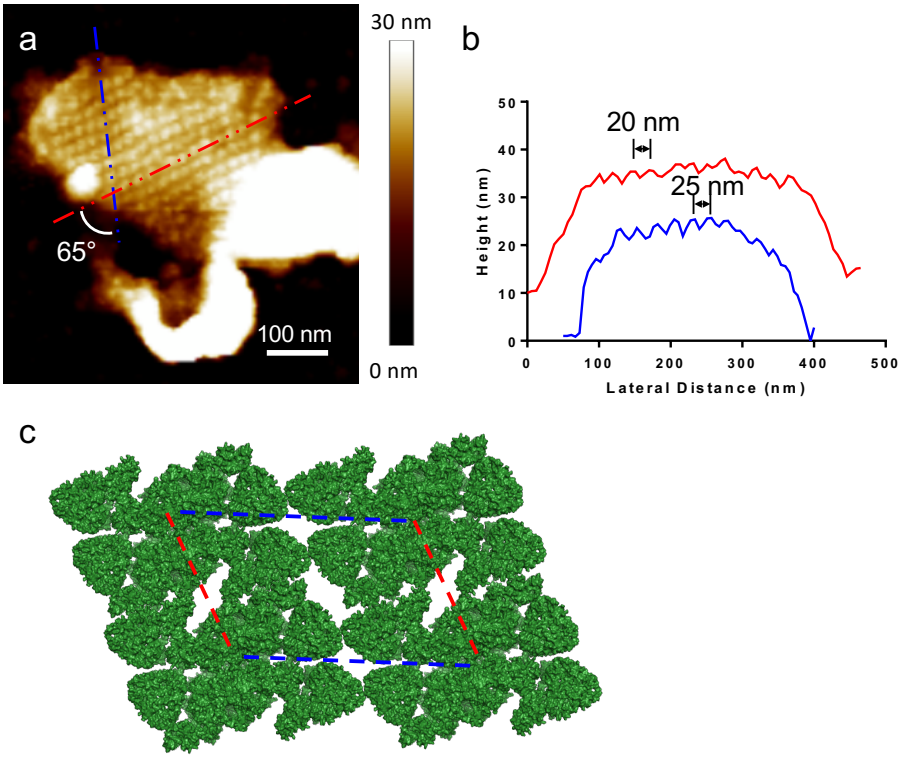
Supplementary Figure 5. Measured ratio of chlorophyll *a* to chlorophyll *b* in *Spinacia oleracea*. Error bars represent mean and standard deviation.



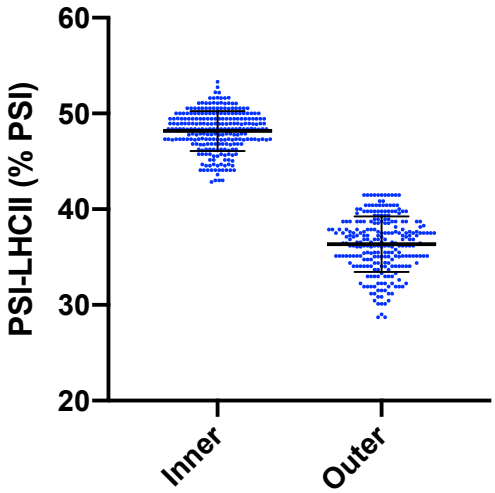
Supplementary Figure 6. The area of grana as a percentage of total grana area is determined by the PSI to PSII ration in this model. A PSI to PSII ratio of 1 was used in this study, resulting in a thylakoid composed of 55% grana by area.



Supplementary Figure 7. AFM analysis of the density of grana and stromal lamellae membranes. A,B, AFM images of stromal lamellae (A) grana (B) with colour threshold to show lipid zones in red (regions <0.5 nm above the membrane surface), proteins of low topographic relief (regions >0.5 nm and <2.5 nm above the membrane surface) and large extrinsic membrane protrusions (regions >2.5 nm above the membrane) in blue/white. **C,** A comparison of lipid, protein and protruding regions in grana (red) and stromal lamellae (blue) thylakoids (displaying mean and standard error, N = 5, * indicates $p \leq 0.001$ by t-test).



Supplementary Figure 8. Crystalline regions of PSII-LHCII in AFM images of *Arabidopsis thaliana* grana. a. PSII-LHCII crystalline region in double-layer grana membrane. **b.** Cross section height profile of red and blue dashed lines in (a). **c.** Model of PSII-LHCII ($C_2S_2M_2$, PDB : 5MDX) crystal (20x25 nm).



Supplementary Figure 9. The difference in the percentage of PSI-LHCII complexes in state II between the inner and outer stromal lamellae. PSI complexes were classed as belonging to the "inner" stromal lamellae if their centre point was closer to the outer edge of the grana than the outer edge of the stromal lamellae and classed as "outer" if they were closer to the outer edge of the stromal lamellae than to the grana.