SUPPORTING INFORMATION:
Membrane Proteins Diffuse as Dynamic Complexes with Lipids

Perttu S. Niemelä, Markus Miettinen, Luca Monticelli, Henrik Hammaren,
Pär Bjelkmar, Teemu Murtola, Erik Lindahl, and Ilpo Vattulainen

May 4, 2010

Contents

1 Simulation Details

1.1 Membrane protein Kv1.2

The starting structure of Kv1.2 was based on the crystal structure determined by MacKinnon and co-workers [1] (PDB code: 2A79). Coordinates were taken from a previous study of the same protein [2], see details therein for the assembly of the protein system. The protein was embedded in a palmitoyoleoylphosphatidylcholine (POPC) lipid bilayer consisting of 910 lipids, using the approach described by Kandt *et al.* [3]. The extracellular leaflet (lower z coordinate) consisted of 451 POPC molecules and the intracellular leaflet (higher z coordinate) of 459 POPC molecules. The system was solvated with 53,188 water molecules. 160 K+ and 144 Cl- ions were added to achieve near-physiological salt concentration and electrical neutrality. The total number of atoms in the system was 225,300 and the size was approximately 19 x 19 x 8.5 nm^3.

The OPLS-AA force field [4] was used for the protein, in combination with the Berger force field [5] for the lipids and the SPC water model [6]. This combination was previously tested and shown to reproduce reasonably well the behavior of membrane peptides [7]. Simulations were carried out in the NpT ensemble using semi-isotropic pressure coupling (Parrinello-Rahman [8, 9] algorithm, with $p = 1$ bar both in the membrane plane and perpendicular to it; time constant of 4 ps) and the Nose-Hoover thermostat[10, 11] ($T = 310$ K, time constant of 1 ps). Periodic boundary conditions were applied in all dimensions. A cutoff of 1.0 nm was applied for the calculation of Lennard-Jones interactions. The PME algorithm [12] was used for electrostatic interactions, with a cutoff of 1.0 nm for interactions in the real space. Reciprocal space interactions were evaluated on a 0.16 nm grid, using a fourth-order B-spline interpolation. All bond lengths of non-water molecules were constrained with the LINCS algorithm [13, 14], while the SETTLE algorithm was used to constrain bond lengths and angles in the water molecules [15]. A time step of 4 fs was used and the neighbor list for non-bonded interactions was updated every 5 steps. All simulations were carried out with Gromacs 4.0.516.

After energy minimization (steepest descent algorithm, 500 steps), a short equilibration run (20 ns) was carried out with the protein frozen in the initial conformation (Fc = 1000 kJ mol^{-1} nm^{-2}), to allow the relaxation of the membrane. Next, a production run of 600 ns was carried out, and the final 500 ns were used for analysis.

In order to gauge finite size effects, we carried out an additional simulation for a system that was largely similar to the first one. The main differences were its larger size (1 protein + 2220 POPC lipids) and the use of a hexagonal periodic box in the membrane plane instead of a rectangular one. This system was hydrated

with 132,245 water molecules, together with 160 K+ and 144 Cl- ions. After equilibration, a production run of 150 ns was carried out, which was used for analysis.

1.2   Membrane protein LacY

This system, containing the E. Coli lactose permease LacY embedded in a bilayer of POPC molecules, was simulated using the CG Martini force field with the GROMACS package 4.0.3. Of the total 1470 POPC molecules 730 were situated in leaflet 1 (lower z) and 740 in leaflet 2 (higher z). The system was solvated with a total of 63440 Martini water beads. Analysis was carried out for data over an effective time of 4 μs. No flip flops were observed during this period.

The system was simulated using a time-step of 20 fs at a temperature of $T = 310\,\mathrm{K}$ on the weak-coupling (Berendsen) thermostat, and a pressure of $p = 1.0\,\mathrm{bar}$, using semi-isotropic coupling on the weak-coupling barostat. The coupling time for the thermostat was set to 0.5 ps, and the time constant for the barostat to 5 ps. For compressibility, we used a value of 3e-5 bar-1. Bond lengths in the simulation were constrained using the LINCS algorithm.

In order to gauge finite size effects, we carried two additional simulations that were smaller in size than the above described original one. One system consisted of 1 protein + 999 POPC molecules, and the other one of 1 protein + 96 POPC molecules. Both systems were simulated for 5 μs, of which 3 μs was used for analysis.

1.3   Membrane peptide WALP23

Simulations of the WALP23 [16, 17] peptide dimer were carried out in a dipalmitoylphosphatidylcholine (DPPC) lipid bilayer of 324 + 324 lipids, using an antiparallel arrangement of the peptides. The termini (Gly1 and Ala23) were not charged, consistent with the capped neutral termini used in experiments [18]. The peptides were embedded in the bilayers using the approach described by Kandt *et al.* [3].

The simulations were carried out using the MARTINI coarse-grained force field [19, 20, 21]. In this force field, each particle represents 4 non-hydrogen atoms, with the exception of ring-containing molecules, which are mapped with higher resolution (up to two non-hydrogen atoms per particle). Both electrostatic and Lennard-Jones interactions were calculated using a 1.2 nm cutoff with switch function; the distance to start switching Coulomb and Lennard-Jones interactions was 0 and 0.9 nm, respectively. The neighbor list was updated every 10 steps and the relative dielectric constant for the medium was set to 15. This is the standard procedure for the MARTINI force field [20]. Periodic boundary conditions

were applied in all dimensions. Simulations were carried out in the NpT ensemble using the Nose-Hoover thermostat [10, 11] ($T = 300$ K, $\tau = 1$ ps) and the Parrinello-Rahman barostat [8, 9] with semi-isotropic pressure coupling (pressure of 1 bar both in the plane of the membrane and perpendicular to the membrane; time constant of 4 ps). The integration time step was 40 fs (formal simulation time), and structures were saved every 100 ps for analysis. Production runs were carried out for 10 μs using GROMACS 4.0.516.

1.4   Two-dimensional LJ system

We simulated a simplified two dimensional system comprising one large and heavy "protein" particle surrounded by several small and light "lipid" particles. In the following description we use reduced units.

The mass of a lipid was 1 and the mass of the protein 150. The lipid-lipid interaction was described by a potential of the from

$$V(r) = 1/r^{12} - 2/r^6. \tag{1}$$

This potential has a strong repulsive interaction for $r < 1.0$ and an attractive tail for $r > 1.0$. The protein-lipid interaction potential was

$$V(r) = 1/(r - 6.38)^{12} - 2/(r - 6.38)^6, \tag{2}$$

such that for $r < 7.38$ there is repulsion and for $r > 7.38$ attraction. The depth of both potentials was 1, and they were taken to zero at $r = 8.25$. For the simulation of hard disks (HD), the lipid–lipid potential (1) was truncated and shifted to zero at $r = 1.0$ and the protein–lipid potential (2) at $r = 7.38$.

The simulations were done in the microcanonical ensemble (constant energy, volume (40x40) and number of particles), with average temperatures of 0.55 and 1.9 (LJ-system), and 0.45 and 1.8 (HD-system).

The simulations were done using Gromacs 4.0.4. No neighbor-list algorithm was applied, but all pair distances were measured at every time step. The starting structure was in all the cases a square lattice of lipids (lipid-lipid nearest-neighbor-distance 1.2), within which the protein was embedded such that all lipid-protein distances > 7.38. Each simulation comprised 10^6 time steps (leap-frog algorithm) of length 0.01. The particle positions were saved every 100 steps. The latter half of the simulation was used for analysis.

4

2 Analysis Details

Before diffusion analysis, the movement of the centre of mass (COM) of the membrane (lipids+protein) was always removed from the trajectories that were being analyzed. This means that all movements of the protein and lipids were measured with respect to the COM of the membrane. As the membrane may undergo relatively large lateral movements with respect to water during the simulation, it is important to remove this effect.

To qualitatively illustrate the local correlations of lipid and protein movements, we plotted the two-dimensional (2D) displacement vectors of the molecular COMs over specified time intervals. To quantify the effect of protein on lipid diffusion, the 2D displacement distributions of lipid COMs were plotted by first centering the protein into the middle of the box. The distributions were averaged over different fixed time intervals, Δt. Next, the radial averages of the 2D-plot were calculated, keeping protein COM in the origin of the coordinate system. Also, the radial and tangential components of the lipids were calculated separately. On the basis of this plot, two cutoffs were decided: $r < r_n$ defines lipids that are protein neighbors and most slowed down by protein, and $r > r_o$ defines lipids that are not affected by the protein.

To quantify lateral diffusion of different components, we calculated the 2D displacement distributions and fitted them to the expected 2D random-walk distribution:

$$P(r, \Delta t) = \frac{r}{2D\Delta t} \exp\left(-\frac{r^2}{4D\Delta t}\right), \tag{3}$$

where r is the lateral displacement, Δt the time period, and D is the lateral diffusion coefficient. The distributions were calculated separately for the COM of the protein, and for the COMs of lipids next to the protein (displacement within $r < r_n$) and other lipids (displacement within $r > r_o$). Examples of distributions are shown in Fig S2 and the diffusion coefficients at different time intervals in Fig S3. At large values of Δt, we should approach the diffusive regime, and the above approach should yield converging values of D.

Local correlations in the movements of lipids and the protein were analyzed in the spirit of our recent publication [22]. The average 2D displacement correlation plot ove a fixed period, Δt, was calculated by centering the displacement vector of the protein into the origin and rotating the whole system such that this vector points to the positive x-direction. Finally, the cosine of each lipid displacement vector with the protein displacement vector was binned in the xy-plane of the rotated coordinate system.

The rotation diffusion of the protein was addressed by calculating the rotated angle over a given time interval, $\Delta\alpha_i$, of each atom around an axis that is parallel to membrane normal (z-axis) and goes through the COM of the protein. The

contributions of the atoms were weighted by their moments of inertia to get the effective rotation of the whole protein:

$$\Delta\alpha_{\text{eff}} = \frac{\Sigma m_i r_i \Delta\alpha_i}{\Sigma m_i r_i},$$ (4)

where m_i is the mass of atom i, and r_i is the distance from the COM of the protein. The resulting displacement distribution of the effective angle can be expected to follow a one-dimensional gaussian distribution that describes the rotational diffusion of the protein.

Finally, the tangential component of lipid displacement vectors (r_t) were corrected by subtracting the theoretical component (r_α) that could be caused by the angular movement of the protein if the lipid moved together with protein surface. Both corrected and non-corrected plots are shown in Fig. S5 together with the correlation between uncorrected r_t and r_α components in Fig. S4.

The packing of lipids around the protein was analyzed by first binning the electron density of the lipid acyl chains into a 2D grid after centering the COM of the protein in the box. Second, the deuterium order paramteter S_{CD} of the carbons 5-7 along the acyl chains of lipids were binned to quantify the ordering of the acyl chains. Third, the sn-2 carbons of all lipids were used to determine the average surface of each monolayer, which were then used to obtain the 2D-plot of the thickness of the bilayer. These three plots are shown in Fig. S6.
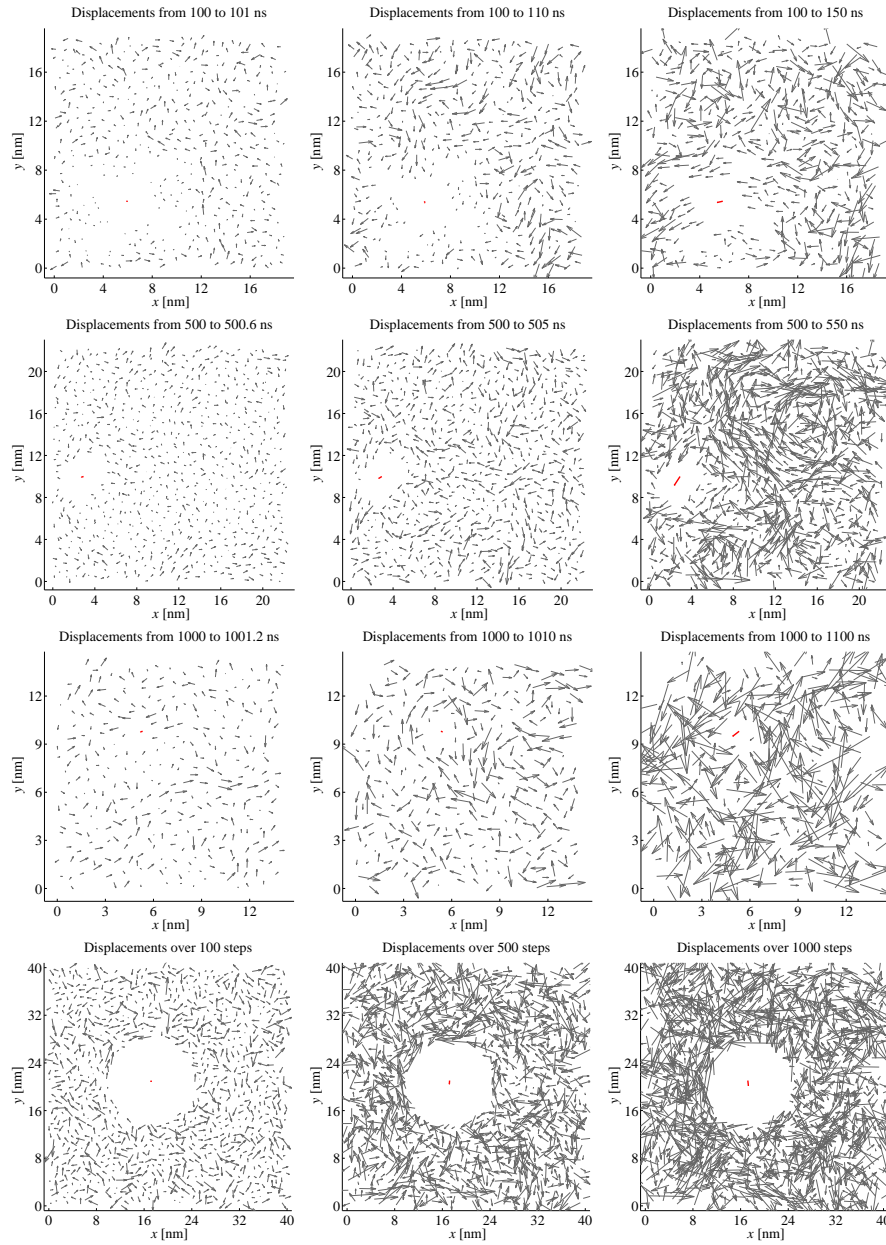
3 Additional Results

Figure S1: [All systems] Displacement vectors of lipids and protein (red) over different time intervals. The plots are shown for leaflet 2 in each of the shown systems: Kv1.2 (top row; intracellular leaflet), LacY (2nd row), WALP23 (3rd row), and 2D-LJ with temperature $T = 1.9$ (bottom row).
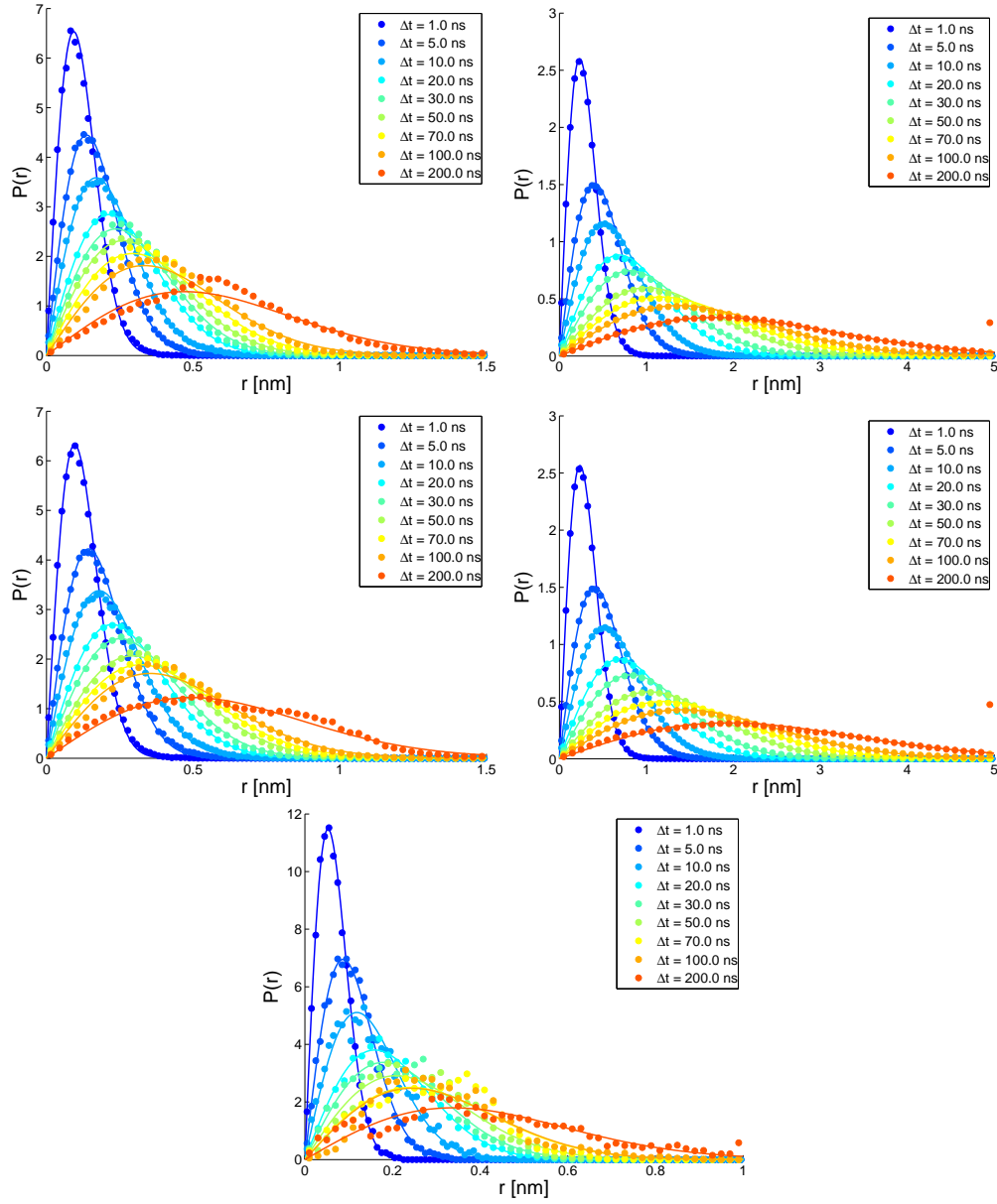
7

Figure S2: [Kv1.2] Measured displacement distributions together with fits to equation (3): lipids in extracellular leaflet (top), lipids in intracellular leaflet (middle), and protein (bottom). The lipids are divided into two groups: protein neighbors ($r < 3$ nm, left) and non-neighbors ($r > 7$ nm, right).
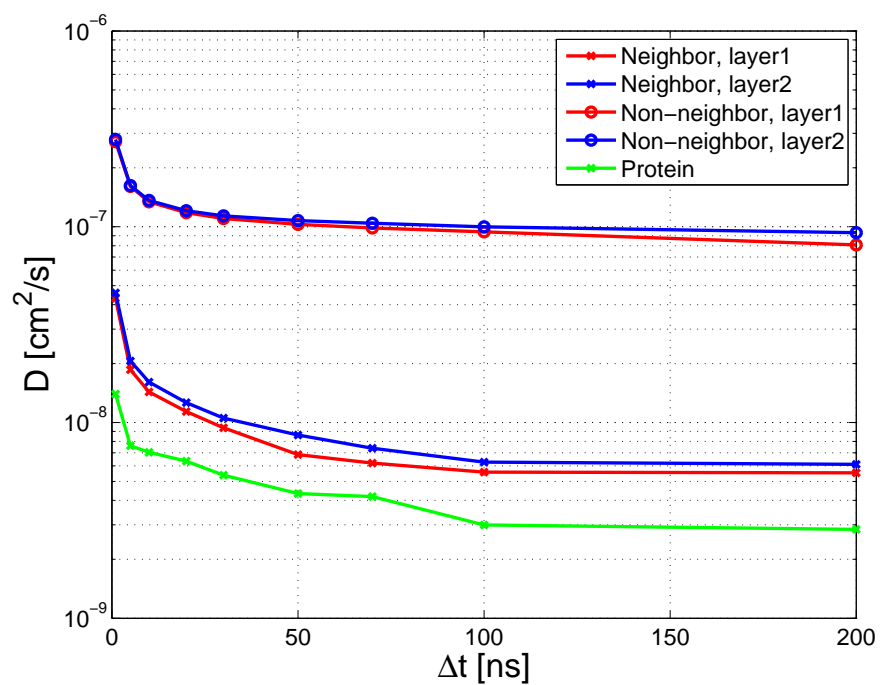
Figure S3: [Kv1.2] Diffusion coefficients from fitting 2-D gaussians to the displacement distributions. In the caption "layer 1" denotes extracellular leaflet and "layer 2" denotes intracellular leaflet.
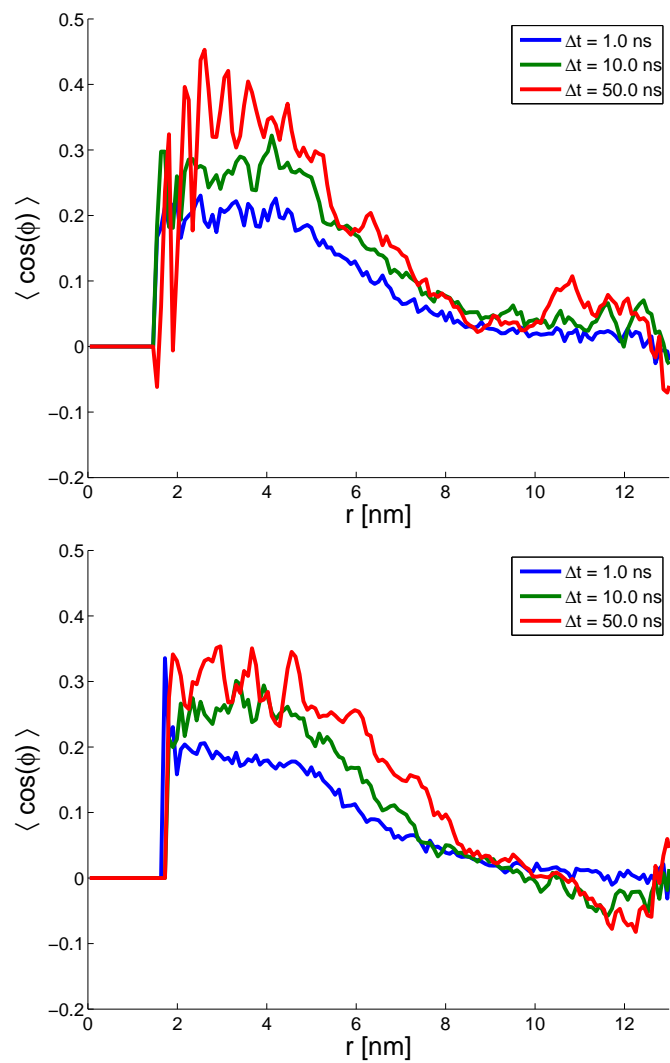
Figure S4: [Kv1.2] Correlation between the (uncorrected) tangential component of the lipid displacement and the theoretical tangential component caused by the angular movement of the protein. Panels correspond to extracellular (top) and intracellular (bottom) leaflets.
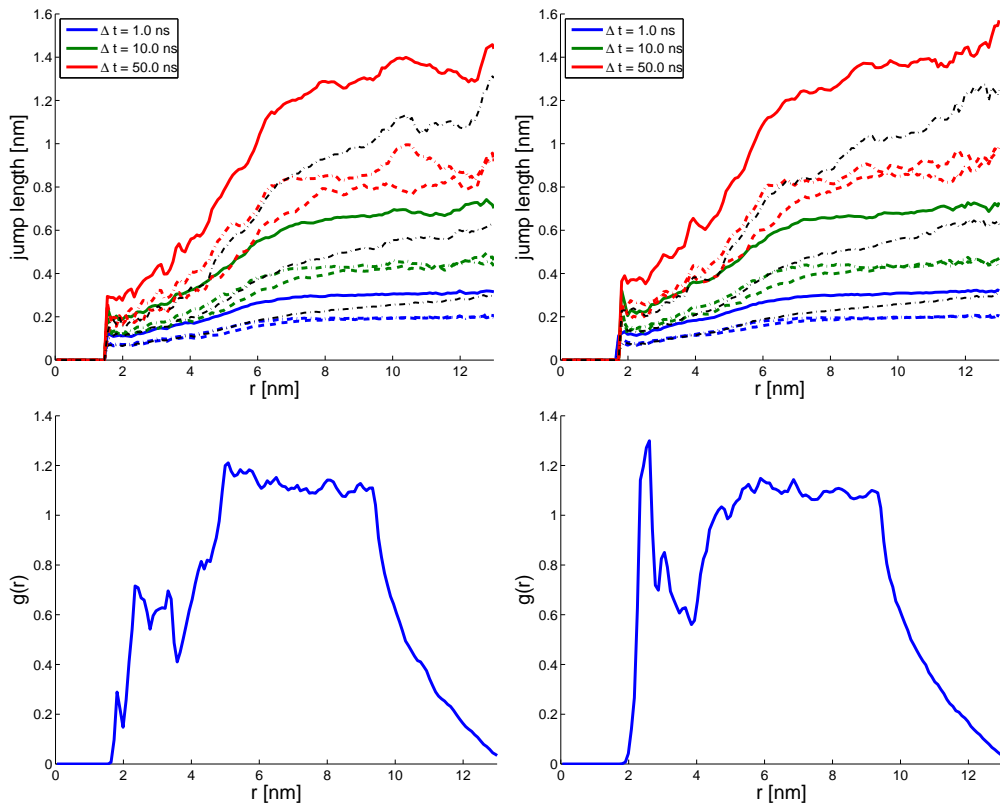
Figure S5: [Kv1.2] Average displacement lengths (solid) of lipids together with the radial (dashed) and tangential (dot-dashed) components in a coordinate system where the COM of protein is in the origin (top). The black lines are the corrected tangential components of lipids, with the estimated effect of protein subtracted. Note that at long distances, where lipid movements are uncorrelated with the rotation of the protein, this correction artificially increases the tangential component of lipids. The radial distribution functions of the lipids around the proteins have been also shown (bottom).
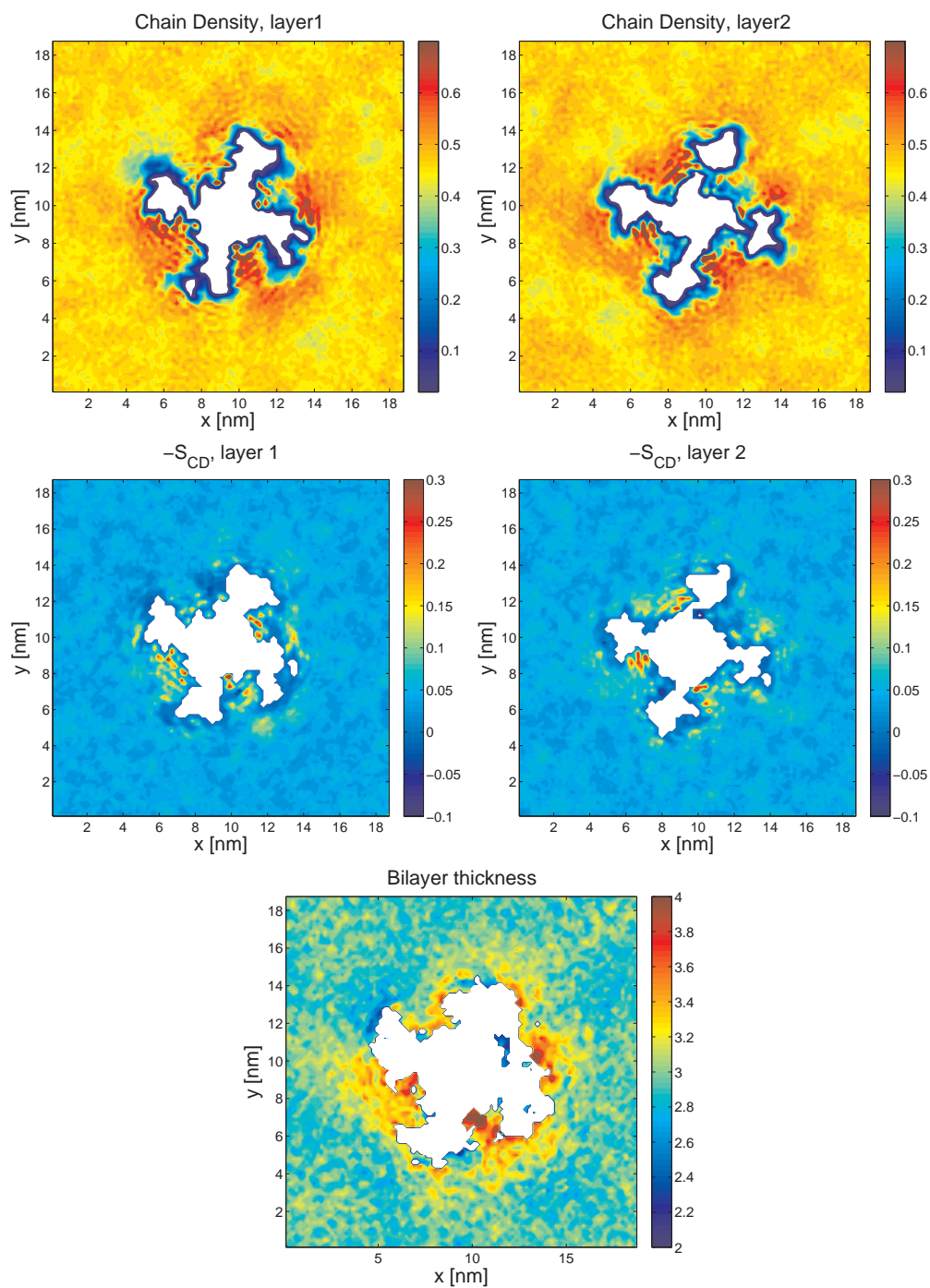
Figure S6: [Kv1.2] Two-dimensional electron densities of lipid acyl chains (top), deuterium order parameters of the acyl chains (middle), and the bilayer thickness (bottom). Extracellular leaflet is on the left panel, intracellular on the right.
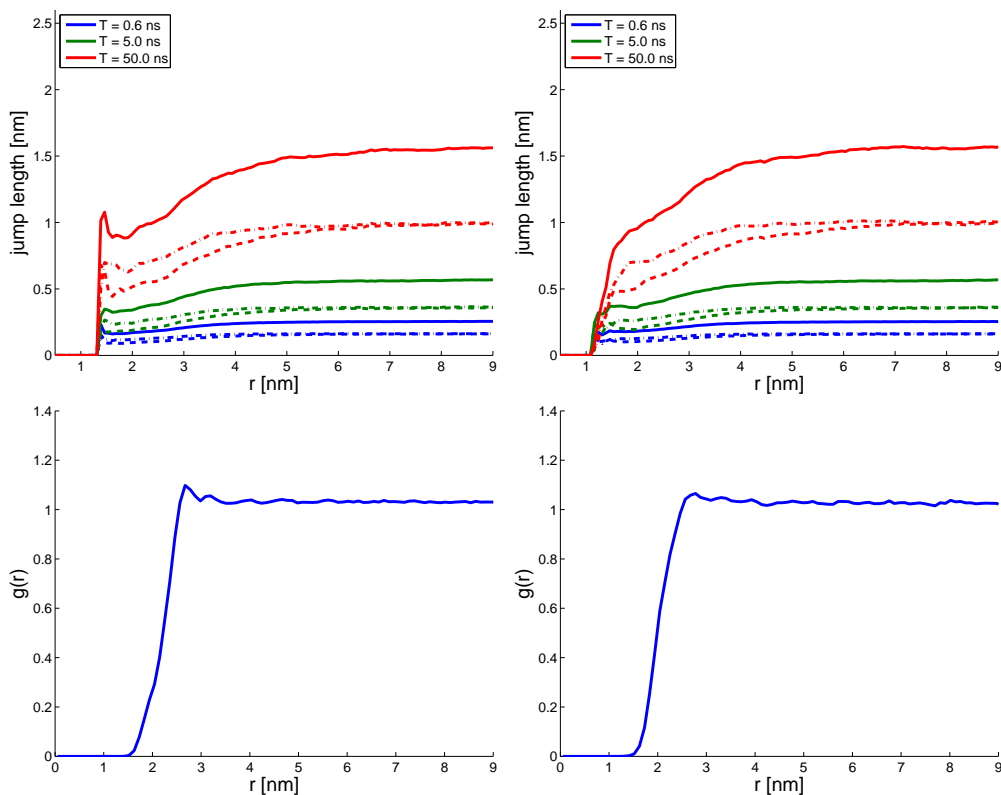
Figure S7: [LacY] Average displacement lengths (solid) of lipids together with the radial (dashed) and tangential (dot-dashed) components in a coordinate system where the COM of protein is in the origin (top). The radial distribution functions of the lipids around the proteins have been also shown (bottom). Left panel shows data for leaflet 1 and right panel for leaflet 2.
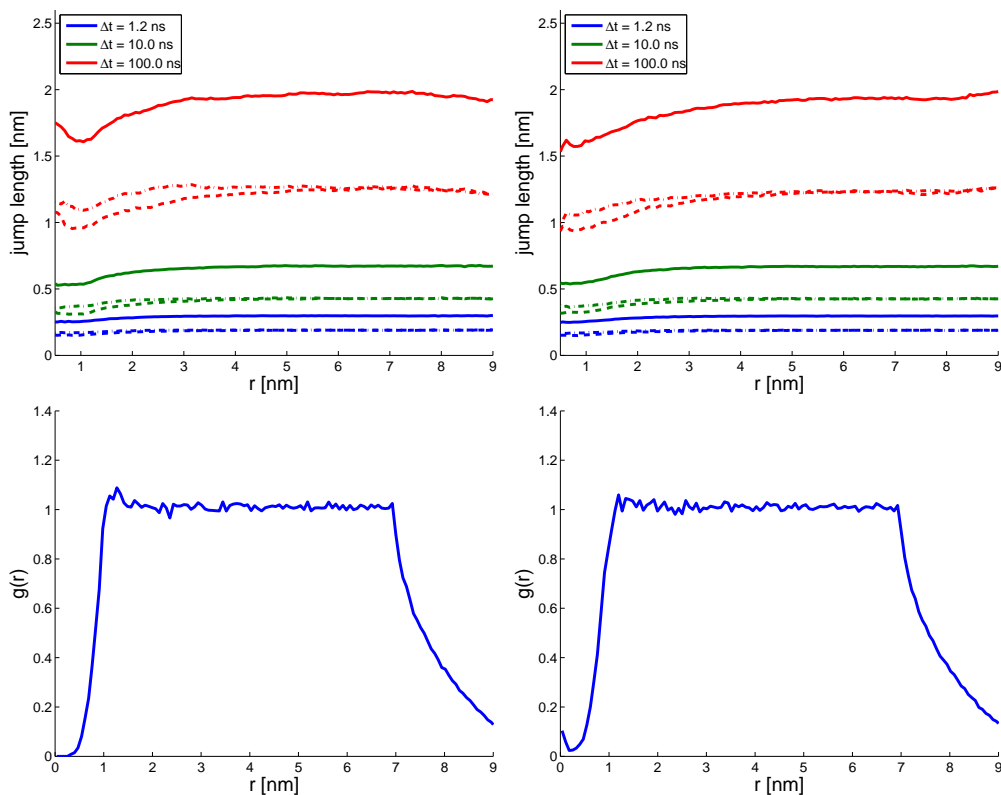
Figure S8: [WALP23] Average displacement lengths (solid) of lipids together with the radial (dashed) and tangential (dot-dashed) components in a coordinate system where the COM of protein is in the origin (top). The radial distribution functions of the lipids around the proteins have been also shown (bottom). Left panel shows data for leaflet 1 and right panel for leaflet 2.
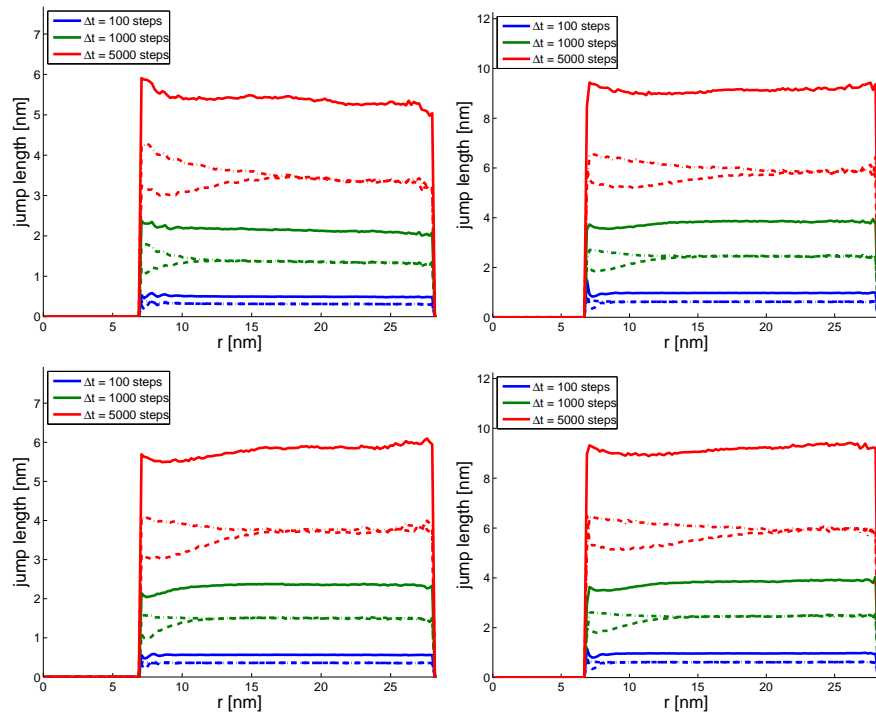
Figure S9: [2D-LJ] Average displacement lengths (solid) of lipids together with the radial (dashed) and tangential (dot-dashed) components in a coordinate system where the COM of protein is in the origin (top). The plots represent different temperatures of the LJ-system: 0.55 (top-left), 1.9 (top-right), and of the HD-system: 0.45 (bottom-left), and 1.8 (bottom-right).

4 Evaluation of Finite Size Effects
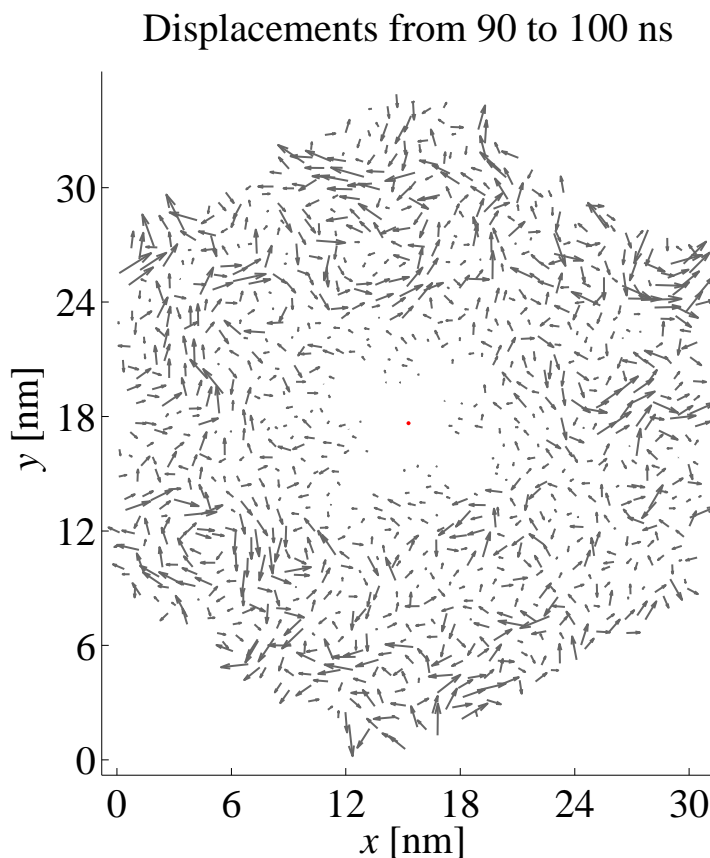
Displacements from 90 to 100 ns

Figure S10: [Kv1.2] Displacement vectors of lipids and protein (red) over one time interval, of leaflet 2 (intracellular) of an extended Kv1.2 simulation system. The extended system is otherwise identical to the original one, but contains 1 protein + 2220 POPC molecules (and hexagonal periodic boundary conditions). We find the same conclusions as in Figure 1 in the main article: about 100 lipids close to the protein are slowed down, and "bulk-like" diffusion of lipids is found about 10 nm from the COM of the protein.
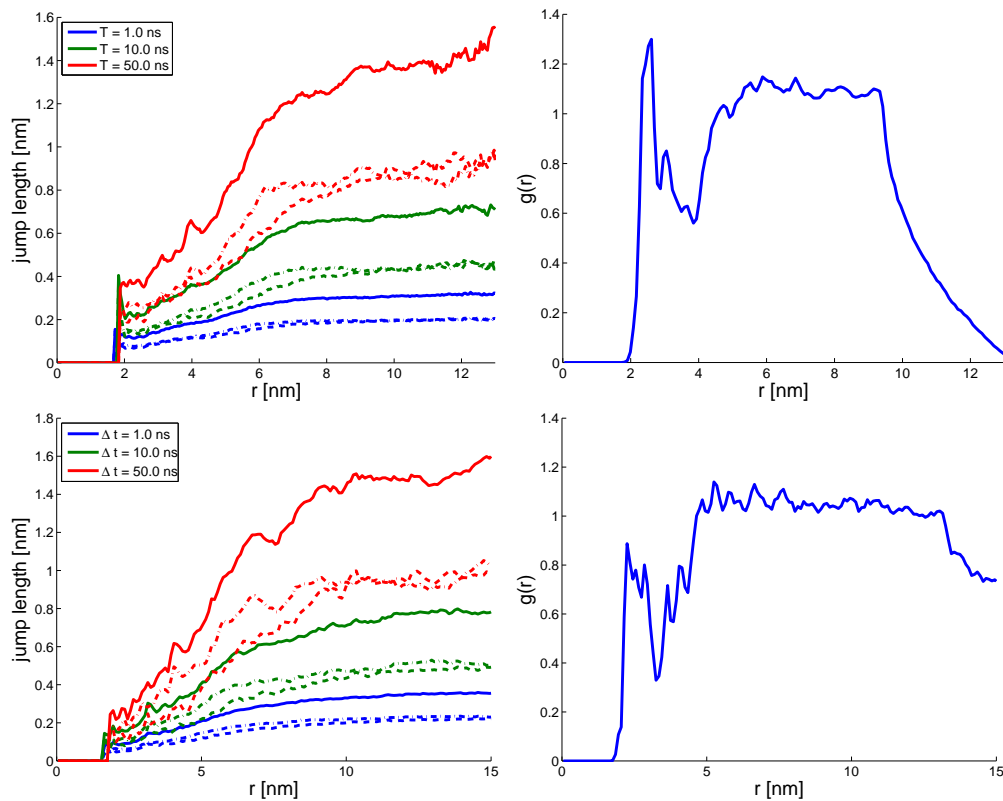
Figure S11: [Kv1.2] Average displacement lengths (solid) of lipids together with the radial (dashed) and tangential (dot-dashed) components in a coordinate system where the COM of protein is in the origin (left column). The radial distribution functions of the lipids around the proteins have been also shown (right column). All plots are shown for leaflet 2 (intracellular) of different system sizes: 1 protein + 910 POPC molecules (top row), and 1 protein + 2220 POPC molecules (bottom row).
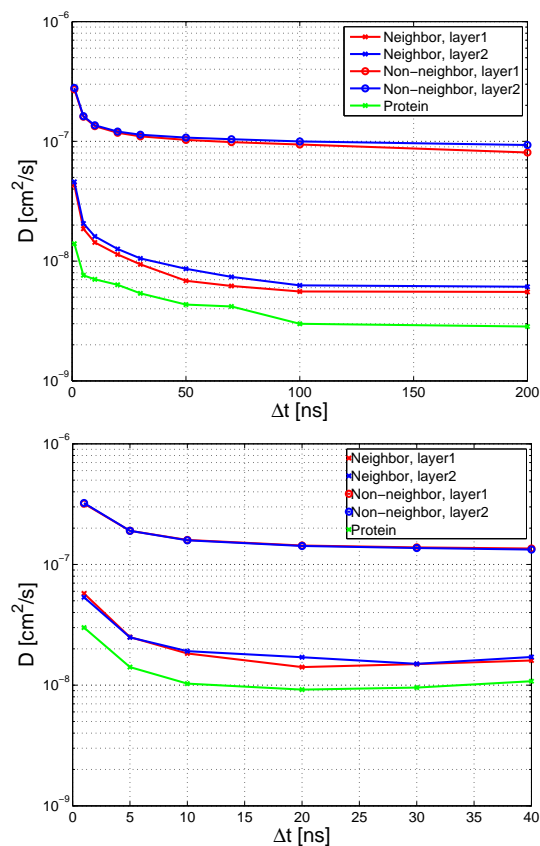
Figure S12: [KV1.2] Diffusion coefficients from fitting 2-D gaussians to the lateral displacement distributions. Plots are shown for systems of different sizes: 1 protein + 910 POPC molecules (top) and 1 protein + 2220 POPC molecules (bottom). Typical lateral diffusion coefficients of proteins under related conditions are of the order of 10^{-8} cm^2/s [Ramadurai et al., J Am Chem Soc 131, 12650-12656 (2009); Gambin et al., J Phys Chem B 114, 3559-3566 (2010)].
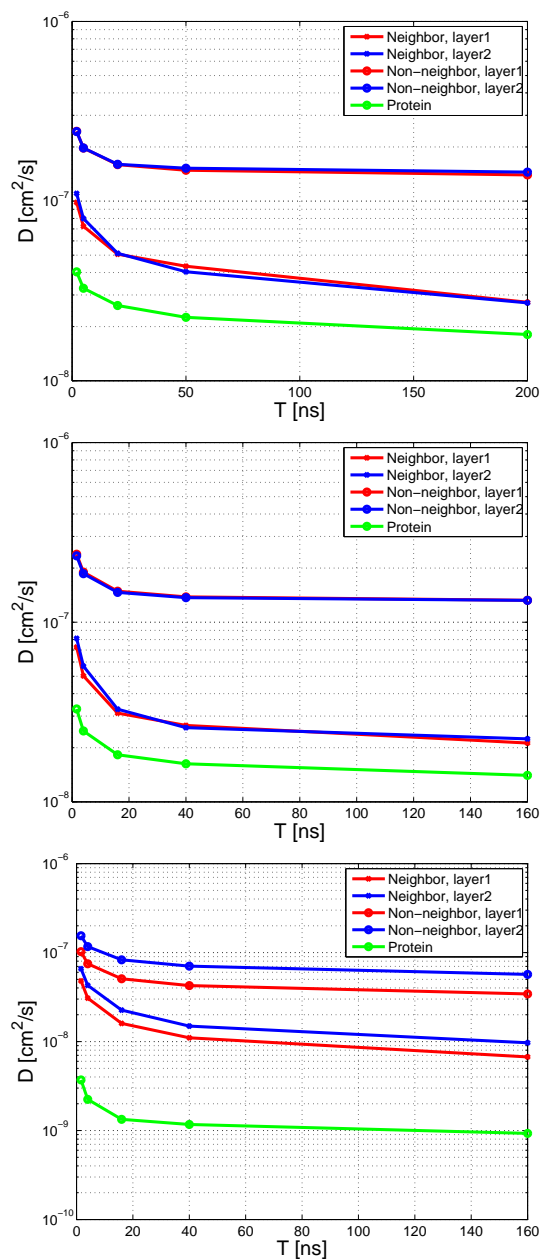
Figure S13: [LacY] Diffusion coefficients from fitting 2-D gaussians to the jump size distributions. Plots are shown for systems of different sizes: 1 protein + 1470 POPC (top), 1 protein + 999 POPC (middle), and 1 protein + 96 POPC (bottom). Lateral diffusion coefficient for LacY under fairly related conditions is about 4 x 10^{-8} cm^2/s [Ramadurai et al.,J Am Chem Soc 131, 12650-12656 (2009)].

References

[1] S. B. Long, E. B. Campbell, and R. MacKinnon, *Science*, **2005**, *309*, 897–903.

[2] P. Bjelkmar, P. S. Niemla, I. Vattulainen, and E. Lindahl, *PLoS Comput Biol*, **2009**, *5*.

[3] C. Kandt, W. Ash, and D. P. Tieleman, *Methods*, **2007**, *41*, 475–488.

[4] G. A. Kaminski, R. A Friesner, J. Tirado-Rives, and W. L. J. Jorgensen, *J. Phys. Chem. B*, **2001**, *105*, 6474–6487.

[5] O. Berger, O. Edholm, and F. Jähnig, *Biophys. J.*, **1997**, *72*, 2002–2013.

[6] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans; Interaction models for water in relation to protein hydration; In B. Pullman, Eds., *Intermolecular Forces*, pages 331–342. Reidel, Dordrecht, 1981.

[7] D. P. Tieleman, J. L. MacCallum, W. L. Ash, C. Kandt, Z. Xu, and L. Monticelli, *J. Phys.: Condensed Matter*, **2006**, *18*, S1221–S1234.

[8] M. Parrinello and A. Rahman, *J. Appl. Phys.*, **1981**, *52*, 7182–7190.

[9] S. Nose and M. L. Klein, *Mol. Phys.*, **1983**, *50*, 1055–1076.

[10] S. Nose, *Mol. Phys.*, **1984**, *52*, 255–268.

[11] W. G. Hoover, *Phys. Rev. A*, **1985**, *31*, 1695–1697.

[12] T. Darden, D. York, and L. Pedersen, *J. Chem. Phys*, **1993**, *98*, 10089–10092.

[13] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, *J. Comput. Chem.*, **1997**, *18*, 1463–1472.

[14] B. Hess, *J Chem Theory Comput*, **2008**, *4*, 116–122.

[15] S. Miyamoto and P. A. Kollman, *J Comput Chem*, **1992**, *13*, 952–962.

[16] M. R. R. de Planque, D. B. Greathouse, R. E. Koeppe, H. Schafer, D. Marsh, and J. A. Killian, *Biochemistry*, **1998**, *37*, 9333–9345.

[17] M. R. R. de Planque and J. A. Killian, *Mol. Membr. Biol.*, **2003**, *20*, 271–284.

[18] E. Strandberg, S. Ozdirekcan, D. T. S. Rijkers, P. C. A. van der Wel, R. E. Koeppe, R. M. J. Liskamp, and J. A. Killian, *Biophys. J.*, **2004**, *86*, 3709–3721.

[19] S. J. Marrink, A. H. de Vries, and A. E. Mark, *J. Phys. Chem. B*, **2004**, *108*, 750–760.

[20] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, *J. Phys. Chem. B*, **2007**, *111*, 7812–7824.

[21] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S. J. Marrink, *J. Chem. Theory Comput.*, **2008**, *4*, 819–834.

[22] T. Apajalahti, P. Niemela, P. N. Govindan, M. S. Miettinen, E. Salonen, S. J. Marrink, and I. Vattulainen, *Faraday Discuss.*, **2010**, *144*, 411–430.