PNAS
www.pnas.org

Supplementary Information for

Clustering and Dynamics of Crowded Proteins near Membranes and Their Influence on Membrane Bending

Grzegorz Nawrocki[1], Wonpil Im[2], Yuji Sugita[3,4,5], Michael Feig[1,5]*

[1]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, USA
[2]Department of Biological Sciences and Bioengineering Program, Lehigh University, Bethlehem, PA 18015, USA
[3]Cluster for Pioneering Research, RIKEN, Wako, Saitama 351-0198, Japan
[4]Center for Computational Research, RIKEN, Kobe, Hyogo 640-0047, Japan
[5]Center for Biosystems Dynamics Research, RIKEN, Kobe, Hyogo 640-0047, Japan

*Corresponding author.
Michael Feig
603 Wilson Road, Room 218 BCH
East Lansing, MI 48824, USA
E-mail: mfeiglab@gmail.com
Phone: 517-432-7439

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1910771116

Supplementary Methods

System Setup. The systems were assembled using the CHARMM-GUI framework and Membrane Builder (1-3). Initial protein structures were taken from the experimental structures deposited in the Protein Data Bank: 1VII for villin (4), 3GB1 for protein G (5), and 1UBQ for ubiquitin (6). The orientations of the proteins were randomized before assembling the crowded systems by randomly packing proteins inside the cubic (non-membrane) or rectangular (membrane) system volumes. The number of proteins and system sizes were adjusted to limit the total number of atoms to less than 400,000 atoms so that sufficiently long simulations could be carried out on the Anton2 hardware. Therefore, the number of protein copies was decreased from 10% to 5% volume fraction instead of increasing the system size (Table S1). Furthermore, the membrane systems were constructed under the constraint of keeping the number of lipids (and therefore the x-y box dimension) constant for all protein concentrations. Snapshots of the initial systems are shown in **Figure 1A**.

Setup of NAMD and Anton2 simulations. The initial step involved four cycles of minimization with 50 steps each (250 steps for membrane systems) using the steepest descent and adopted basis Newton-Raphson methods. Minimization was followed by simulations at 300 K (303.15 K for membrane systems) under restraints on heavy atoms (with a force constant of 1.0 kcal/mol/Å^2 for backbone atoms and 0.1 kcal/mol/Å^2 for non-backbone atoms). In the membrane systems, water molecules were also restrained from entering the hydrophobic core of the lipid bilayer and lipids were restrained to remain oriented with the head groups near the water phase and the tails inside the hydrophobic core. For non-membrane systems, 10 ps were initially simulated in the NVT ensemble with a 1-fs time step, followed by 20 ps with a 2-fs time step, and another 20 ps in the NPT ensemble. The equilibration of the membrane systems involved longer simulations of two 250 ps runs in the NVT ensemble with a 1-fs time step followed by another 250 ps in the NPT ensemble before the time step was increased to 2 fs and another three simulations were carried out over 700 ps with decreasing restraints on lipids and water molecules.

Production runs on Anton2 were begun after 5000 steps of minimization and 500 ps of equilibration without restraints using NAMD. The 5% simulations were started from the systems at 10% after the initial equilibration after replacing half of the proteins with water molecules and adjusting ion concentrations accordingly.

Additional details for NAMD and Anton simulations. Proteins were described by a modified version of the CHARMM36 force field (7), where protein-water Lennard-Jones interactions were increased by a factor of 1.09 to avoid aggregation artefacts as introduced previously (8). Lipid interactions were described by the CHARMM36 lipid force field (9), and explicit water was modeled with the CHARMM version of the TIP3P water model (10). Protein-lipid interactions were not altered based on a previous study that found good agreement with experiment for protein-membrane interactions with the (unmodified) CHARMM36 protein and lipid parameters (11). Initial ion parameters (12) were modified via NBFIX based on osmotic pressure corrections (13). Energies were matched between Anton2 and the Desmond software and between Desmond and CHARMM to ensure correct implementation of the NBFIX modifications and enhanced water scaling.

Changes of the box size were controlled isotropically in the x and y dimensions in the membrane systems and for all dimensions in the non-membrane simulations. A Berendsen thermostat and barostat (14) were used in the NAMD simulations with a thermal coupling constant of τ=1 ps and an isothermal compressibility β=0.0000457 bar^{-1}. A time step of 2 fs was used in the NAMD simulations in combination with SHAKE applied to all bonds involving hydrogen atoms. In the Anton2 simulations, integration was carried out via the "multigrator" algorithm (15) with a 2.5 ps time step applied to bonded and near-range non-bonded interactions.

Far-range interactions were evaluated every third time step based on a RESPA scheme. All bonds involving hydrogen atoms were constrained with M-SHAKE (16). Pressure was controlled according to a Martyna-Tobias-Klein (MTK) barostat (17) with an interval length of 1.2 ps and temperature was maintained via a Nose-Hoover thermostat (18) with an interval length of 60 fs. A relaxation time of τ=0.041667 ps was used for baro- and thermostats.

In NAMD simulations, Lennard-Jones interactions were truncated with a switching function that was effective from 10 to 12 Å. Particle-mesh Ewald summation was used to estimate long-range electrostatic interactions with a grid spacing of 1 Å and a spline interpolation order of 6. In Anton2 simulations, Lennard-Jones interactions were truncated at 9 Å and the u-series version of Gaussian-split Ewald summation was used to calculate long-range Electrostatic interactions (19). The direct space cutoff was optimized for accuracy and speed based on system size and varied between 9.12 to 12.65 Å.

Analysis. The results presented here were extracted from the Anton2 simulations. The first 100 ns of each trajectory were omitted from analysis as equilibration based on variations in contact formation at the beginning of the trajectories (see Figure S10). Statistical uncertainties were estimated from variations in the reported results between different protein copies (where possible) or from block averaging along the trajectory.

Protein contacts were determined based on two different criteria to facilitate comparisons with previous work: 1) Closest Cα-Cα distances <7 Å (8) and 2) closest heavy atom distances <2.7 Å (an additional criterion introduced when analyzing the detailed effect of protein contacts on the rotational diffusion of villin (20)). The parameters were chosen so that cluster-size distributions extracted from all-atom simulations based on protein contacts were similar independent of the chosen criterion. Clusters were determined from contacts based on any protein being in contact with at least one other protein in the cluster.

Translational and rotational diffusion coefficients were evaluated as described previously (8). To describe translational diffusion, the mean square displacement (MSD) was calculated as a function of time from $\mathrm{MSD}(\boldsymbol{\tau}) =< (\mathbf{r}(\boldsymbol{t_0}+\boldsymbol{\tau}) - \mathrm{r}(\boldsymbol{t_0}))^2 >_{\boldsymbol{t_0}}$ based on the center of mass of a given protein after unwrapping coordinates due to periodic boundary conditions, r(t), and averaged over different initial times t_0 along the trajectory. From the slope of the linear fits over a given time interval $\Delta\tau$, $s(\Delta\tau)$, translational diffusion coefficients were then obtained according to the Einstein relationship: $\boldsymbol{D_t = s(\Delta\tau)/(2N\Delta\tau)}$ where N is the dimension ('3' for 3D diffusion, '2' for diffusion in the x-y plane, and '1' for diffusion along the membrane normal z). In order to analyze transient and anomalous behavior of D_t, we considered time intervals $\Delta\tau$ of 0 – 1 ns, 1 ns – 10 ns, and 10 ns – 100 ns. Translational diffusion coefficients were estimated separately for each protein and as a function of distance from the membrane based only on the position of the protein at time t_0, irrespective of where it may diffuse afterwards.

Estimates of translational diffusion are subject to finite-size effects in the presence of periodic boundary conditions. To obtain infinite-size values, we added the correction term given in Eq. 1 (21):

$$D_{t,PBCcorr} = \frac{k_B T}{6\pi\eta L}\left(\xi - \frac{4\pi R_h^2}{3L^2}\right) \qquad (1)$$

where $\xi = 2.837$, k_B is the Boltzmann constant, T the temperature of the system, L is the length of the cubic simulation box, η is the shear viscosity of the solvent, and R_h is the hydrodynamic radius of a given molecule, estimated with HYDROPRO (22), as R_h(villin) = 9 Å, R_h(protein G) = 11 Å, and R_h(ubiquitin) = 12 Å.

For crowded systems, the viscosity was further adjusted from the viscosity of pure solvent, η_w. Instead of a simple hard-sphere based estimate used earlier (8, 23) for how the effective viscosity varies with the protein volume fraction, φ, we used here the expanded formalism in Eq. 2 that was introduced recently (24):

$$\eta = \eta_w(1 + 2.5\varphi + b\varphi^2) \qquad (2)$$

where the parameter b catpures effective increases in viscosity not just due to volume exclusion but also as a result of increased clustering (8, 24). Von Bülow et al. determined different values of b for different proteins depending on their propensity to form clusters (24), but because similar values were found for villin, protein G, and ubiquitin, the three proteins simulated here as well in a mixture, we used an average value of 58.2.

In the presence of the membrane, there are additional confinement effects to consider. Following the analysis by Simonnin et al. (25), the finite size correction for translational diffusion parallel to the membrane in a fluid that is infinite in x-y directions but constrained in the z direction by a membrane is:

$$D_{t,\parallel,PBCcorr} = \frac{k_B T}{\eta}\left(\frac{3\ln(1+\sqrt{2})}{4\pi L} - \frac{3H}{40L^2}\right) \qquad (3)$$

where L is the box size in x and y directions, and H is the width of the fluid slab that is calculated as the box length in the z direction minus the width of the membrane layer (2 x 27 Å = 54 Å). Other parameters were set as in Eq. 1.

The diffusion parallel to the membrane in a slab of width H for infinite box dimensions in the x-y directions that is obtained after correction based on Eq. 3 can then be compared to bulk diffusion in the absence of a constraining slab (i.e., H=∞) according to Eq. 4 (25):

$$D_{\parallel}(H, L=\infty) = D_{bulk}\left(1 + \frac{9R_H}{8H}\ln\left(\frac{R_H}{H}\right)\right) \qquad (4)$$

Rotational diffusion coefficients were estimated following the method introduced by Wong and Case (26), where randomly distributed unit vectors are rotated along with the protein and a correlation function is obtained based on the rotation of the vectors. The correlation functions up to 100 ns (up to 24 ns for z-dependent analysis of rotational diffusion) were then fitted with double-exponential functions $\boldsymbol{C_O(t) = S_R^2 e^{-\frac{t}{\tau_{Rs}}} + (1 - S_R^2) e^{-\frac{t}{\tau_{Rf}}}}$ to obtain slow and fast correlation times, τ_{Rf} and τ_{Rs}, weighted by S_R^2. An overall relaxation time τ was determined according to $\boldsymbol{\tau = \left(\frac{S_R^2}{\tau_{Rs}} + \frac{1-S_R^2}{\tau_{Rf}}\right)^{-1}}$ and the rotational diffusion coefficient D_r was calculated as D_r=1/6τ. Because rotational correlation functions converge slowly for a single protein, we averaged correlation functions from multiple proteins before fitting the exponential functions. Statistical uncertainties were estimated from comparing results obtained with different subsets of proteins. We did not correct the rotational diffusion estimates for periodic boundary conditions (27) because there is little change for the large systems studied here.

The MMTSB Tool Set (28), analysis functions in CHARMM (29), and custom-written programs in C/C++ and perl were used for all of the analysis. VMD (30) and gnuplot (http://www.gnuplot.info) were used for visualization and plotting. Gnuplot was also used for fitting linear and exponential functions for the determination of diffusion coefficients.

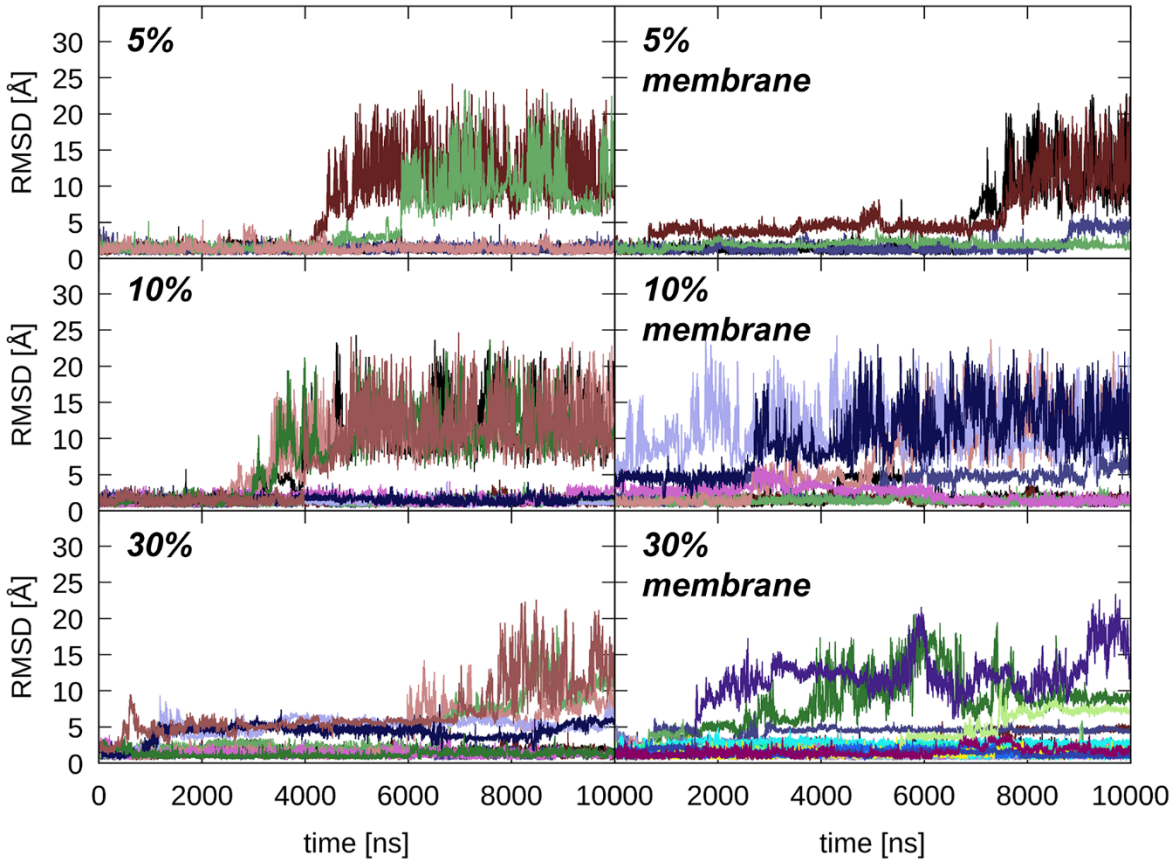

Figure S1: Root mean square deviation (RMSD) of villin molecules for Cα atoms after optimal structural superposition compared to the experimental reference structure (PDB code: 1WY3 (31)) as a function of simulation time in the Anton2 simulations. Different colors indicate different molecules.

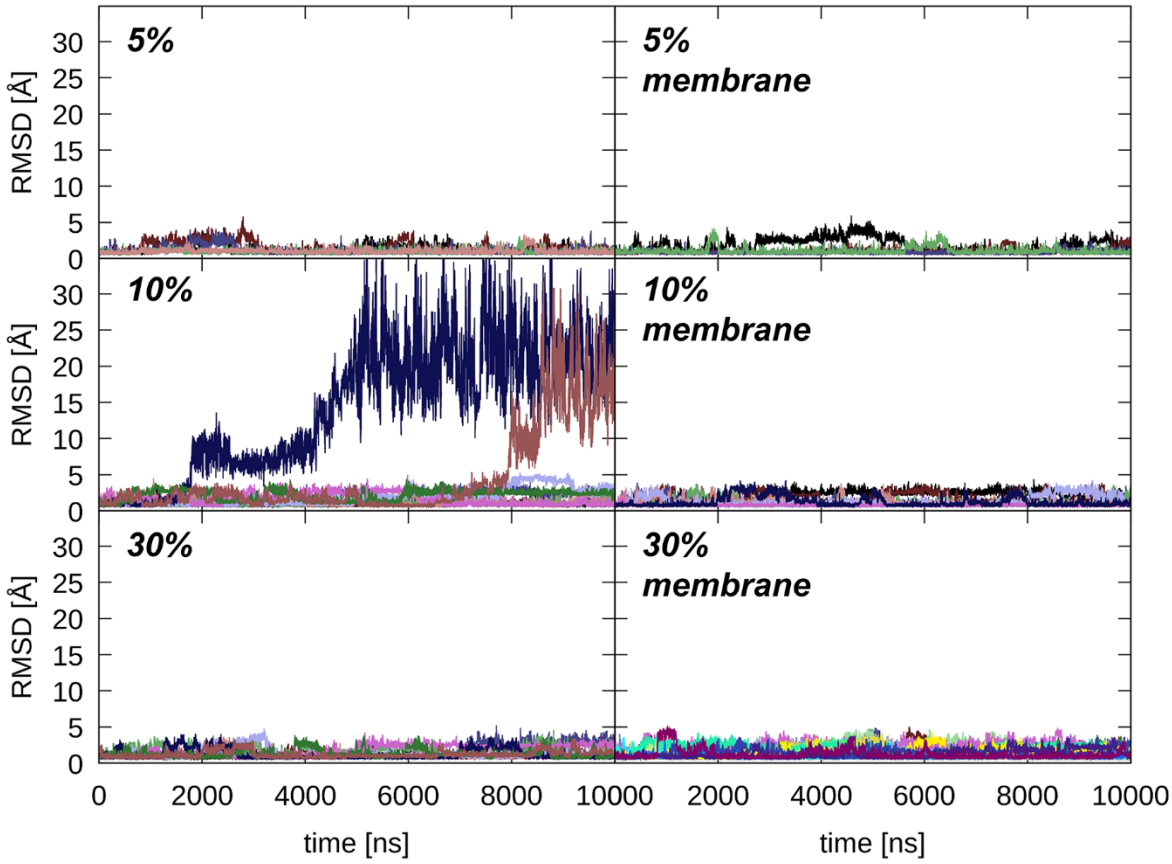

Figure S2: RMSD of protein G molecules for Cα atoms after optimal structural superposition compared to the experimental reference structure (PDB code: 3GB1 (5)) as a function of simulation time in the Anton2 simulations. Different colors indicate different molecules.

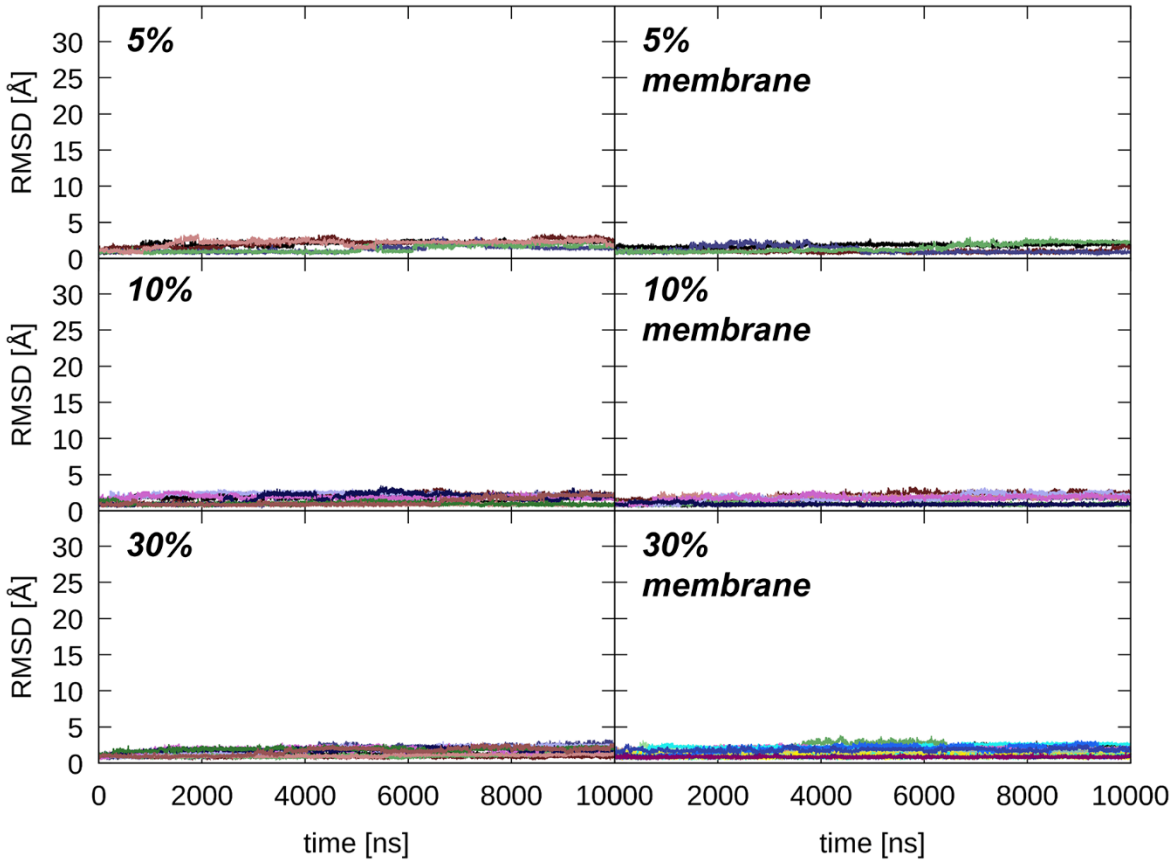

Figure S3: RMSD of ubiquitin molecules for Cα atoms after optimal structural superposition compared to the experimental reference structure (PDB code: 1UBQ (6)) as a function of simulation time in the Anton2 simulations. Only residues 1-72 were considered to exclude the flexible C-terminus. Different colors indicate different molecules.

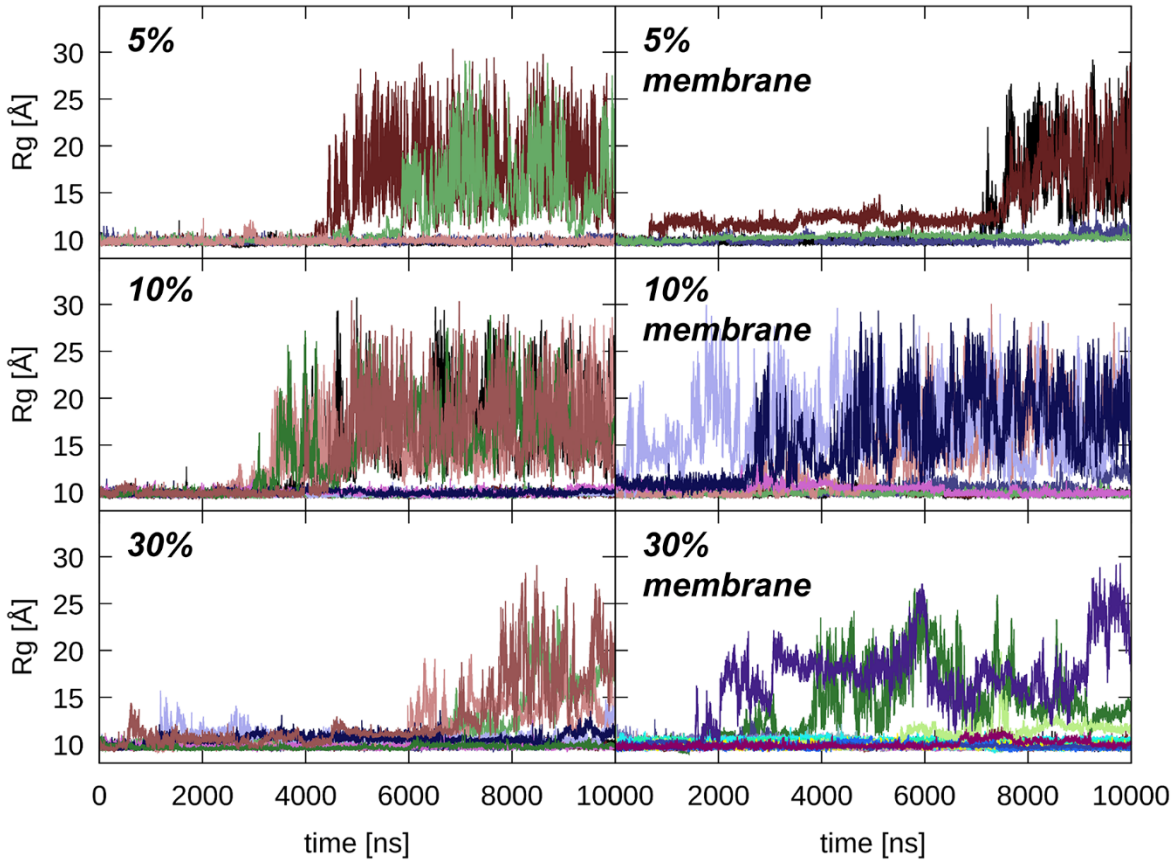

Figure S4: Radius of gyration R_g of villin molecules based on Cα atoms as a function of simulation time in the Anton2 simulations. Different colors indicate different molecules.

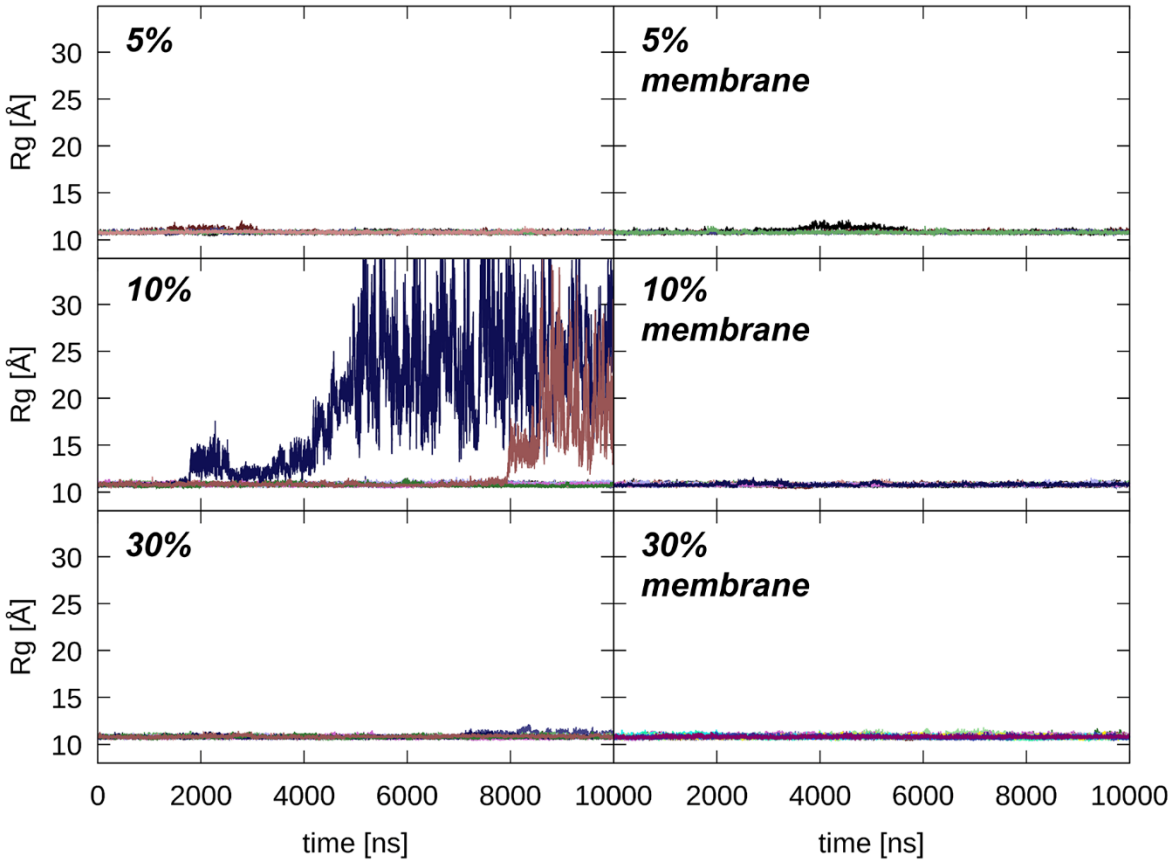

Figure S5: R_g of protein G molecules based on Cα atoms as a function of simulation time in the Anton2 simulations. Different colors indicate different molecules.

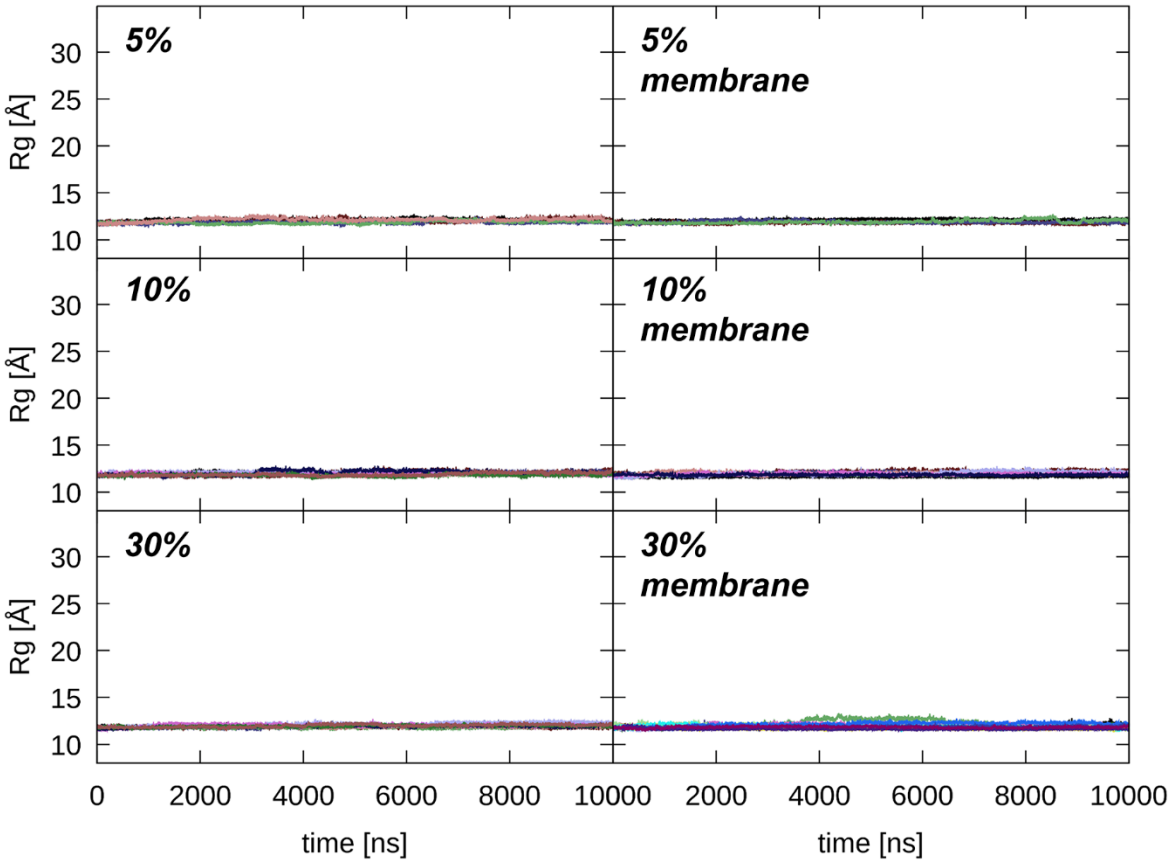

Figure S6: R_g of ubiquitin molecules based on Cα atoms as a function of simulation time in the Anton2 simulations. Different colors indicate different molecules.

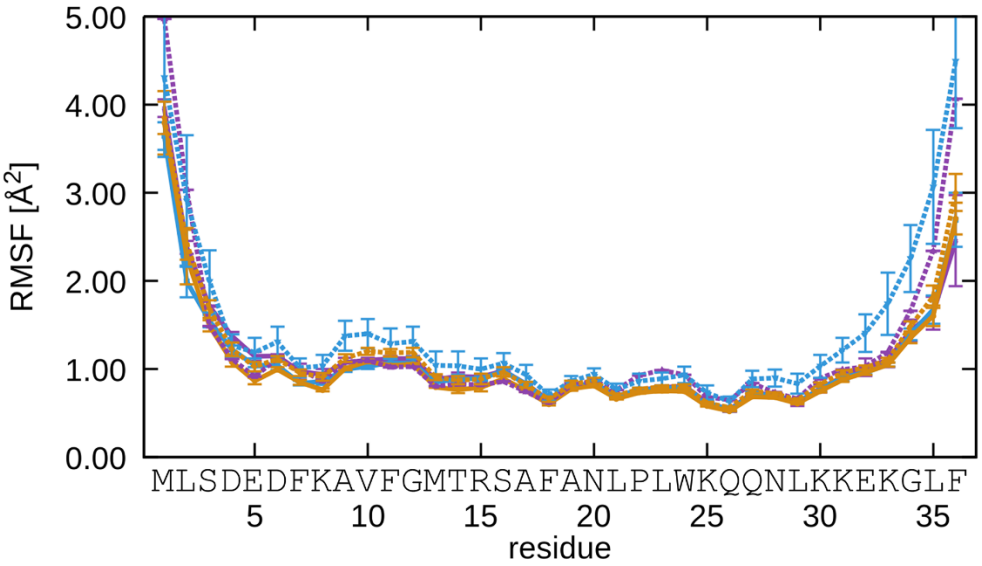

Figure S7: Root mean square fluctuations (RMSF) for Cα atoms in villin molecules at 5% (purple), 10% (light blue), and 30% (tan) protein concentration in the absence (solid lines) and presence of the membrane (dashed lines). Error bars reflect uncertainties obtained from variations between different protein molecules in a given system. Only protein molecules where the Cα RMSD at the end of the trajectory was below 2.5 Å were included.

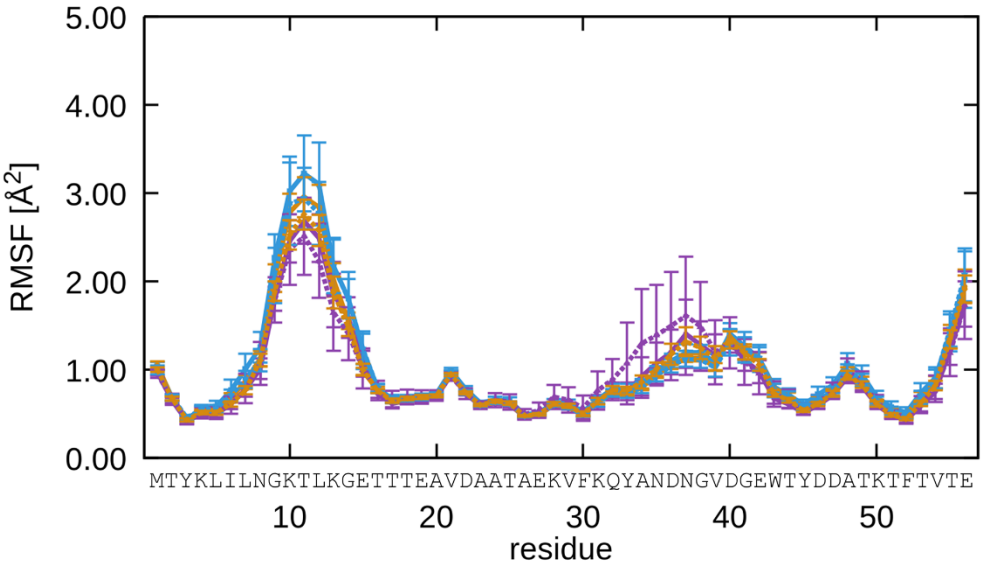

Figure S8: Root mean square fluctuations (RMSF) for Cα atoms in protein G molecules at 5% (purple), 10% (light blue), and 30% (tan) protein concentration in the absence (solid lines) and presence of the membrane (dashed lines) as in Figure S7

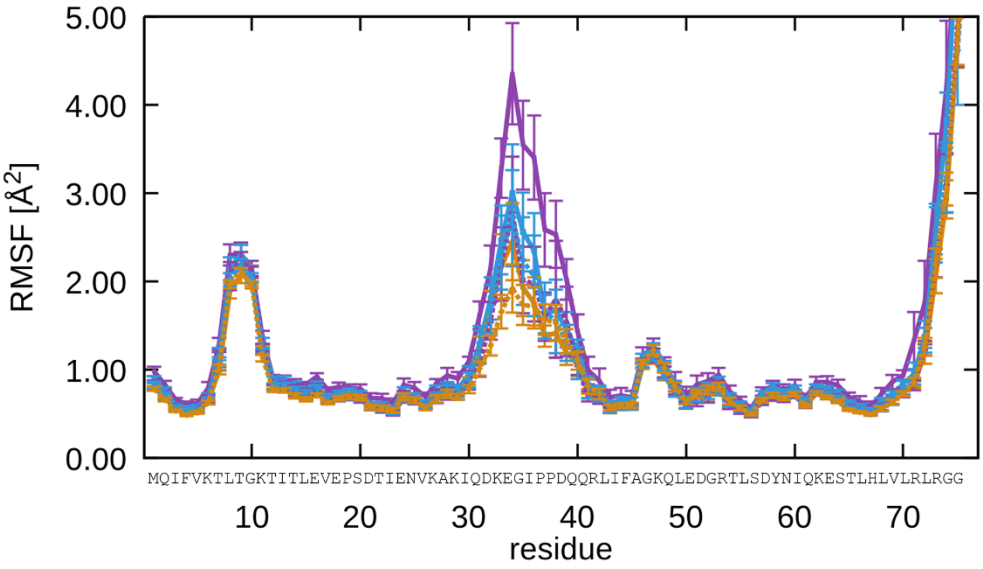

Figure S9: RMSF for Cα atoms in ubiquitin molecules at 5% (purple), 10% (light blue), and 30% (tan) protein concentration in the absence (solid lines) and presence of the membrane (dashed lines) as in Figure S7.

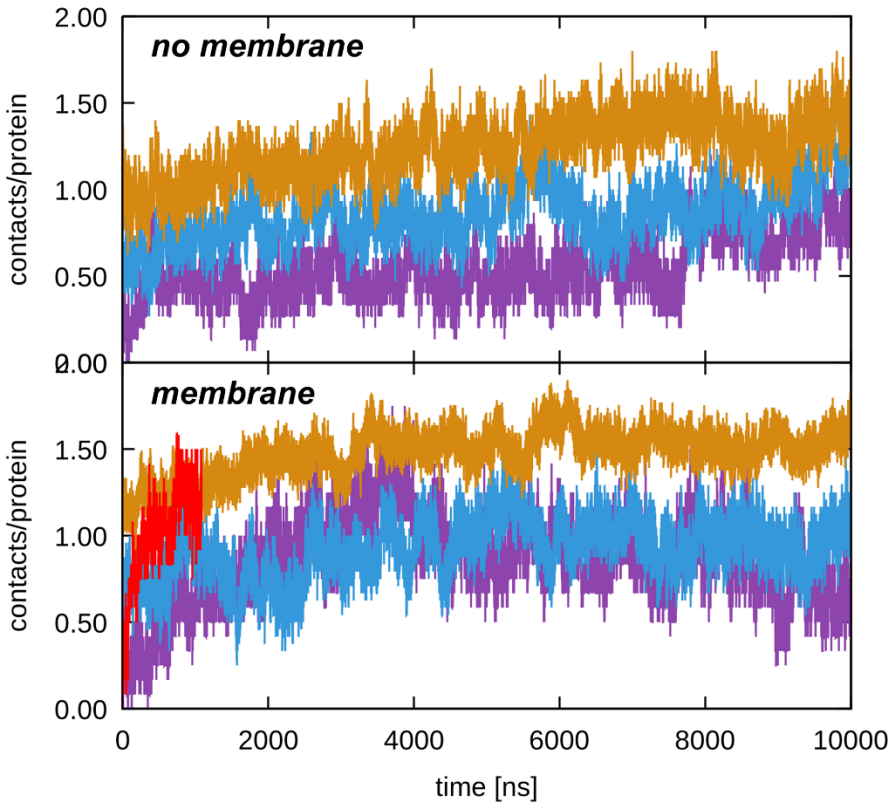

Figure S10: Protein contacts normalized by number of proteins as a function of simulation time at 5% (purple), 10% (light blue), and 30% (tan) protein concentration in the absence (top) and presence of the membrane (bottom). Results from a simulation at 5% with the original CHARMM c36 force field without scaling protein-water interactions are shown in red. Contacts were defined as minimum Cα distances of less than 7 Å.

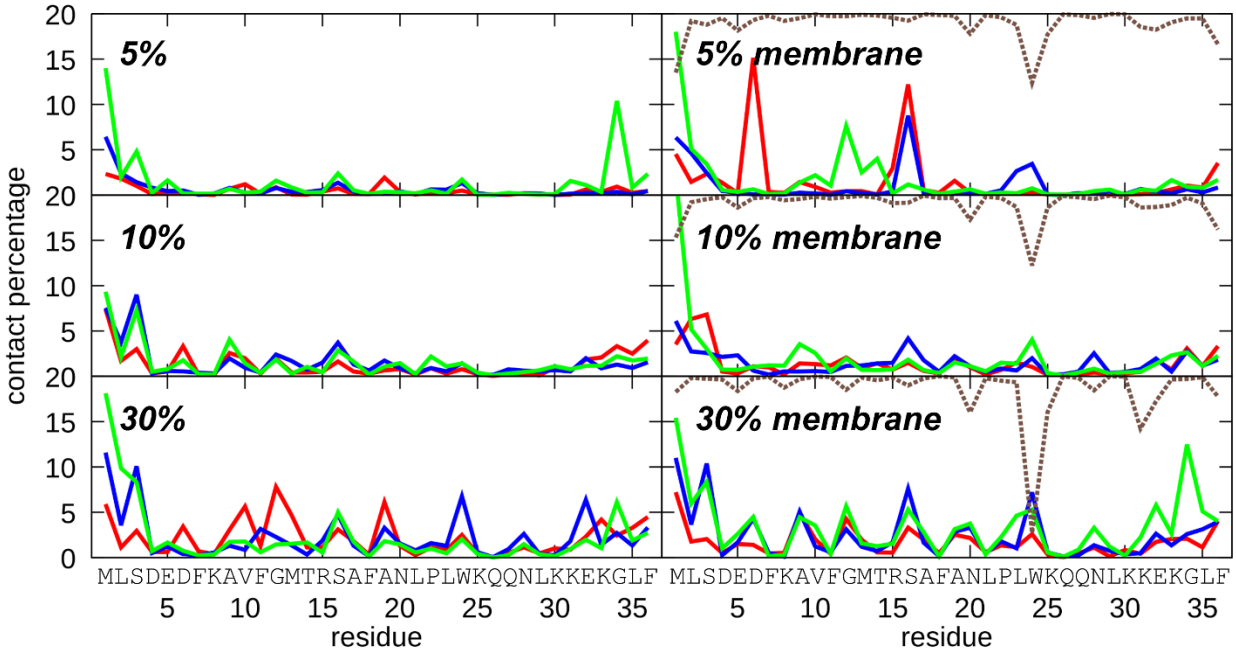

Figure S11: Percentage of residues (x10) involved in minimum heavy-atom distances below 5 Å between villin and other villin (red), protein G (blue), ubiquitin (green), or the membrane (black dashed line, shown as 20-10*percentage).

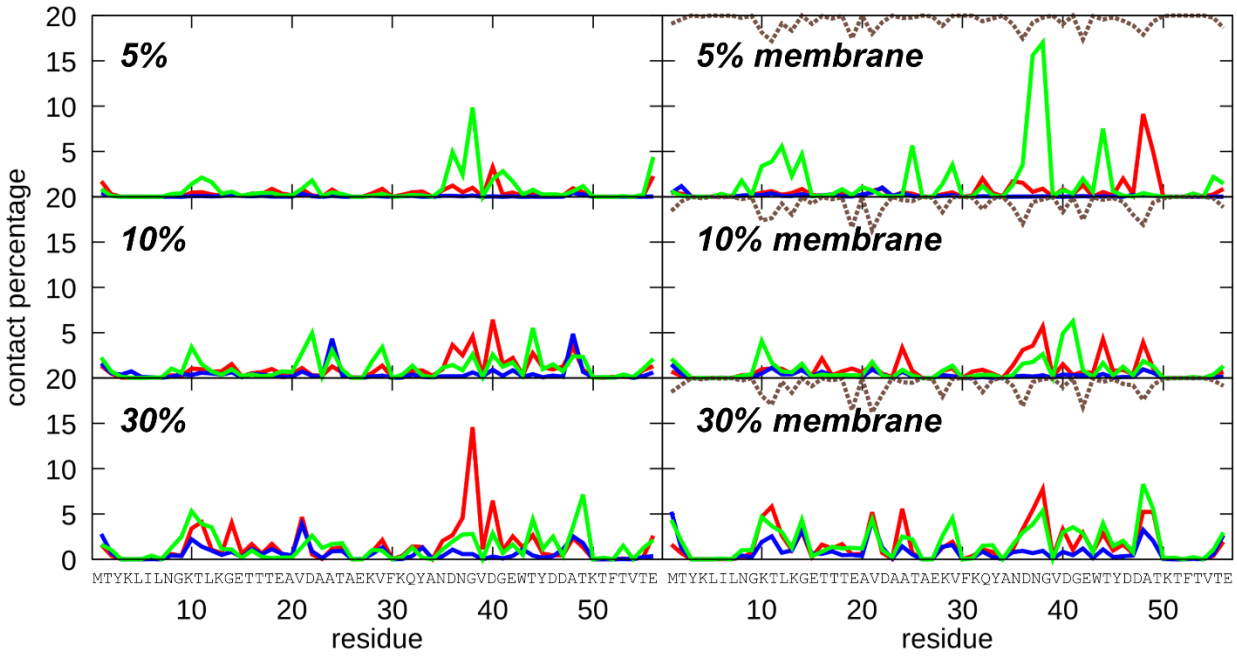

Figure S12: Percentage of residues (x10) involved in minimum heavy-atom distances below 5 Å between protein G and villin (red), other protein G (blue), ubiquitin (green), or the membrane (black dashed line, shown as 20-10*percentage).

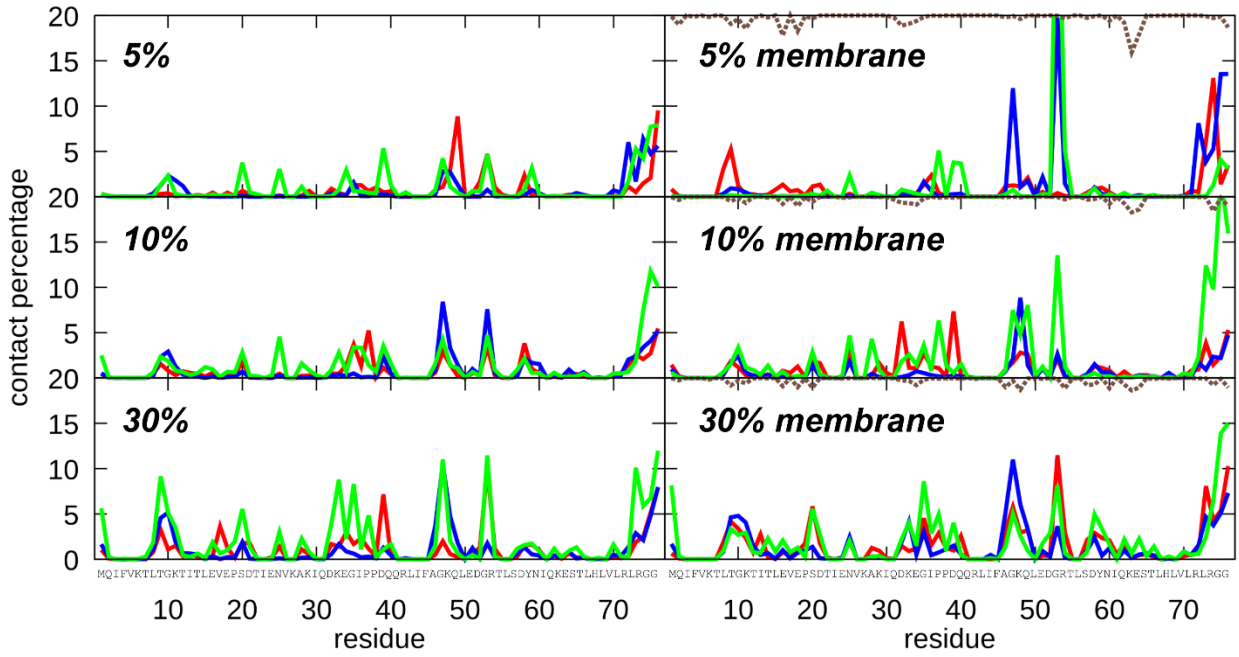

Figure S13: Percentage of residues (x10) involved in minimum heavy-atom distances below 5 Å between ubiquitin and villin (red), protein G (blue), other ubiquitin (green), or the membrane (black dashed line, shown as 20-10*percentage).

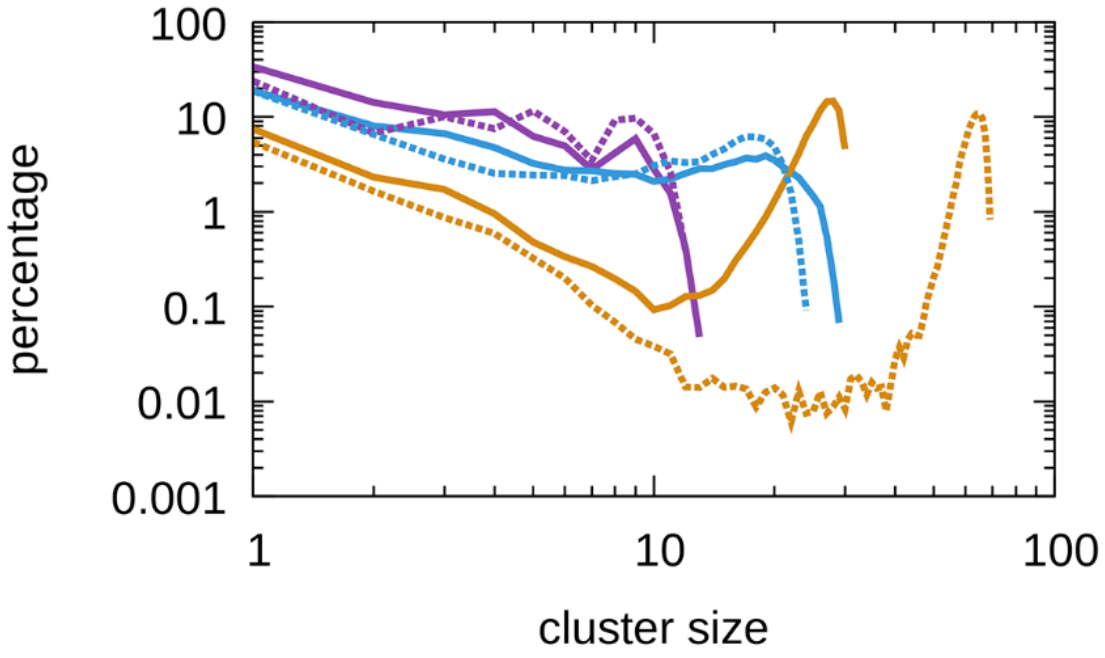

Figure S14: Cluster size distributions between all proteins at 5% (purple), 10% (light blue), and 30% (tan) in the absence (solid lines) and presence (dashed lines) of the lipid bilayer based on protein contacts defined as minimum heavy atom distances of less than 2.7 Å.

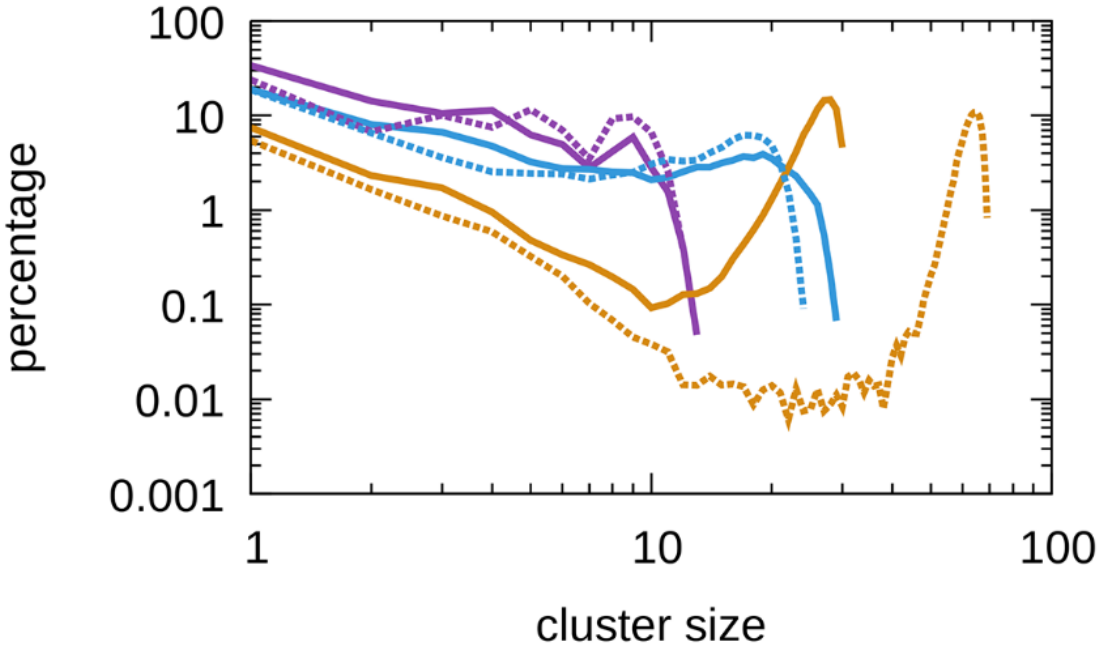

Figure S15: Cluster size distributions between all proteins at 5% (purple), 10% (light blue), and 30% (tan) in the absence (solid lines) and presence (dashed lines) of the lipid bilayer based on protein contacts defined as minimum Cα-Cα atom distances of less than 7 Å as in Figure 1B but using only data from the first 2 μs in order to exclude frames with partially unfolded proteins.

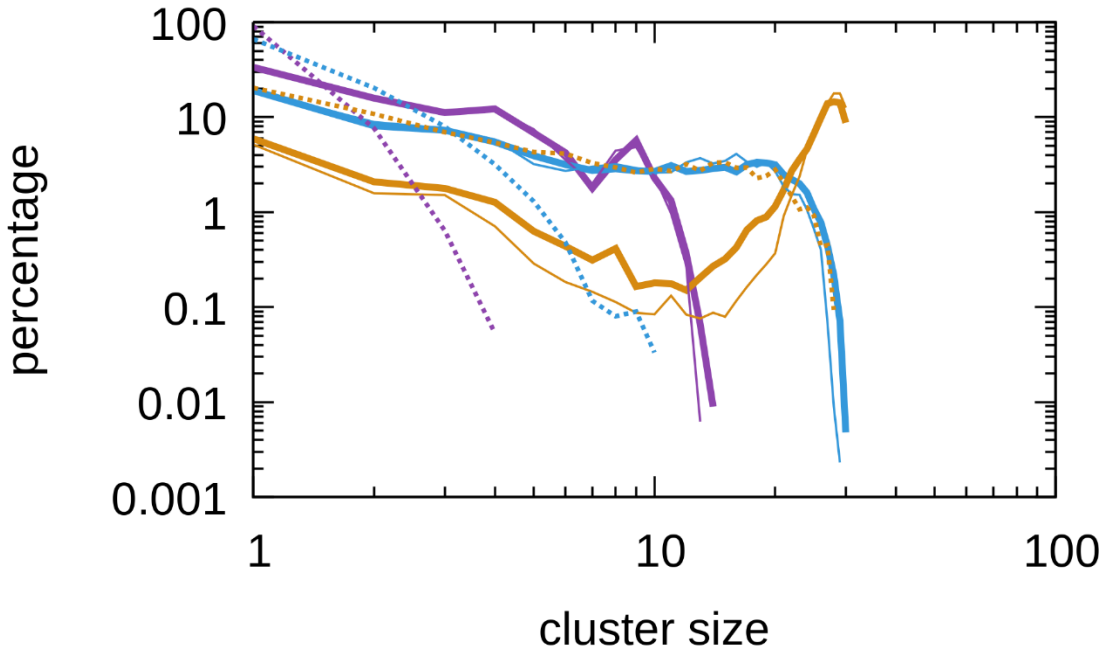

Figure S16: Cluster size distributions between all proteins in the absence of a membrane surface at 5% (purple), 10% (light blue), and 30% (tan) based on all-atom simulations using a Cα-Cα distance criterion of 7 Å (thick solid lines), based on all-atom simulations using an equivalent center-of-mass based criterion (thin solid lines), and based on coarse-grained (CG) simulations using the same center-of-mass based criterion (dashed lines). According to the center-of-mass criterion two proteins A and B were considered in contact when the distance between their centers was less than $(\sigma_A + \sigma_B)/2 + 7\text{Å}$ with values for σ given below. The CG simulations involved spherical models with 10 particles for each protein type at box volumes that were adjusted to result in the same volume fractions as in the atomistic simulations. The interaction potential for the CG model consisted of a purely repulsive short-range Lennard-Jones type potential as in the work by Mani et al. (32): $U_{ij}(r_{ij}) = 4\varepsilon\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{100}$. The value of ε was set to 4 kJ/mol for protein-protein interactions. The size-dependent parameter σ_{ij} was calculated as $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ with σ_i determined from the diameter of a sphere with a volume equivalent to the molecular volume of molecule i. Specifically, we set σ_{villin}=18.13 Å, $\sigma_{proteinG}$=20.67 Å, and $\sigma_{ubiquitin}$=23.34 Å. The CG simulations were run with OpenMM for 1 μs at 298 K using a Langevin thermostat.

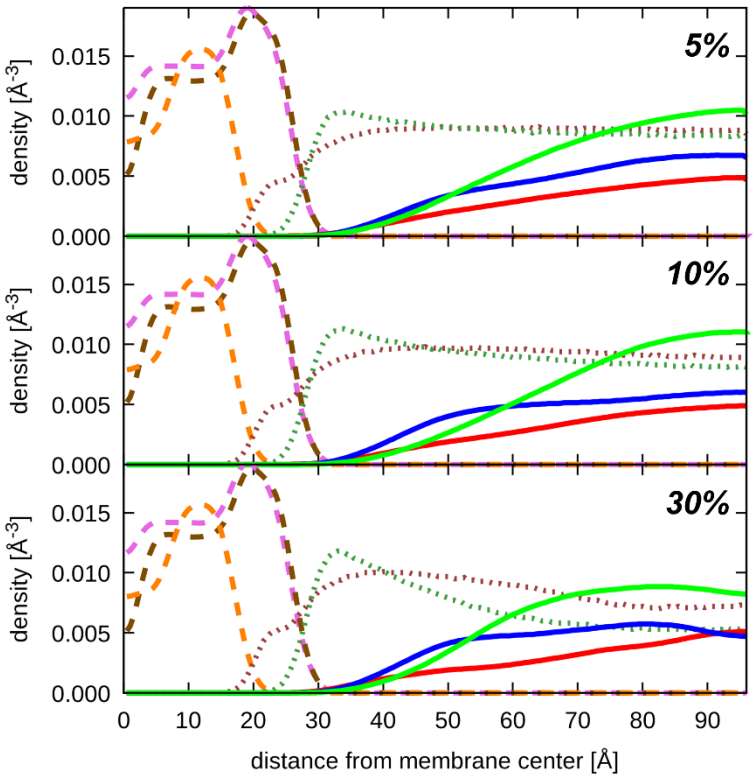

Figure S17: Heavy-atom density distributions along the membrane normal for different proteins (solid lines, villin – red, protein G – blue, ubiquitin – green), individual lipid types (long dashed lines; POPC – dark brown, sphingomyelin - purple, cholesterol – orange), and ions (short dashed lines, Na^+ - red, Cl^- - green). Ion concentrations are shown at 100x of the actual densities. Protein concentrations are shown at 3x for 10% protein concentration and 6x for 5% concentration to facilitate comparisons.

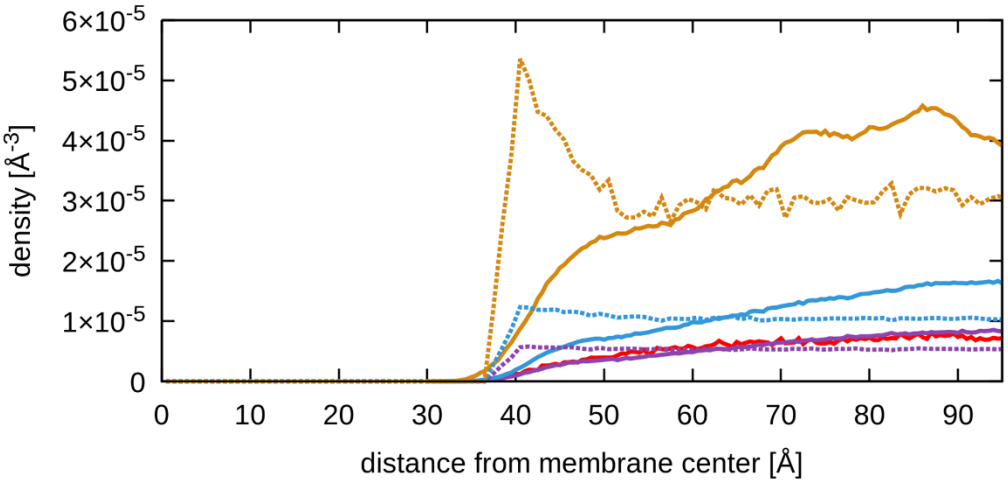

Figure S18: Density distribution of all protein center of mass positions at 5% (purple), 10% (light blue), and 30% (tan) from all-atom (solid lines) and hard-sphere (dashed lines) simulations. Results from a simulation at 5% with the original CHARMM c36 force field of the protein-membrane system without scaling protein-water interactions are shown in red. The hard-sphere results were obtained from coarse-grained (CG) simulations using spherical models with 10 particles for each protein type at box volumes as described in the caption for Figure S16. Interactions with the membrane were also repulsive using the same potential form based on the distance from the membrane plane along the membrane normal. The membrane plane was set at 2.8 nm from the center of the membrane. The value of ε was set to 4 kJ/mol for protein-membrane interactions. For interactions with the membrane, the value of σ for different proteins was used directly. The increase in density near the membrane with the CG model, especially at the higher concentrations, has been observed before and is due to anisotropic collisions with other molecules at a planar surface (33, 34).

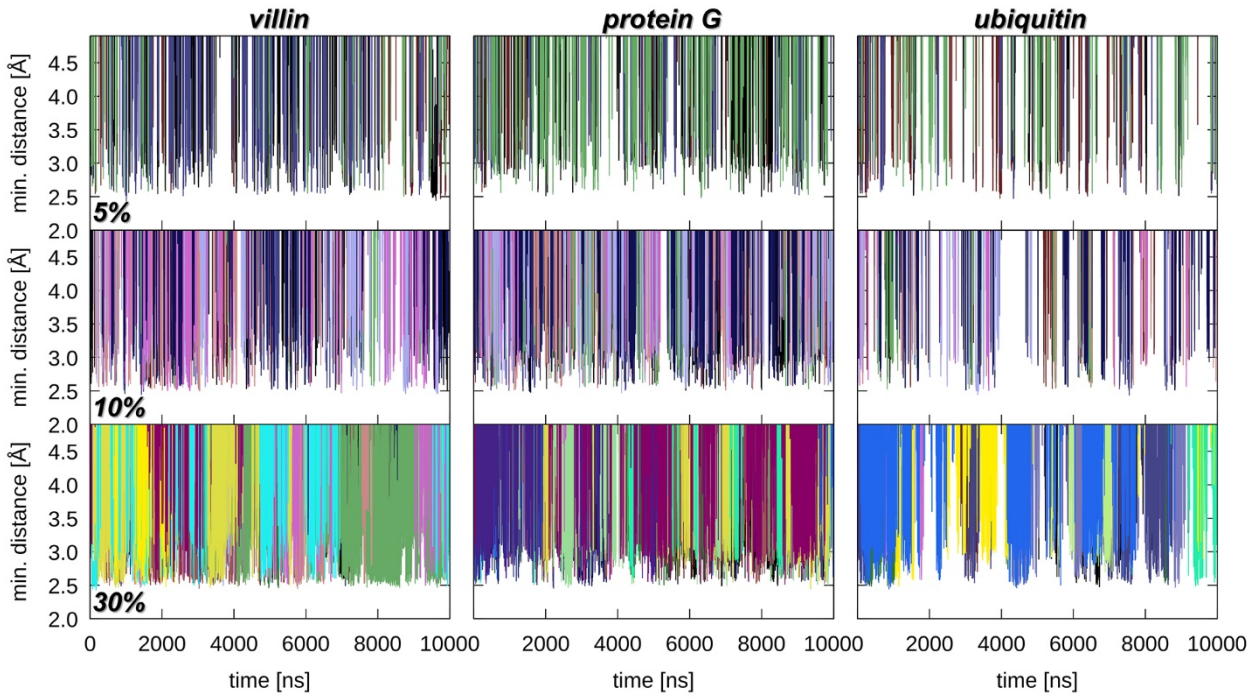

Figure S19: Minimum heavy-atom distances between individual proteins and membrane lipids as a function of simulation time. Different colors indicate different proteins.

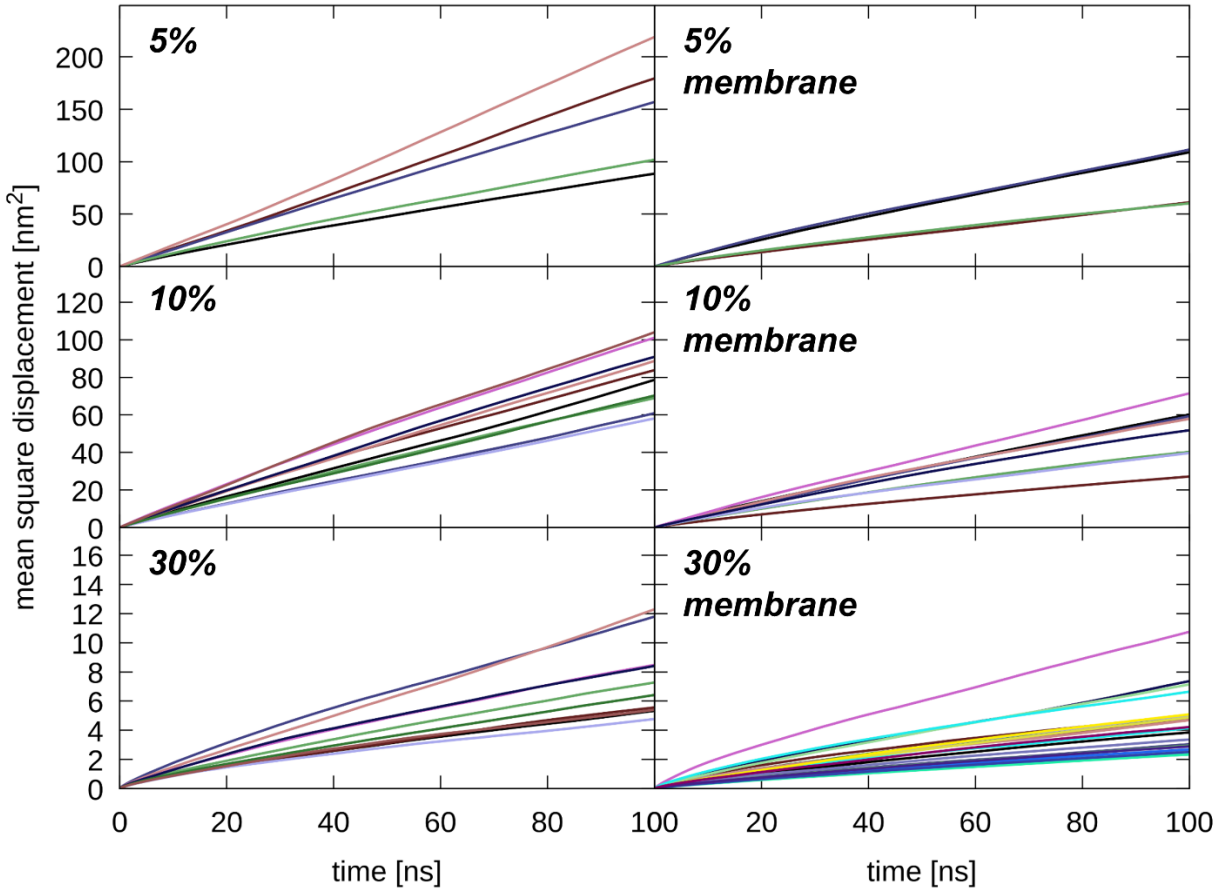

Figure S20: Mean-square displacement (MSD) curves for individual villin molecules (shown with different colors) in the absence (left) and presence (right) of a membrane bilayer at total protein concentrations of 5%, 10%, and 30%.

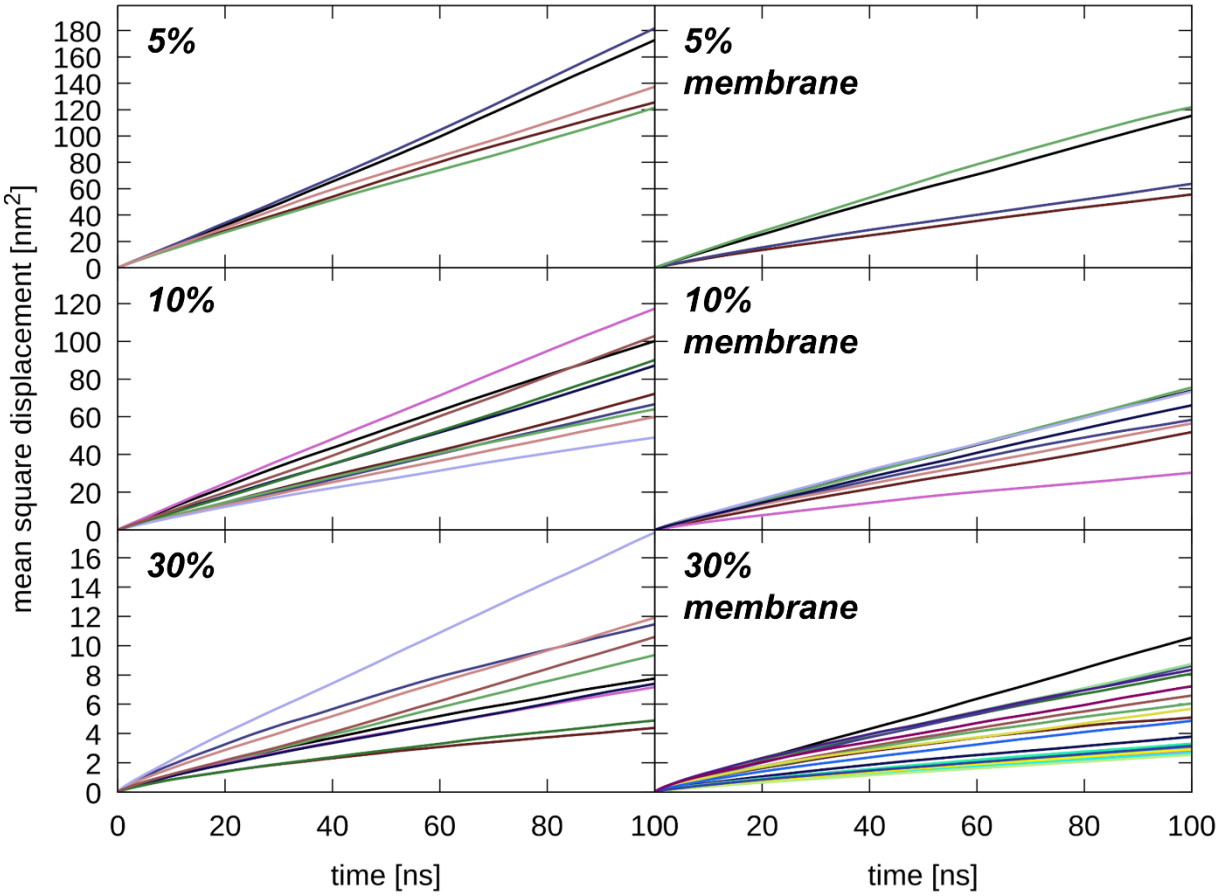

Figure S21: MSD curves for individual protein G molecules (shown with different colors) in the absence (left) and presence (right) of a membrane bilayer at total protein concentrations of 5%, 10%, and 30%.

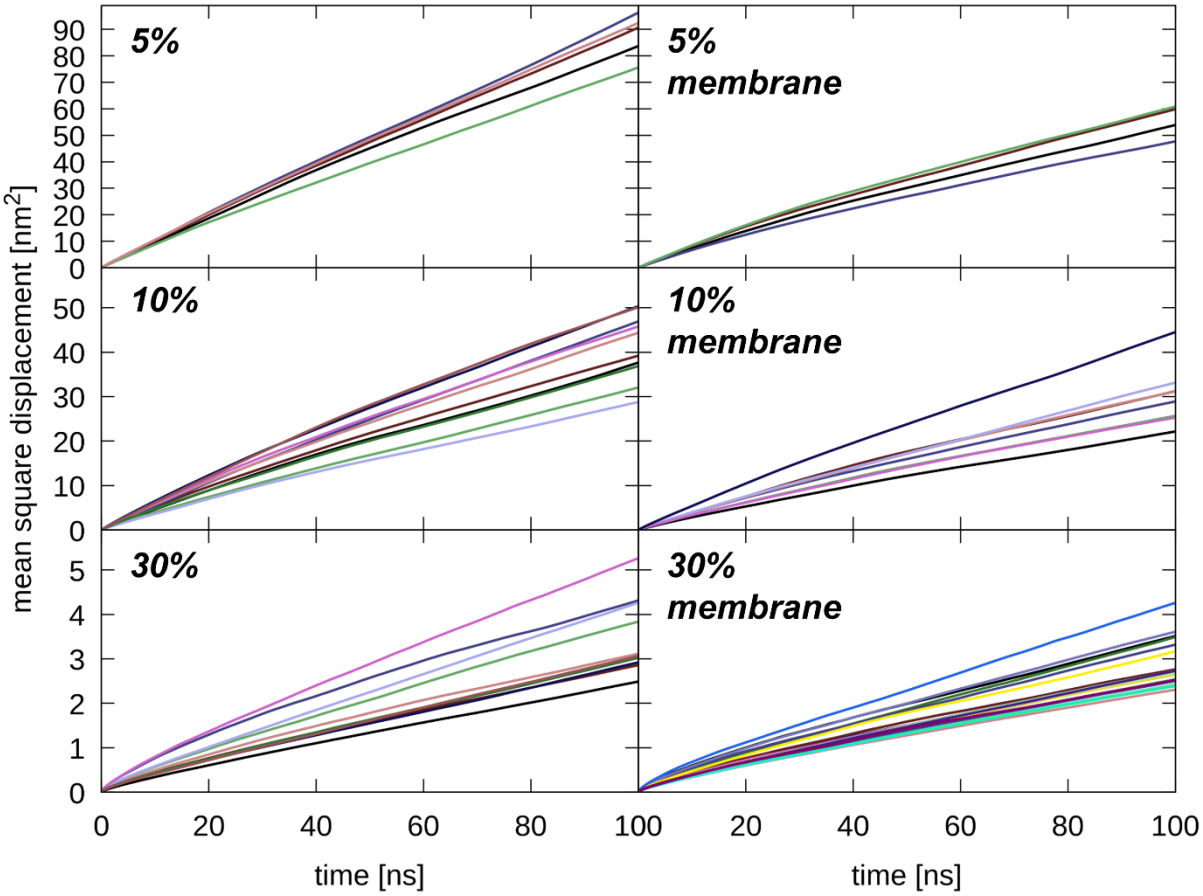

Figure S22: MSD curves for individual ubiquitin molecules (shown with different colors) in the absence (left) and presence (right) of a membrane bilayer at total protein concentrations of 5%, 10%, and 30%.

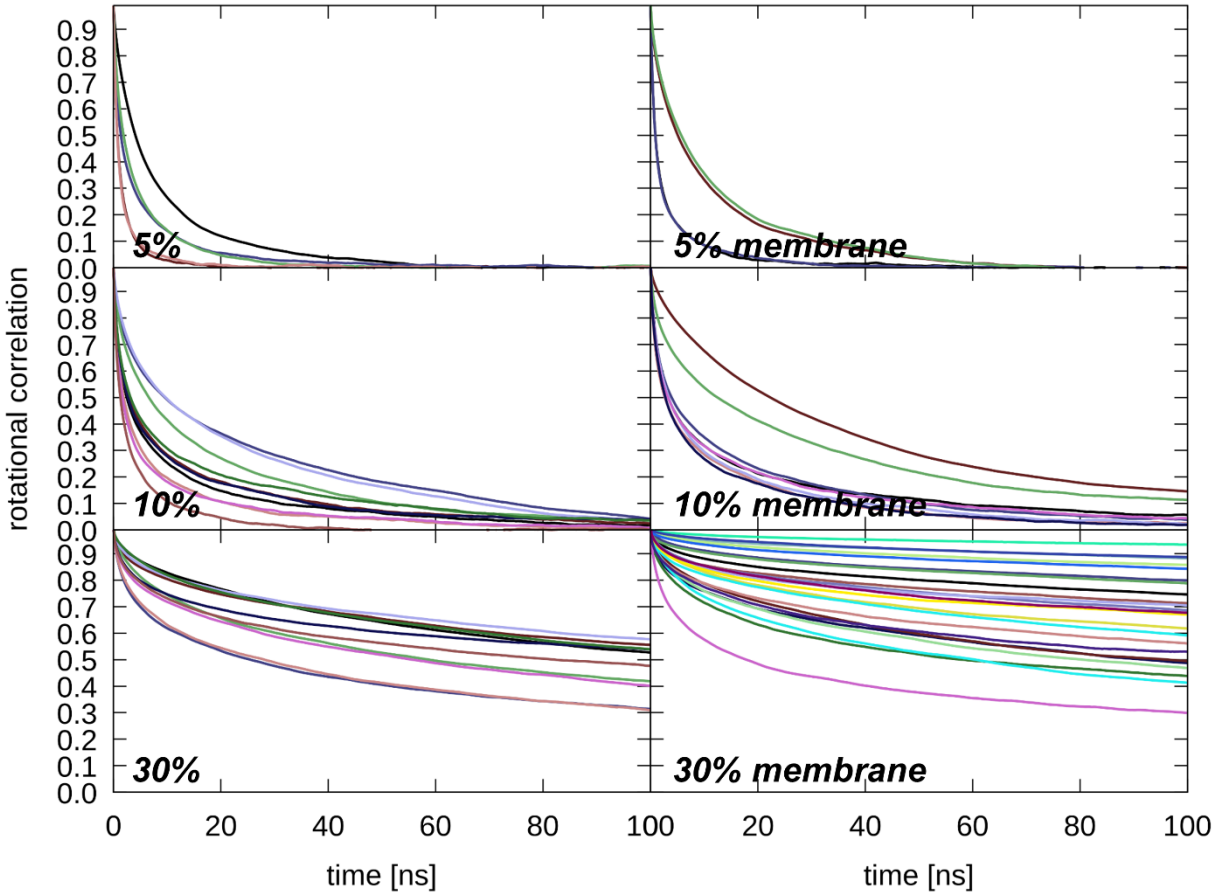

Figure S23: Rotational correlation functions for individual villin molecules (shown with different colors) in the absence (left) and presence (right) of a membrane bilayer at total protein concentrations of 5%, 10%, and 30%.

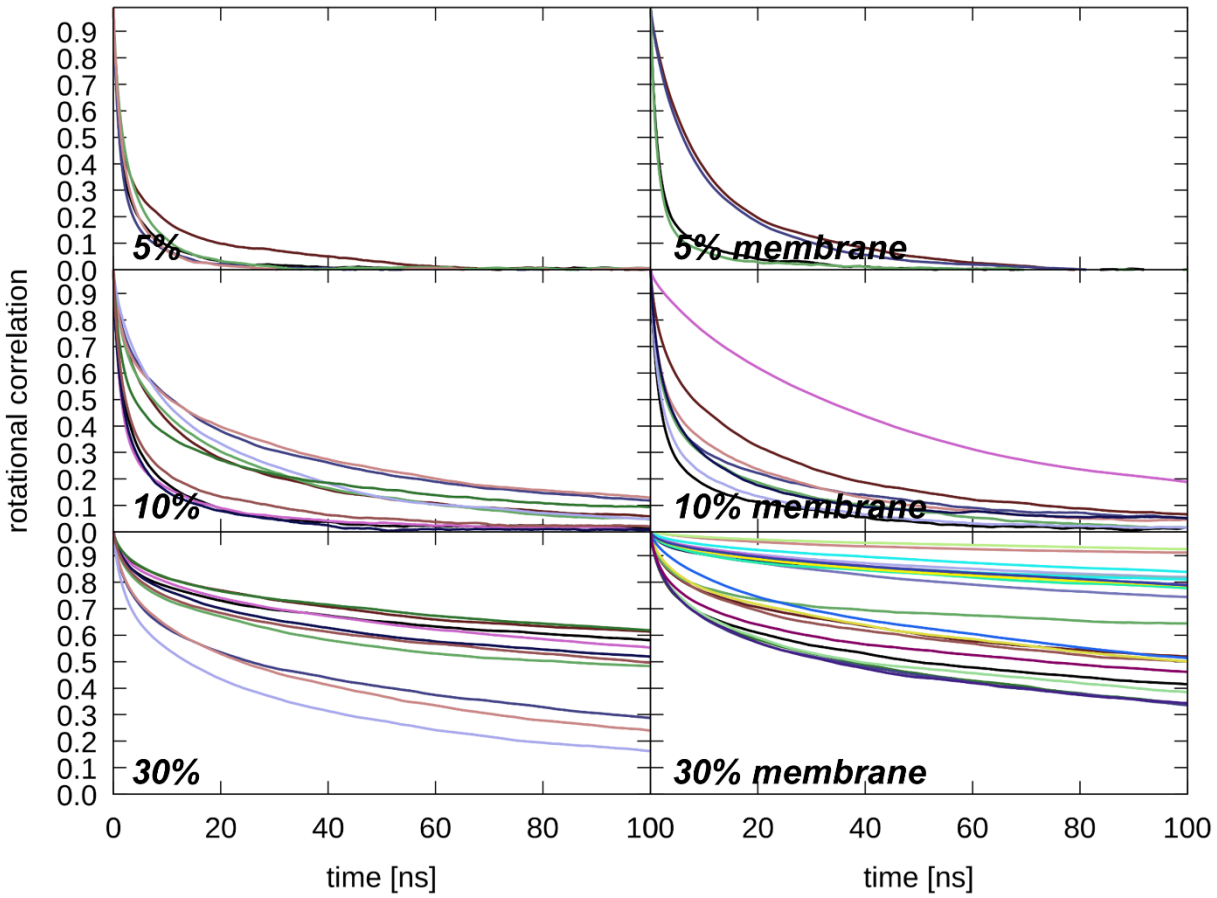

Figure S24: Rotational correlation functions for individual protein G molecules (shown with different colors) in the absence (left) and presence (right) of a membrane bilayer at total protein concentrations of 5%, 10%, and 30%.

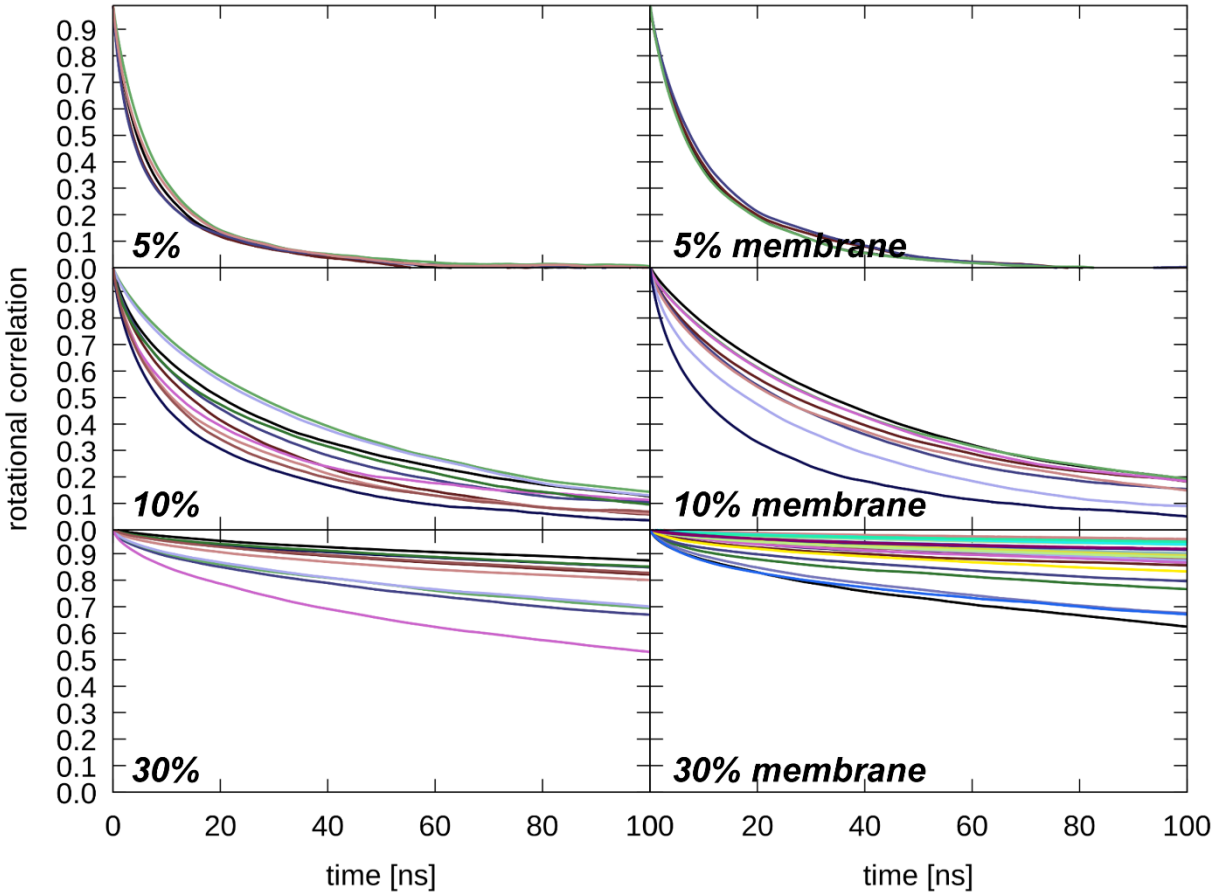

Figure S25: Rotational correlation functions for individual ubiquitin molecules (shown with different colors) in the absence (left) and presence (right) of a membrane bilayer at total protein concentrations of 5%, 10%, and 30%.

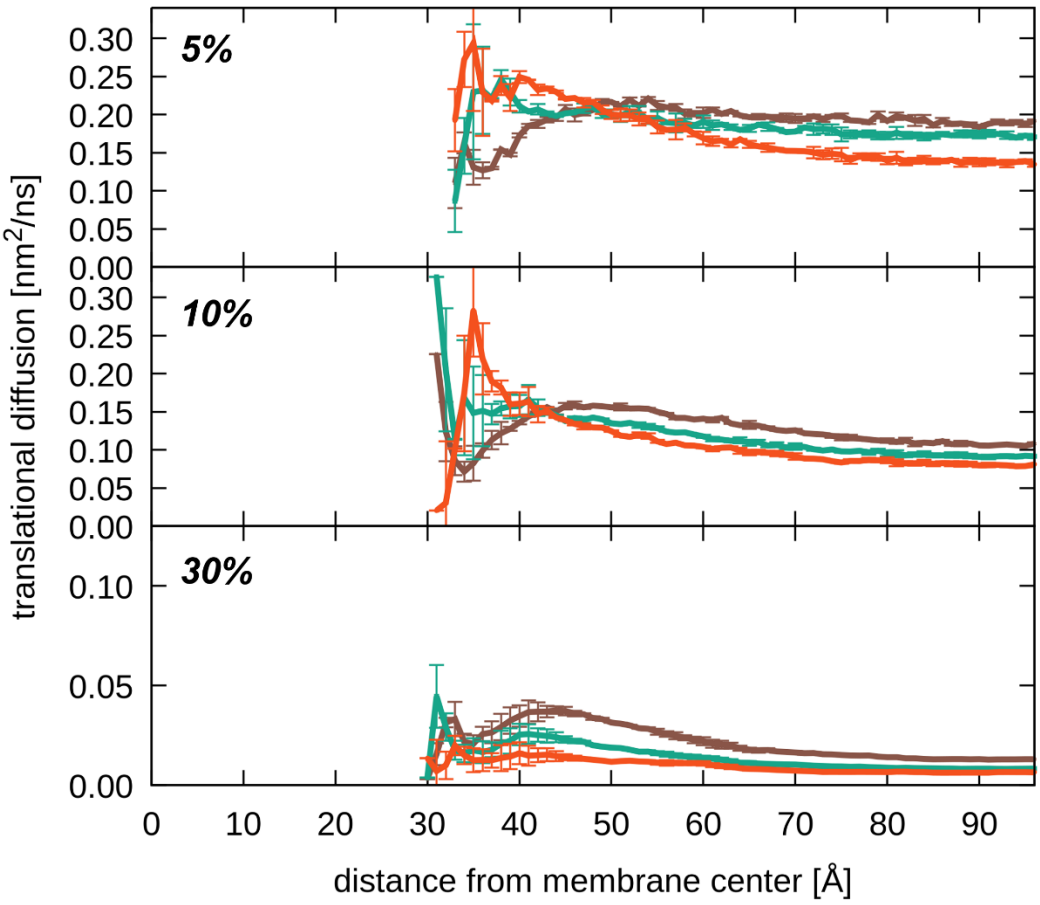

Figure S26: Translational diffusion parallel to the membrane (in the x-y plane) as a function of the distance from the membrane center extracted from mean-square displacement curves at different time scales: 0-1 ns (dark brown), 1-10 ns (green), 10-100 ns (red) for protein concentrations of 5%, 10%, and 30%. MSD curves were generated by combining data from all proteins and the results are not corrected for PBC artefacts.

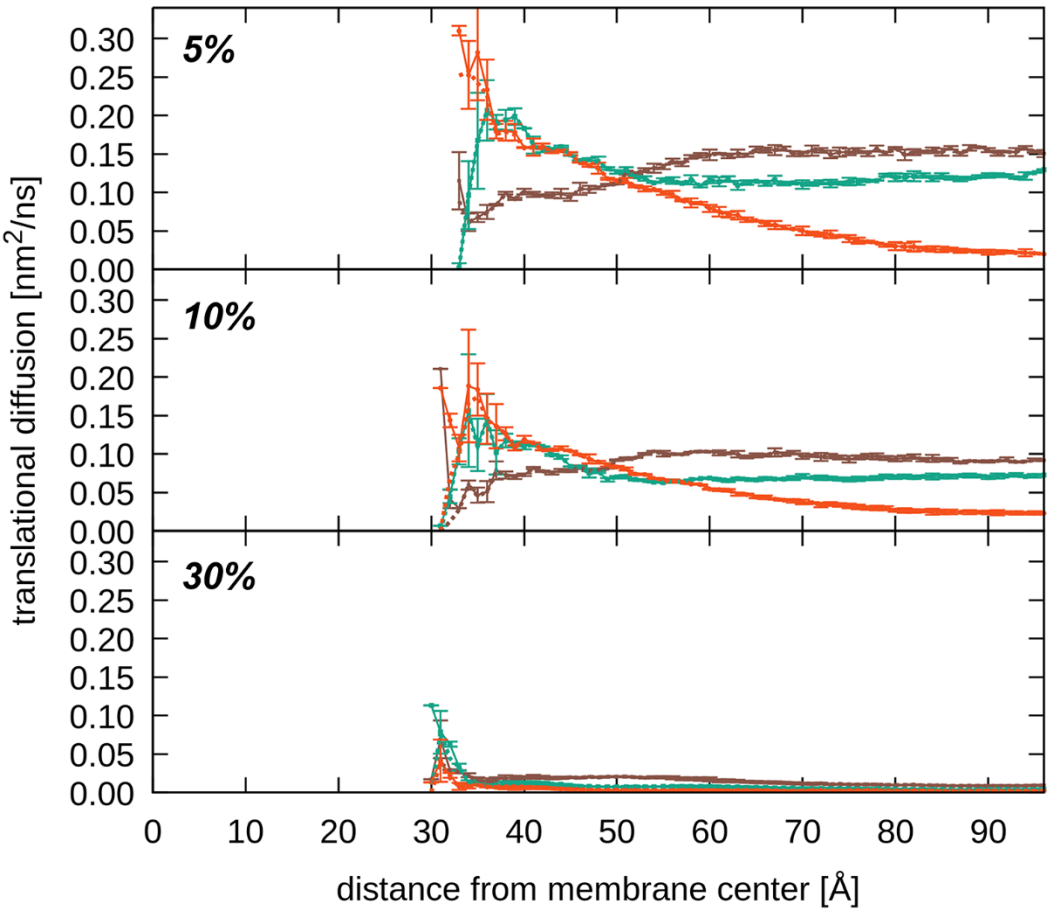

Figure S27: Translational diffusion perpendicular to the membrane (along the z-axis) as a function of the distance from the membrane center extracted from MSD curves at different time scales: 0-1 ns (dark brown), 1-10 ns (green), 10-100 ns (red) for protein concentrations of 5%, 10%, and 30%. MSD curves were generated by combining data from all proteins and the results are not corrected for PBC artefacts. Dashed lines indicate results from a modified analysis where potential drift along z due to a potential gradient was subtracted according to $\langle(dz-\langle dz\rangle)^2\rangle$.

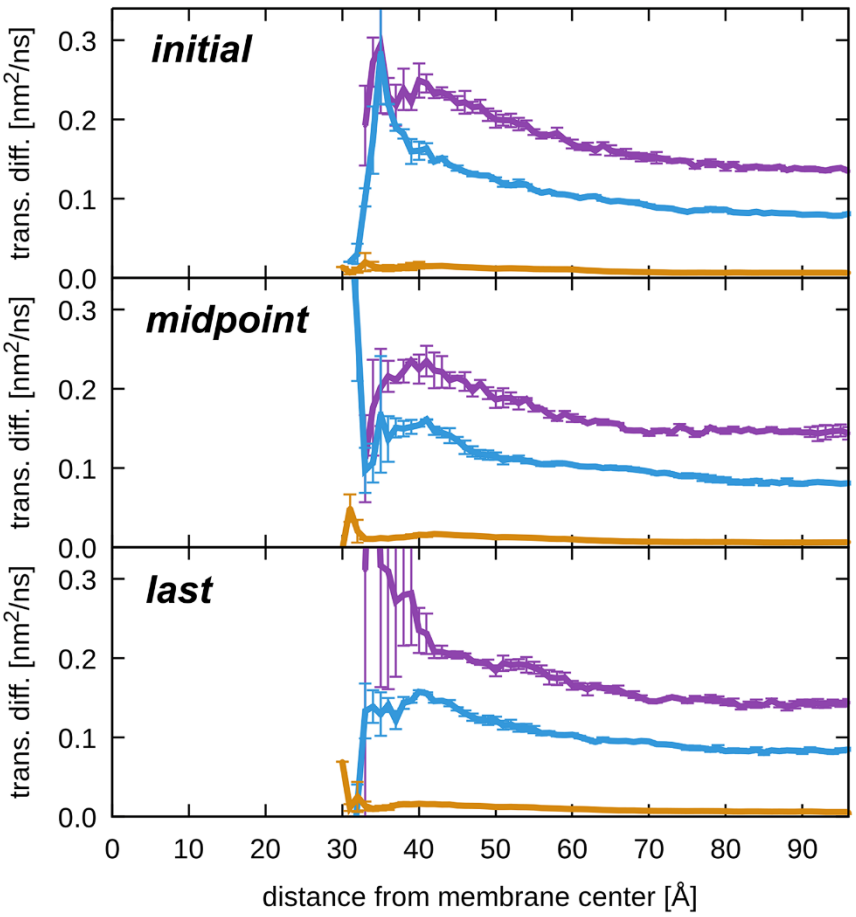

Figure S28: Translational diffusion constants parallel to the membrane (in the x-y plane) as a function of the distance from the membrane center extracted from MSD curves at 10-100 ns for all proteins at total protein concentrations of 5% (purple), 10% (light blue), and 30% (tan). Top, center, and bottom panels compare the choice of the initial, mid, or last point of the time interval over which diffusion is measured in assigning the distance from the membrane center.

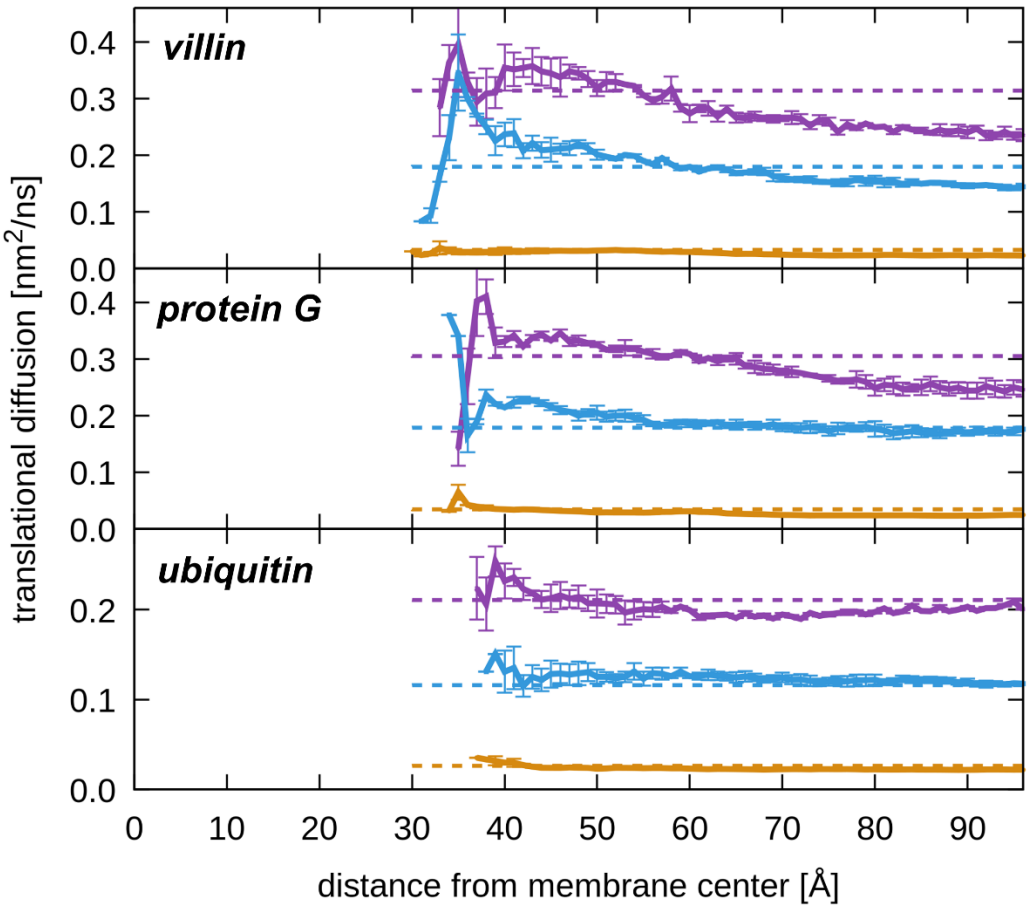

Figure S29: Translational diffusion constants parallel to the membrane (in the x-y plane) as a function of the distance from the membrane center extracted from MSD curves at 10-100 ns for only villin, protein G, or ubiquitin molecules at total protein concentrations of 5% (purple), 10% (light blue), and 30% (tan). Diffusion values were corrected for PBC artefacts according to Eq. 3 with viscosity estimated by Eq. 2. The dashed lines indicate the predicted diffusion rates parallel to the membrane according to Eq. 4 based on PBC-corrected diffusion for different proteins in non-membrane systems at the same concentration (see data in Table S7).

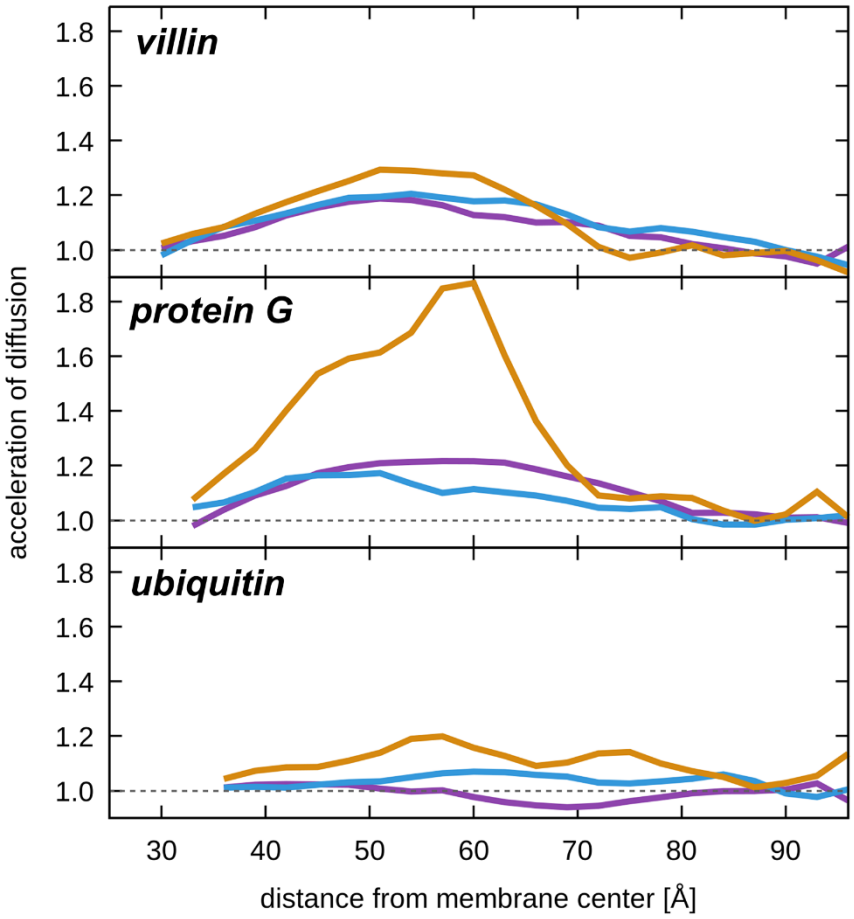

Figure S30: Population-weighted acceleration of translational diffusion parallel to the membrane (in the x-y plane) as a function of the distance from the membrane center extracted from MSD curves at 10-100 ns for only villin, protein G, or ubiquitin molecules at total protein concentrations of 5% (purple), 10% (light blue), and 30% (tan). Acceleration was calculated as the ratio of relaxation times τ_{bulk}/τ_z where $\tau_{bulk}=1/D_{bulk}$ and $\tau_z=(1-p)/D_z+ p/D_{bulk}$ and $p=\rho_z/\rho_{bulk}$ is the probability of a protein to be found at a given distance from the membrane center based on the z-dependent density profiles shown in Figure S18 and the z-dependent diffusion profiles shown in Figure S29.

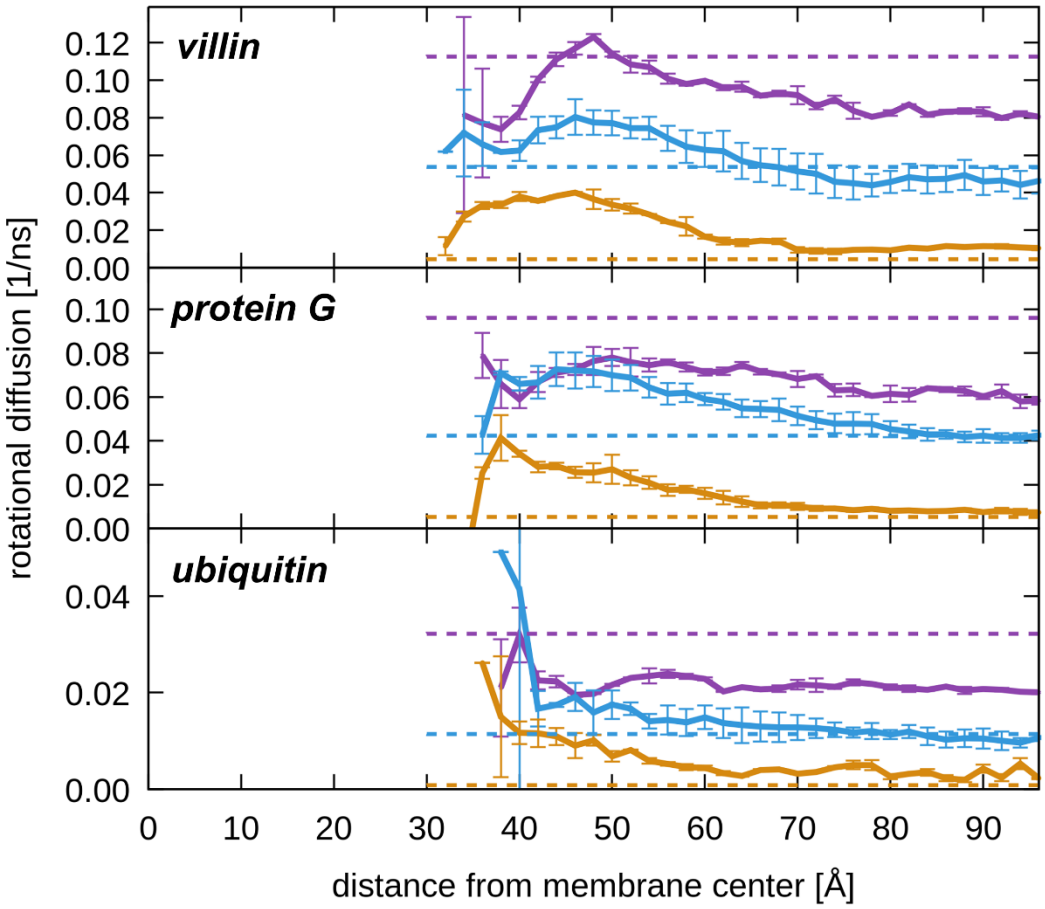

Figure S31: Rotational diffusion constants as a function of the distance from the membrane center for only villin, protein G, or ubiquitin molecules at total protein concentrations of 5% (purple), 10% (light blue), and 30% (tan). The dashed lines indicate rotational diffusion for different proteins in non-membrane systems at the same concentration (see data in Table S8).

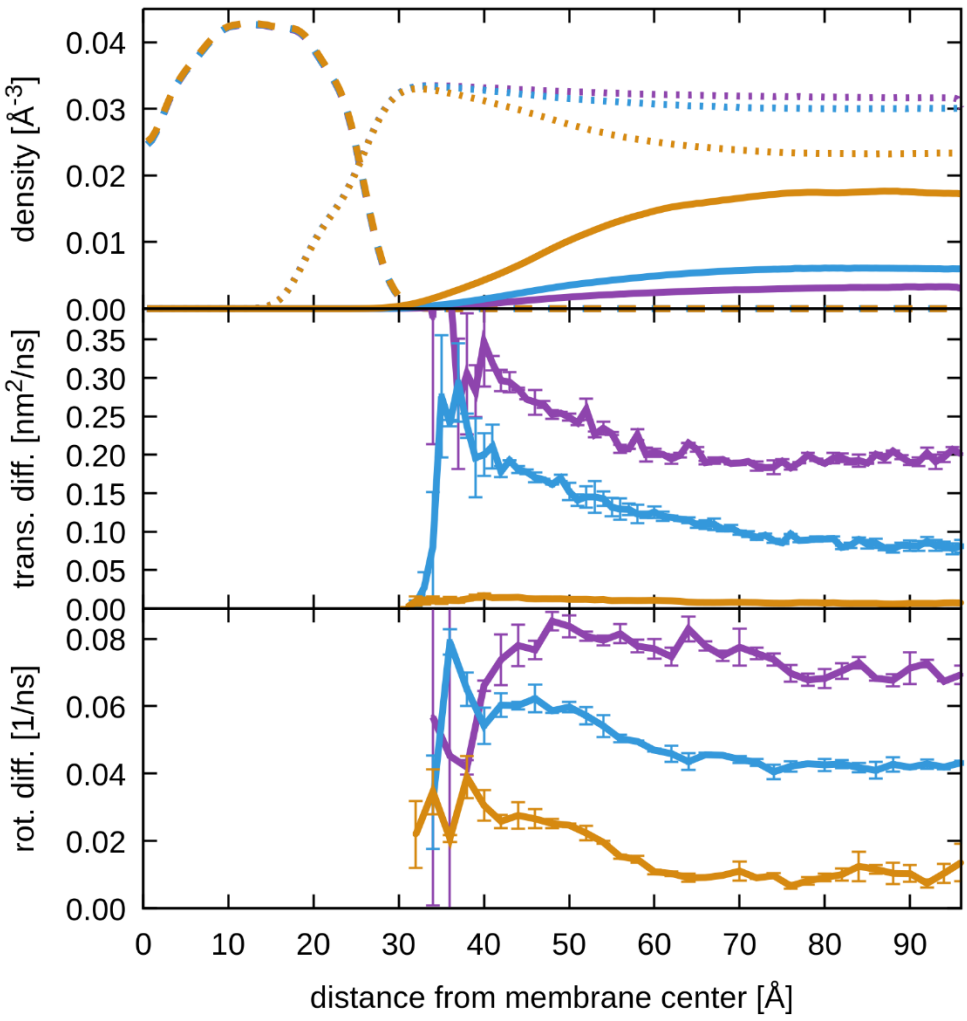

Figure S32: Heavy-atom density distributions of molecular components (top), translational diffusion constants parallel to the membrane (center), and rotational diffusion constants (bottom) as a function of the distance from the membrane center along the membrane normal as in Figure 3 but based on only the first 2 μs before some proteins unfold. Results are shown for systems with proteins at 5% (purple), 10% (light blue), and 30% (tan). Densities are shown for proteins (solid lines), lipids (long dashes), and water molecules (short dashes). Translational and rotational diffusion constants were assigned to the center of mass of a given protein at the beginning of the intervals for which mean-square displacements (MSD) and rotational correlation functions were obtained. Translational diffusion was estimated from MSD vs. time during 10-100 ns. Statistical errors for density distributions are less than 1%.

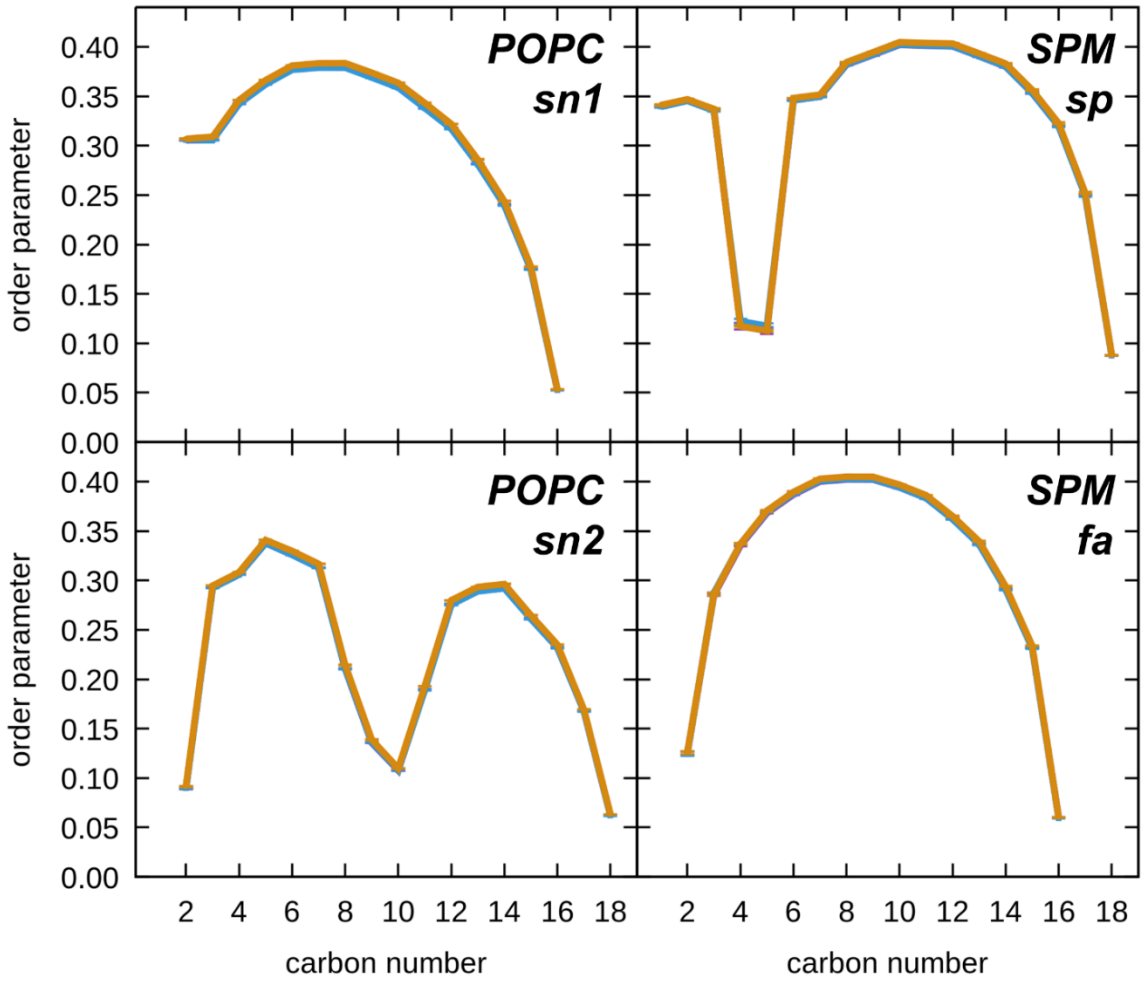

Figure S33: Lipid acyl chain order parameters for POPC (sn1 and sn2 chains) and sphingomyelin (sphingosine (sp) and fatty acid (fa) chains) with protein concentrations of 5% (purple), 10% (light blue), and 30% (tan). Order parameters were calculated as $S=|\langle(3\cos^2\theta-1)/2\rangle|$, where θ is the time-dependent angle of a given C-H bond vectors along the acyl chains relative to the membrane normal.

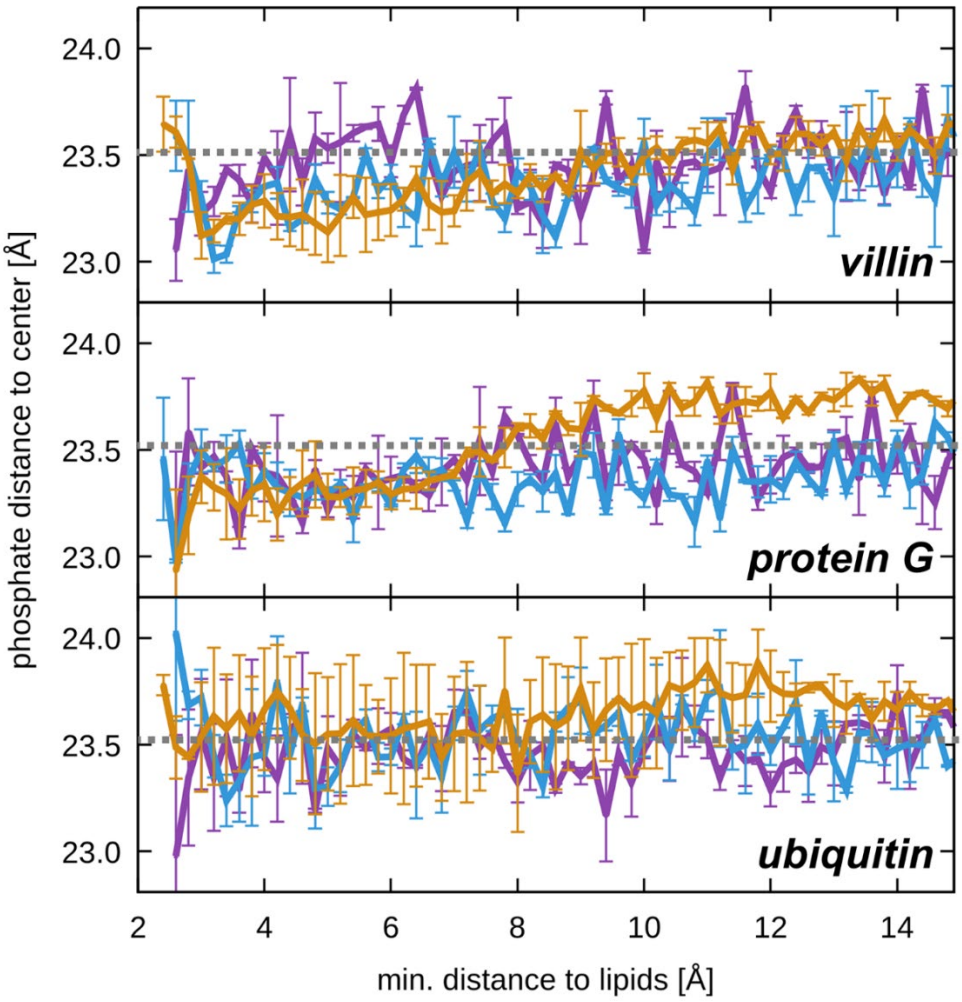

Figure S34: Membrane distortion as a function of protein interactions at protein concentrations of 5% (purple), 10% (light blue), and 30% (tan) as in Figure 4 but based on only the first 2 us of each trajectory before some proteins unfold. Protein-membrane distances are defined based on minimum heavy-atom distances between proteins and lipids. Membrane distortions are characterized by average phosphate distances from the membrane center for phosphate atoms within a 15 Å radius from the lipid atom in closest contact with the protein. The average phosphate distance to the center irrespective of any protein contact is indicated as a grey line.

Table S1. Simulated all-atom systems.

System	%vol	#proteins[1]	#lipids[2]	#K^+/Cl^-	#atoms	Box x-y *[Å]*	Box z *[Å]*	NAMD time *[μs]*	Anton2 time *[μs]*
a5	5	15 (5,5,5)	0	306/296	364721	153.32	153.32	0	10
a10	10	30 (10,10,10)	0	326/306	371813	154.06	154.06	0.5	10
a30	30	30 (10,10,10)	0	103/83	124602	106.31	106.31	0.8	10
a5m	5	12 (4,4,4)	828 (276,276,276)	232/224	363360	134.21	193.71	0	10
a10m	10	24 (8,8,8)	828 (276,276,276)	248/232	367575	134.18	195.34	0.3	10
a30m	30	69 (23,23,23)	828 (276,276,276)	227/181	369657	133.74	194.83	0.7	10

[1]number of villin, protein G, and ubiquitin molecules given in parentheses; [2]number of POPC, sphingomyelin, and cholesterol molecules given in parentheses.

Table S2. Protein stability vs. concentration.

	%vol	Villin			Protein G			Ubiquitin		
		avg.[1]	err.[2]	p[3]	avg.[1]	err.[2]	p[3]	avg.[1]	err.[2]	p[3]
RMSD	5	3.36	*0.81*		1.14	*0.12*		1.56	*0.13*	
[Å]	10	4.41	*0.87*		2.17	*0.65*		1.48	*0.11*	
	30	2.79	*0.41*	0.59	1.29	*0.06*	0.39	1.44	*0.08*	0.53
RMSD	5	1.31	*0.08*		1.04	*0.07*		1.53	*0.13*	
(<2.5 Å)[4]	10	1.32	*0.11*		1.20	*0.07*		1.45	*0.11*	
[Å]	30	1.42	*0.05*	0.35	1.18	*0.05*	0.24	1.42	*0.08*	0.54
$\mathbf{R_g}$	5	11.26	*0.596*		10.80	*0.017*		11.96	*0.048*	
[Å]	10	11.82	*0.624*		11.25	*0.396*		11.93	*0.032*	
	30	10.53	*0.231*	0.37	10.80	*0.009*	0.90	11.90	*0.022*	0.37
$\mathbf{R_g}$	5	9.84	*0.059*		10.79	*0.006*		11.95	*0.047*	
(<2.5 Å)[4]	10	9.21	*0.616*		10.78	*0.005*		11.92	*0.031*	
[Å]	30	9.85	*0.037*	0.88	10.79	*0.007*	0.73	11.90	*0.020*	0.37

[1]averages over Anton2 simulations with and without membrane; [2]standard errors obtained via error propagation formula from trajectory errors; [3]p-values for null hypothesis of values at 5% and 30% not being significantly different; [4]considering only folded snapshots with an RMSD of less than 2.5 Å.

Table S3. Protein stability vs. membrane presence.

	mem.	Villin			Protein G			Ubiquitin		
		avg.[1]	err.[2]	p[3]	avg.[1]	err.[2]	p[3]	avg.[1]	err.[2]	p[3]
RMSD	no	3.45	*0.591*		1.82	*0.431*		1.59	*0.094*	
[Å]	yes	3.59	*0.599*	0.88	1.25	*0.093*	0.32	1.39	*0.090*	0.27
RMSD (<2.5 Å)[4]	no	1.34	*0.033*		1.15	*0.041*		1.56	*0.088*	
[Å]	yes	1.36	*0.088*	0.81	1.13	*0.058*	0.73	1.37	*0.087*	0.28
$\mathbf{R_g}$	no	11.13	*0.400*		11.11	*0.264*		11.95	*0.026*	
[Å]	yes	11.28	*0.440*	0.83	10.79	*0.016*	0.35	11.91	*0.032*	0.39
$\mathbf{R_g}$ **(<2.5 Å)**[4]	no	9.81	*0.026*		10.79	*0.005*		11.94	*0.025*	
[Å]	yes	9.46	*0.412*	0.48	10.78	*0.005*	0.56	11.90	*0.031*	0.41

[1]averages over Anton2 simulations at all three concentrations; [2]standard errors obtained via error propagation formula from trajectory errors; [3]p-values for null hypothesis of values without and with membrane not being significantly different; [4]considering only folded snapshots with an RMSD of less than 2.5 Å.

Table S4. Protein contacts by protein.

System	%vol	Villin			Protein G			Ubiquitin			All	
		$f_{cont.}$[1]	err.[2]	f/A[3]	$f_{cont.}$[1]	err.[2]	f/A[3]	$f_{cont.}$[1]	err.[2]	f/A[3]	$f_{cont.}$[1]	err.[2]
a5	5	0.08	*0.04*	0.77	0.02	*0.002*	0.15	0.36	*0.16*	2.10	0.55	*0.07*
a10	10	0.24	*0.05*	2.32	0.13	*0.05*	0.97	0.48	*0.08*	2.80	0.86	*0.06*
a30	30	0.39	*0.09*	3.78	0.18	*0.03*	1.34	0.76	*0.06*	4.44	1.24	*0.08*
a5m	5	0.27	*0.03*	2.61	0.04	*0.01*	0.30	0.39	*0.05*	2.28	0.86	*0.06*
a10m	10	0.25	*0.04*	2.42	0.07	*0.01*	0.52	0.74	*0.10*	4.32	0.92	*0.07*
a30m	30	0.30	*0.02*	2.91	0.24	*0.03*	1.79	0.75	*0.08*	4.38	1.48	*0.06*

[1]fraction of contacts calculated as the number of contacts between the same type of proteins normalized by the number of proteins; contacts were defined as minimum Cα distances of less than 7 Å; [2]standard errors based on variation between different sets of molecules; [3]fraction of contacts divided by the surface areas of a sphere with an equivalent volume to the respective protein structures (A_{villin}=10.33 nm^2, $A_{proteinG}$=13.42 nm^2, $A_{ubiquitin}$=17.11 nm^2) multiplied by 100.

Table S5. Amino acid preferences in protein-membrane contacts[1].

Amino acid	5%	err.	10%	err.[2]	30%	err.[2]
Alkanes[3]	19.1	*11.7*	11.8	*1.0*	18.2	*2.0*
Polar[4]	16.6	*1.1*	15.1	*2.9*	22.6	*7.2*
Acidic[5]	6.0	*0.1*	6.2	*0.4*	5.7	*1.6*
Basic[6]	4.0	*0.01*	4.9	*1.5*	5.9	*2.3*
Sulfur[7]	4.6	*1.2*	3.9	*1.4*	2.3	*0.2*
Phe	3.1	*0.7*	2.7	*0.6*	2.6	*0.1*
Tyr	16.6	*2.3*	12.4	*2.3*	18.2	*10.2*
Trp	6.5	*0.6*	9.6	*1.4*	10.8	*4.3*
Gly	1.7	*0.1*	1.7	*0.3*	2.8	*0.6*
Pro	1.9	*0.1*	2.6	*1.1*	3.1	*0.6*

[1]Percentages of closest protein-lipid heavy atom contacts within 5 Å as a function of amino acid type, e.g. in 6.5% of all instances where Trp is the closest amino acid in a protein to the membrane at 5% concentration; [2]Statistical uncertainties from comparing subsets of proteins; [3]Ala, Val, Leu, Ile; [4]Asn, Gln, His, Ser, Thr; [5]Asp, Glu; [6]Arg, Lys.

Table S6. Protein-membrane contact residence times.

System	%vol	Protein	τ_1 [ns]	err.	τ_2 [ns]	err	a	err.
a5m	**5**	villin	5.28	*1.10*	1585	*727*	0.95	*0.02*
		protein G	1.51	*0.41*	1843	*650*	0.97	*0.001*
		ubiquitin	3.87	*0.95*	945	*408*	0.95	*0.003*
		all	**3.20**	*0.18*	**1144**	*11.5*	**0.94**	*0.008*
a10m	**10**	villin	3.93	*0.42*	1622	*1302*	0.91	*0.02*
		protein G	2.22	*0.13*	1296	*118*	0.93	*0.004*
		ubiquitin	1.65	*0.04*	75	*12.3*	0.81	*0.03*
		all	**3.09**	*0.37*	**829**	*420*	**0.92**	*0.01*
a30m	**30**	villin	5.63	*1.05*	684	*15.4*	0.78	*0.03*
		protein G	2.28	*0.41*	622	*278*	0.83	*0.03*
		ubiquitin	1.82	*0.47*	489	*31.1*	0.79	*0.02*
		all	**2.97**	*0.29*	**589**	*78.6*	**0.80**	*0.02*

Residence times from fitting contact survival functions to a double exponential expression with parameters a, τ_1, and τ_2: $a*\exp(-t/\tau_1)+(1-a)*\exp(-t/\tau_2)$. Protein-membrane contacts were defined based on protein-lipid heavy-atom distances within 5 Å.

Table S7. Translational diffusion rates for proteins in non-membrane systems.

System	%vol	Protein	D_t *0-1 ns* [nm^2/ns]	err.	D_t *1-10 ns* [nm^2/ns]	err	D_t *10-100 ns* [nm^2/ns]	err.
a5	**5**	villin	0.36	*0.03*	0.35	*0.03*	0.34	*0.04*
		protein G	0.36	*0.01*	0.34	*0.01*	0.33	*0.02*
		ubiquitin	0.27	*0.007*	0.25	*0.005*	0.23	*0.006*
a10	**10**	Villin	0.23	*0.01*	0.21	*0.01*	0.19	*0.01*
		protein G	0.22	*0.01*	0.21	*0.01*	0.20	*0.01*
		ubiquitin	0.16	*0.005*	0.15	*0.005*	0.13	*0.004*
a30	**30**	villin	0.051	*0.002*	0.041	*0.002*	0.035	*0.001*
		protein G	0.055	*0.002*	0.043	*0.002*	0.038	*0.002*
		ubiquitin	0.037	*0.001*	0.031	*0.001*	0.029	*0.0004*

Diffusion from mean-square center of mass displacement curves with correction for periodic boundary artefacts according to Eq. 1 with viscosity estimated by Eq. 2 using η_w=0.35 cP (for CHARMM TIP3P with Ewald summation) (35).

Table S8. Rotational diffusion rates for proteins in non-membrane systems.

System	%vol	Protein	D_r [1/ns]	err.	τ_f [ns]	err	τ_s [ns]	err.	S_R^2	err.
a5	**5**	villin	0.112	*0.027*	1.36	*0.42*	8.79	*2.01*	0.38	*0.05*
		protein G	0.096	*0.007*	1.21	*0.07*	9.68	*2.51*	0.37	*0.04*
		ubiquitin	0.032	*0.003*	3.66	*0.60*	17.59	*1.49*	0.42	*0.04*
a10	**10**	villin	0.054	*0.008*	2.22	*0.22*	30.50	*3.13*	0.40	*0.04*
		protein G	0.042	*0.008*	3.31	*0.50*	43.61	*7.51*	0.39	*0.03*
		ubiquitin	0.011	*0.001*	9.07	*1.28*	58.25	*4.33*	0.54	*0.03*
a30	**30**	villin	0.005	*0.001*	21.49	*3.34*	547.5	*64.8*	0.55	*0.03*
		protein G	0.005	*0.001*	21.62	*2.81*	708.1	*124.7*	0.54	*0.03*
		ubiquitin	0.0008	*0.0002*	72.47	*10.8*	1502.9	*276.8*	0.77	*0.02*

Diffusion rates were extracted from fitting rotational correlation functions to double exponentials: $S_R^2*\exp(-t/\tau_s)+(1-S_R^2)*\exp(-t/\tau_f)$.

References

1. S. Jo, T. Kim, V. G. Iyer, W. Im, CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J Comput Chem* **29**, 1859-1865 (2008).
2. S. Jo, J. B. Lim, J. B. Klauda, W. Im, CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophys J* **97**, 50-58 (2009).
3. E. L. Wu *et al.*, CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J Comput Chem* **35**, 1997-2004 (2014).
4. C. J. McKnight, P. T. Matsudaira, P. S. Kim, NMR Structure of the 35-Residue Villin Headpiece Subdomain. *Nat Struct Biol* **4**, 180-184 (1997).
5. A. M. Gronenborn *et al.*, A Novel, Highly Stable Fold of the Immunoglobulin Binding Domain of Streptococcal Protein G. *Science* **253**, 657-661 (1991).
6. S. Vijaykumar, C. E. Bugg, W. J. Cook, Structure of Ubiquitin Refined at 1.8 a Resolution. *J Mol Biol* **194**, 531-544 (1987).
7. R. B. Best *et al.*, Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain $\chi1$ and $\chi2$ Dihedral Angles. *J Chem Theory Comput* **8**, 3257-3273 (2012).
8. G. Nawrocki, P.-h. Wang, I. Yu, Y. Sugita, M. Feig, Slow-Down in Crowded Protein Solutions Correlates with Transient Oligomer Formation. *J Phys Chem B* **121**, 11072-11084 (2017).
9. J. B. Klauda *et al.*, Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J Phys Chem B* **114**, 7830-7843 (2010).
10. A. D. MacKerell, Jr. *et al.*, All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Phys Chem B* **102**, 3586-3616 (1998).
11. A. Sandoval-Perez, K. Pluhackova, R. A. Böckmann, Critical Comparison of Biomembrane Force Fields: Protein–Lipid Interactions at the Membrane Interface. *J Chem Theory Comput* **13**, 2310-2321 (2017).
12. D. Beglov, B. Roux, Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J Chem Phys* **100**, 9050-9063 (1994).
13. Y. Luo, B. Roux, Simulation of Osmotic Pressure in Concentrated Aqueous Salt Solutions. *J Phys Chem Lett* **1**, 183-189 (2010).
14. H. J. C. Berendsen, J. P. M. Postma, W. F. v. Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath. *J Chem Phys* **81**, 3684-3690 (1984).
15. R. A. Lippert *et al.*, Accurate and efficient integration for molecular dynamics simulations at constant temperature and pressure. *J Chem Phys* **139**, 164106 (2013).
16. V. Kräutler, W. F. Van Gunsteren, P. H. Hünenberger, A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J Comput Chem* **22**, 501-508 (2001).
17. G. J. Martyna, D. J. Tobias, M. L. Klein, Constant pressure molecular dynamics algorithms. *J Chem Phys* **101**, 4177-4189 (1994).
18. S. Nose, A molecular dynamics method for simulations in the canonical ensemble. *Mol Phys* **52**, 255-268 (1984).

19. Y. Shan, J. L. Klepeis, M. P. Eastwood, R. O. Dror, D. E. Shaw, Gaussian split Ewald: A fast Ewald mesh method for molecular simulation. *J Chem Phys* **122**, 054101 (2005).
20. G. Nawrocki, A. Karaboga, Y. Sugita, M. Feig, Effect of Protein–Protein Interactions and Solvent Viscosity on the Rotational Diffusion of Proteins in Crowded Environments. *Phys Chem Chem Phys* **21**, 876-883 (2019).
21. I. C. Yeh, G. Hummer, System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. *J Phys Chem B* **108**, 15873-15879 (2004).
22. A. Ortega, D. Amorós, J. García de la Torre, Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. *Biophys J* **101**, 892-898 (2011).
23. C. Tanford, *Physical Chemistry of Macromolecules* (John Wiley & Sons, New York, 1961).
24. S. von Bülow, M. Siggel, M. Linke, G. Hummer, Dynamic cluster formation determines viscosity and diffusion in dense protein solutions. *Proc Natl Acad Sci USA* 10.1073/pnas.1817564116, 201817564 (2019).
25. P. Simonnin, B. Noetinger, C. Nieto-Draghi, V. Marry, B. Rotenberg, Diffusion under confinement: hydrodynamic finite-size effects in simulation. *J Chem Theory Comput* **13**, 2881-2889 (2017).
26. V. Wong, D. A. Case, Evaluating Rotational Diffusion from Protein MD Simulations. *J Phys Chem B* **112**, 6013-6024 (2008).
27. M. Linke, J. Köfinger, G. Hummer, Rotational Diffusion Depends on Box Size in Molecular Dynamics Simulations. *J Phys Chem Lett* **9**, 2874-2878 (2018).
28. M. Feig, J. Karanicolas, C. L. Brooks, MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. *J Mol Graph Modell* **22**, 377-395 (2004).
29. B. R. Brooks *et al.*, CHARMM: The Biomolecular Simulation Program. *J Comput Chem* **30**, 1545-1614 (2009).
30. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual Molecular Dynamics. *J Mol Graph* **14**, 33-38 (1996).
31. T. K. Chiu *et al.*, High-resolution x-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein. *Proc Natl Acad Sci USA* **102**, 7517-7522 (2005).
32. E. Mani, W. Lechner, W. K. Kegel, P. G. Bolhuis, Equilibrium and Non-Equilibrium Cluster Phases in Colloids with Competing Interactions. *Soft Matter* **10**, 4479-4486 (2014).
33. D. Deb *et al.*, Hard sphere fluids at a soft repulsive wall: A comparative study using Monte Carlo and density functional methods. *J Chem Phys* **134**, 214706 (2011).
34. R. Sibug-Aga, B. B. Laird, Structure of a soft-sphere fluid at a soft repulsive wall: A comparison of weighted density-functional theories. *Phys Rev E* **69**, 051502 (2004).
35. S. E. Feller, R. W. Pastor, A. Rojnuckarin, S. Bogusz, B. R. Brooks, Effect of Electrostatic Force Truncation on Interfacial and Transport Properties of Water. *J Phys Chem* **100**, 17011-17020 (1996).