

Supporting Information:

Direct evidence for upright and helical structure of alpha-synuclein at lipid membranes

Steven J. Roeters,^{*,†,¶} Kris Strunge,^{†,¶} Kasper B. Pedersen,^{†,¶} Thaddeus W.
Golbek,[†] Mikkel Bregnhøj,[†] Yin Wang,[‡] Mingdong Dong,[‡] Daniel Otzen,[‡] Birgit
Schjøtt,^{†,‡} and Tobias Weidner^{*,†}

[†]*Department of Chemistry, Aarhus University, Langelandsgade 140, 8000 Aarhus C,
Denmark*

[‡]*Interdisciplinary Nanoscience Center (iNANO), Aarhus University, Gustav Wieds Vej 14,
8000 Aarhus C, Denmark*

[¶]*These authors contributed equally to the presented work.*

E-mail: ^{*}s.j.roeters@chem.au.dk; ^{*}weidner@chem.au.dk

The supporting information (35 pages in total) contains 3 tables and 13 figures.

1 Experimental details

1.1 α S expression and purification

Purified α S was expressed and purified as described before in ref. S1. Briefly, the plasmid vector pET11-d was used to express α S in *Escherichia coli* BL21 (DE3). The cells were then pelleted by 20 min of centrifugation at 3500 rpm, at 4 °C. The pellet of 1 liter of culture was resuspended in 100 mL of osmotic shock buffer (composed of 30 mM Tris-HCl, 40% sucrose, 2 mM EDTA, at pH 7.2) and incubated for 10 minutes. Subsequently, the suspension was centrifuged for 30 minutes at 9000 g and 20 °C. The resulting pellet was then resuspended in 90 mL ice-cold deionised water and 40 μ L of saturated $MgCl_2$ was added. Subsequently, the pellet was incubated on ice for 3 min, and the supernatant, containing the periplasmic preparation, was collected by 20 min centrifugation at 9000 g and 4 °C. The periplasmic preparation was subjected to acid precipitation with drop-wise addition of 1 M HCl to a final pH level of 3.5 and then incubated for 5 min. The supernatant was collected by 20 min of centrifugation at 9000 g and 4 °C. The pH of the supernatant was immediately adjusted to pH 7.5 with drop-wise addition of 1 M NaOH. The solution was filtered (0.45 μ m) and loaded on a Q-Sepharose column (HiTrap Q HP) pre-equilibrated with 20 mM Tris-HCl pH 7.5. The column was washed with three column volumes of 0.1 M NaCl in buffer followed by elution of α S with a NaCl gradient from 0.1-0.5 M. SDS-PAGE analysis was used to identify fractions with α S and ensure protein purity. Finally the α S was dialyzed exhaustively against deionised water, lyophilized and stored at -20 °C.

Prior to use, the lyophilized powder was dissolved in phosphate buffered saline (PBS; 0.01 M phosphate buffer, 0.0027 M KCl, and 0.137 M NaCl, pD 7.2, Sigma Aldrich) in D_2O (99.9%D, Eurisotop) to avoid interference from H_2O bending modes. The α S solution was prepared either by filtration with 100 kDa Nanosep filters (Pall Corporation, USA) or without this filtration step, because we observed that this did not affect any of the kinetic or spectral features of the measurements.

1.2 Lipid monolayer preparation and protein injection

A lipid monolayer was assembled at the air-water interface within a trough at room temperature. This trough is made of polyether ether ketone (PEEK) with a glass bottom and contained approximately 1.5 ml of PBS in D₂O to avoid the interference with the OH-bending mode of H₂O.^{S2} During the construction of the lipid monolayer, the surface pressure at the air-water interface was monitored by a Dyne probe instrument (Kibron, Finland). Throughout the entire experiment, the water level was held constant by a syringe pump (New Era Pump Systems Inc.) with the cannula submerged at the bottom of the trough. Phospholipid 1,2-dipalmitoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (DPPG, Avanti Polar Lipids, Inc.) was chosen as a model system due to physiological relevance of its headgroup chemistry and characterization as a model system in current literature. The DPPG was first dissolved in chloroform (Honeywell, HPLC grade) and then spread by dropwise addition to the air-water interface with the initial surface pressure set to ~ 15 mN/m, corresponding to a liquid-condensed (LC) phase.^{S3} After equilibrium, VSFG spectra were recorded on the pure lipid monolayer. The final protein concentration in the trough after α S injection was ~ 50 μ M and ~ 50 nM for the low- and high-LPR experiments, respectively. After the injection, the surface pressure increased to 30 mN/m. In each experiment, the protein was dissolved at a 245 μ M concentration in 300 μ L of PBS-D₂O buffer before injection into the subphase of the lipid monolayer. The concentration before protein injection of each experiment was measured by UV-vis spectroscopy using the absorbance at 280 nm (NanoDrop 2000c, Thermo Scientific).

The determination of the LPR values of ~ 0.2 and 16, as used in the Main Text is given in the table below. The area/lipid molecule value of 64 $\text{\AA}^2$ is taken from ref. S4). The factor 2 is incorporated to optimally compare our results to the literature studies, which are all bilayer studies. This way, the number of lipids exposed to the protein solution is the same for a given LPR, without affecting the physics behind the α S-lipid interaction, because α S is not expected to be affected by the presence of the distal lipid leaflet.^{S5} Using the same

methodology, the high-LPR value of 16 was determined for the $[\alpha\text{S}] = 50 \text{ nM}$ experiments, and (with a total area based on a round internal reflective element of 2 mm diameter, a sample volume of 20 μL , an αS concentration of 50 μM , and a lipid area of $\sim 32 \text{ \AA}^2$,^{S6} assuming that the supported lipid bilayer adopted a physiological surface pressure of $\sim 30 \text{ mN/m}$) the 0.033 LPR value derived for reference S7.

Table S1: Calculation of lipid-protein ratios. Same method was applied for the determination of $\text{LPR} = 16$ for $[\alpha\text{S}] = 50 \text{ nM}$.

Trough:	
Total area	0.00061 m^2
Volume	0.0035 L
Protein:	
Concentration	50 μM αS
Moles	1.8×10^{-7}
Lipid:	
Surface pressure	13 mN/m
Area/lipid 13 mN/m	$64 \text{ \AA}^2$
Number of lipids	9.5×10^{14} molecules
	1.6×10^{-9} moles
Exposed-lipid correction:	2 (monolayer-to-bilayer factor)
LPR:	0.018 (lipid-to-protein ratio)
PLR:	56 (protein-to-lipid ratio)

1.3 Vibrational sum-frequency generation (VSFG) spectroscopy

The SFG setup has been previously described in ref. S8. In short, the SFG setup is based on a 7W, 35 femtosecond laser system (Astrella, Coherent) with pulses centered at 800 nm and a repetition rate of 1 kHz. One part of the output was used to pump an optical parametric amplifier (OPA) with a non-collinear difference frequency generation (NDFG) extension (TOPAS Prime, Light Conversion) to generate broadband ($\text{FWHM} \sim 300 \text{ cm}^{-1}$) IR pulses at 6.1 μm . A narrowband ($\text{FWHM} \sim 15 \text{ cm}^{-1}$) visible beam was generated by guiding 1 mJ of the fundamental output through a Fabry-Perot etalon. The visible and IR beams were temporally and spatially overlapped at the sample surface. The SFG signal was focused into a spectrograph (Shamrock 303i, Andor) and detected by an EMCCD camera

(Newton 971, Andor). VSFG spectra are recorded using an IR pulse spectral full-width-at-half-max of $\sim 270 \text{ cm}^{-1}$ (from 1540 to 1810 cm^{-1}), for 5 minutes per sample scan (as an optimal acquisition time in terms of dark-noise reduction without obtaining so many cosmic rays per scan that removing them in an automated fashion becomes challenging) and for 15 seconds for the Au intensity calibration spectra in PPP, and in SSP (S-SF, S-visible, P-IR), PPP, SPS and PSP polarization combinations for the sample spectra. The various polarization combinations were recorded in a sequential manner, frequently going back to the SSP polarization combination in order to get an impression of the development of the overall signal intensity. The intensity of the chiral polarization combination, PSP, appears to be very weak as compared to the achiral polarization combinations, but was included in the analysis nonetheless, because (as can be seen in Main Text Figure 3) the spectral calculations of various frames do lead to chiral signals, and including this polarization combination will thus make us more sensitive in our quest of finding which structures are present in our experiment. The sample stage and IR beam path were flushed with nitrogen to avoid artifacts due to IR light adsorption by water vapor.

The spectra are processed using a home-written Python script as follows. First the spectra are frequency-corrected by calibrating the frequency axis with a calibrated water vapor FTIR spectrum by overlapping the water vapor peaks in the 1600-1800 cm^{-1} with the sharp dips in an unpurged VSFG sample spectrum. Then, the cosmic rays are removed by detecting sudden jumps in the spectra when going to a subsequent pixel (threshold = 15 counts, which is sharp enough to identify cosmic rays, while not modifying the VSFG signal's shape, which has a maximum intensity of ~ 125 counts), and after how many pixels the photon count is again within the threshold, after which the intermediate pixels are interpolated. Then, the background spectra (recorded with blocked IR) are subtracted from the sample and the gold spectra. To compensate for minor remaining amounts of visible light due to beam pointing differences or sample evaporation during the experiments, which can create a linear offset when visible light that scatters from the trough bottom hits the detector, subsequently a

linear background (fitted to the average values of the first and last 50 pixels of the 1100-2200 cm^{-1} range of the CCD chip), is subtracted. In these experiments, this only yielded linear offsets of a few counts at the extremes of the CCD chip (while, again, the highest signals were ~ 125 counts). Finally, the processed sample spectra are normalized using a reference spectrum obtained from gold, which has been processed with the same procedure.

1.4 Molecular-dynamics simulation setup

Two systems, denoted S1 and S2, containing α S in the kinked α -helix ‘horseshoe’ conformation (PDB: 2KKW,^{S9} S1: Model 1 and S2: Model 11) were built using the CHARMM-GUI webserver.^{S10} Both S1 and S2 contains a bilayer with 269 DPPG lipids in each leaflet (538 total), 69626 water molecules, 739 Na^+ , and 192 Cl^- ions (0.15 M NaCl), resulting in system dimensions of ~ 13 nm x 13 nm x 16 nm. The α S horseshoe was manually placed in parallel with the bilayer (~ 1.5 nm above). The C-terminal pointed away from the bilayer surface and was not in contact with it in the starting position. All residues were in standard protonation states (pH 7.4) with H50 being neutral and in the δ -tautomer. Protein termini were uncapped and charged.

Molecular-dynamics simulations were performed using GROMACS^{S11} v2019.6 on Nvidia V100 GPUs at the Centre for Scientific Computing Aarhus (CSCAA). We employed a GROMACS port of the IDP protein force field a99SB-disp force field,^{S12} using the TIP4P-disp^{S13} water model, and Charmm22^{S14} ions, provided by Robustelli et al.^{S12} via D.E. Shaw Research. For the lipids we used the AMBER17 lipids,^{S15,S16} ported to GROMACS by Wu et al.^{S17} A non-bonded truncation cutoff of 1 nm was used and the particle mesh Ewald^{S18} method was used for long-range electrostatic summation. Frames were saved every 50 ps. S1 and S2 were first minimized and equilibrated in three steps: (1) 1 ns (1 fs timesteps) in the semiisotropic NPT ensemble using the Berendsen barostat^{S19} and v-rescaling thermostat^{S20} at 1 bar and 300 K with position restraints on the DPPG molecules and the protein backbone; (2) 4 ns (2 fs timesteps) using the same settings and hydrogen bonds restrained

using LINCS; (3) 40 ns (2 fs timestep) using same settings but now only position restraints on the protein backbone and using the Parrinello-Rahman barostat.^{S21} The protein did not come in contact with the bilayer during step (3) and stays ~ 1.5 nm above the surface. After equilibration of the DPPG bilayer, production runs of S1 and S2 were simulated for 400 ns with three repeat simulations of each using the same setting as step (3) but without position restraints on the protein.

There are a few limitations in our implementation of the MD simulations. First off, the DPPG bilayer contracts during equilibration, because the simulation temperature of ~ 300 K (chosen to mimic the room temperature in the VSFG experiments) is below the DPPG phase transition temperature of 314 K. This results in a slightly more rugged surface, which could influence lipid-protein interactions. Improvements to both the experimental and simulation setup would be to use bilayers (possibly vesicles to test dependence on curvature), with a more realistic neuronal lipid composition, including e.g. cholesterol. Secondly, the low simulation time of 400 ns (due to excessively large systems) leads to undersampling and non-equilibrium systems. And finally, the AMBER17 lipids have not been validated in combination with the a99SB-disp protein and water models. Since the LJ interactions between TIP4P-disp and the a99SB-disp protein parameters are increased, this could shift the relative interactions between protein and lipids, which together with low sampling, could explain why we mostly see interaction between the N-terminal region and the DPPG lipids. We stress that the MD simulations are used as a way to generate hypothesis conformations and the fit RSS of the VSFG spectra is used to extract an ensemble consistent with the VSFG experiment.

1.5 Spectral calculations

The spectral calculations are based on a excitonic Hamiltonian approach developed first described in ref. S22. Because they are based on the six 400 ns MD trajectories, with the two different starting structures (S1, S2), each with three (re)production runs, and

50 ps sampling, we performed the calculations for a total of 48,000 different frames, with a varying protein structure and orientation in each frame. Because we did not find an improvement of the spectral match when we averaged over multiple frames (see Table S3), all presented spectral calculation are performed by directly obtaining the eigenmodes from each frame, by evaluating the excitonic Hamiltonian model. The residual sum-of-squares (RSS) between the experimental and calculated spectra was used to find the frames that describe the experimental response best.

In the spectral calculations (also known as frequency mapping^{S23}), the exact spectral simulation parameters that are used can be very important for the outcome. The nearest-neighbor coupling was described using a dihedral-angle parameterized map of *ab initio* B3LYP/6-31G+(d) coupling from N-Methyldacetamide (NMA) deuterated at the amide group’s N atom. Transition dipole momenta were based on a map of Mulliken charges from an *ab initio* B3LYP/6-31G+(d) calculation. The frequency shift applied for the five proline residues in α S was $\delta\omega_{\text{proline}} = 19 \text{ cm}^{-1}$. The refractive indices used to describe the local-field corrections are listed in Table S2 for the bulk-air phase, bulk-water phase and the refractive index of the thin interfacial lipid/protein film. The interfacial refractive index is important in the description of $L_{zz}(\omega_j)$ in the Fresnel factor equations.^{S24,S25} It is a weak point in determining the local mode corrections, since it is particularly hard to determine with a consistent and independent approach, but the frequently-employed estimation of Zhuang et al.^{S24} has resulted in consistent molecular pictures in many studies.^{S25,S26}

Table S2: Refractive indices used for estimation of the local field corrections. Water data from ref.^{S27} and the interfacial refractive index from ref.^{S24}

	ω_{SF}	ω_{VIS}	ω_{IR}
n_1 (air)	1	1	1
n_2 (water)	1.331	1.329	1.315
n'	1.18	1.18	1.18

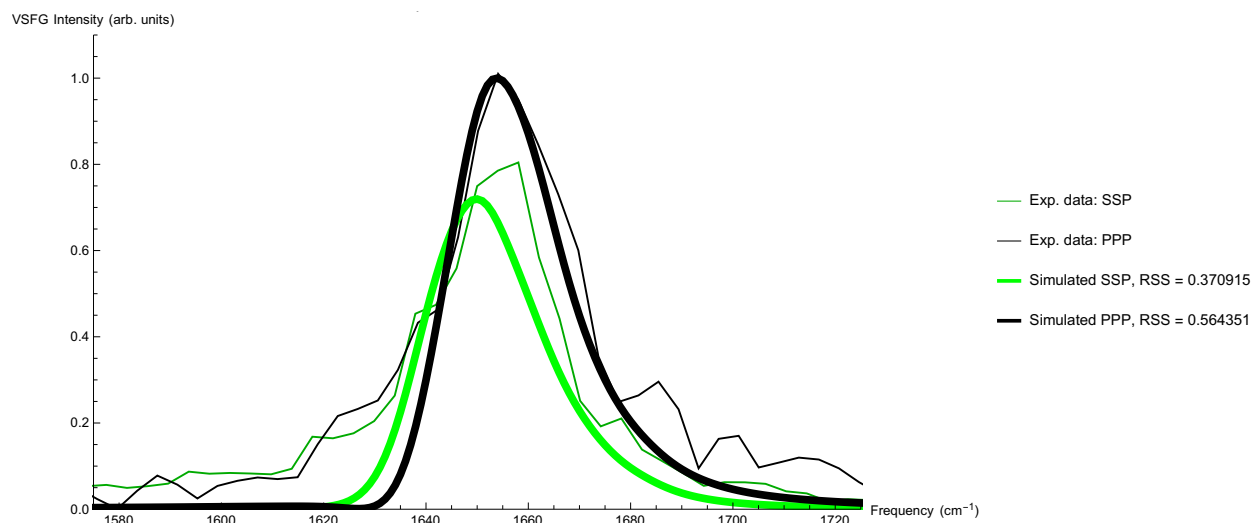


Figure S1: Experimental (rugged) and calculated (smooth) spectra of LK α 14, used to calibrate the local-mode frequency parameters of the spectral calculations.

1.5.1 Calibration of the spectral calculations with LK α 14

To calibrate the local-mode frequency parameters (the frequency of the uncoupled local modes and influence of hydrogenbonding on the local-mode frequency, estimated by a linear correlation with the C=O bond length^{S28,S29}), we compare experimental and calculated spectra of the LK α 14 peptide.^{S30} LK α 14 is a short peptide of leucine and lysine residues (Ac-LKKLLKLLKKLLKL). It is specifically engineered to fold into a stable α -helical structure at a hydrophilic/hydrophobic interface. LK α 14 is simulated for 300 ns in 3 replications on a water-hydrophobic (PG lipid-tails) interface. The experimental data is taken from ref. S30 of LK α 14 at a hydrophobic polystyrene interface in water. The reflections of the SF-, VIS- and IR- beams, as well as the different VIS-wavelength ($\lambda = 532$ nm) beams were compensated for with the appropriate local-field corrections. By optimizing the spectral parameters of the calculations to optimally match the SSP and PPP spectra experiments, we found the slope parameter of the hydrogen-bonding effect on the local-mode frequency to be $\delta\omega_{HB} = 400 \text{ cm}^{-1} \text{ \AA}^{-1} r_{CO}$ and isolated local-mode frequency Ω^0 was 1655 cm^{-1} (see Figure S1). The experimental LK α 14 data is recorded with a picosecond VSFG scanning setup in Seattle, as opposed to the α -synuclein data recorded in Aarhus from the femtosecond Astrella laser in the Weidner Lab. The linewidth for the two setups is different, so this parameter was therefore optimized for the α -synuclein spectrum. However, the best protein structure was found to be independent of the linewidth, within a range around the expected experimental linewidth (see Table S3).

1.5.2 Additional discussion of spectral match

The resulting spectra with the lowest RSS (see Figure 3) closely match the relative SSP and PPP intensities, as well as their lineshape. The SPS match is slightly less good, which can be understood from the fact that selection based on the total RSS of all polarization combinations results in a higher sensitivity to stronger polarization combinations. The lack of any (chiral) PSP intensity is reproduced well, and the transparent band that indicates the

bandwidth of the spectra calculated for the frames within the ensembles shows that including this polarization combination filters out structures that have a chiral response, contrary to the experimental observations. It is interesting to note that most deviations within the transparent band are in the 1660-1680 cm^{-1} region, where turns are thought to absorb IR light.^{S31} Also based on other recent results obtained with this Hamiltonian approach, we think that this is due to force-field imperfections that lead to too-steep turns. Happily, when taking the average of the structures in the ensemble the effect averages out, or at least has a negligible effect on the spectral lineshapes.

1.5.3 Additional discussion of spectral-calculation results for the MD simulations

A more detailed comparison of the molecular-dynamics (MD) simulations and the spectral-calculation derived RSS values reveals that, for all the trajectories, the kinked α -helical starting structure partly unfolds, resulting in shorter helical segments linked by disordered regions that lack a very strong correlation between helicity and RSS (see Figures 2(C) and S10). The α -helical contents (quantified by the DSSP algorithm^{S32}) generally decreases throughout the 400-ns trajectories. In some cases, very disordered structures with less than $\sim 30\%$ helical residues are found after 400 ns (Figure S10). The longest simulation time periods of conformations that result in VSFG spectra with low RSS values (i.e., with a close spectral match) are found at intermediate simulation times; from 50 to 350 ns. These low-RSS conformations thus have $\sim 30\text{-}60\%$ helical contents for both system S1 and S2 (see also Main Text Figure 4) and both for the high- and low-LPR experimental data. As the helical content decreases below $\sim 30\%$, the RSS values become slightly larger, indicating a less optimal match with the experimental structures (Figure S10). Interestingly, there is no clear correlation between the obtained RSS values and the helicity, indicating — as expected — that the VSFG response is sensitive to other structural aspects besides the secondary structure alone.

For S1, in one of the three trajectories, the N-terminus detaches from the surface, which causes it to probe a much larger part of the N-terminal-DPPG COM parameter spaces (see Figure S13). In system S2, the N-terminal remains close to the surface in all three repeats. The fact that one S1 repeat detaches from the surface could be related to the starting configuration where α S is positioned with only one lysine group that is close (within 1 nm) to the anionic DPPG membrane, while in the case of the initial S2 configuration, there are 6 lysine groups in close proximity to the lipids (see Figure S5). The lysine groups are known to be of high importance for the lipid-binding capability of α S^{S33} as they provide the main electrostatic attraction to the anionic DPPG lipids.

1.5.4 Additional discussion of spectral calculations for literature models

For 2KKW, for both LPRs, all available models within the PDB entry were tried in order to find the model with the best match. This resulted in model 26 for the low-LPR dataset, and in model 18 for the high-LPR dataset. For the structures (of α S fragment (residues 9-89) bound to 7:3 POPC/POPS small unilamellar vesicles^{S34}) provided by the Langen group, only the first structure out the identified ensemble was analyzed, as all of the structures were structurally very similar. Although a small part of the N-terminus and the full C-terminus is missing, we expect that this will only minutely influence the result, as random-coiled structures hardly contribute to VSFG signals.^{S25} For the low-LPR dataset, the best 2KKW orientation results in a reasonably extended protein orientation, and also the degree of helicity is not very different in the best-matching MD-simulation frames, which explains why the RSS of this structure and orientation is relatively low. Lorentzian widths of both 8 and 12 cm^{-1} have been tried, finding that a width of 12 cm^{-1} gave the best match. Most importantly, for the low-LPR dataset, the orientations reported in the associated publications do not appear to be consistent with our experimental results at all, given the very large RSS values at y rotation = $\sim 90^\circ$ for the horseshoe structure^{S9} and at both x and y rotation = 90° for the Langen structure,^{S34} while for the high-LPR dataset these orientations actually lead to

a close match, which corroborates the similar orientations derived with the frame-selection methodology described in the main text, thus benchmarking this novel approach.

1.6 Atomic-force microscopy

The AFM samples have been prepared by performing a Langmuir-Shaefer deposition of the sample surface on freshly cleaved mica, followed by drying for 10 min at the ambient environment. Subsequently, the sample on mica was imaged.

The AFM imaging was conducted on the Multimode[®] VIII AFM (Bruker, CA) in a Peakforce tapping mode. An silicon nitride probe Scanasyst-Air with the spring constant of 0.4 N/m was used under the ambient environment. The setpoint is less than 200 pN in order to protect sample. The raw images were processed by the software SPIP[™](Scanning Probe Image Processor) software package (Image Metrology, Denmark).

2 Additional calculation results

2.1 RSS values for various Lorentzian widths and frame-averaging sizes

Table S3: Table of key parameters from the investigation of the dependency of the best matching α -synuclein structure on the broadening parameter Γ . "Best structure" refers to the structure with the lowest RSS value. The half-width-at-half-max (HWHM) $\Gamma = 8 \text{ cm}^{-1}$ is used as it is the expected experimental linewidth of the visible pulse coming out of the etalon, but it is also observed that the conclusion from the calculation is stable within $\sim 2 \text{ cm}^{-1}$. The structure in frame 2599 is shown in main text figure 2(C). The best matching structures from $\Gamma = 6 - 12 \text{ cm}^{-1}$ are all very similar in structure and orientation.

$\Gamma [\text{cm}^{-1}]$	Lowest RSS	MD frame of best structure	RSS of frame 2599
6	1.04	2423	1.29
8	1.16	2599	1.16
10	1.10	2599	1.10
12	1.08	7710	1.53
16	1.32	397	5.45
20	1.62	526	5.22

Table S4: RSS as a function of the number of successive frames over which the local-mode frequencies (that are a function of the hydrogenbonds) are averaged, for each of the production runs.

Frames per segment	Min RSS s1_r1	Min RSS s1_r2	Min RSS s1_r3	Min RSS s2_r1	Min RSS s2_r2	Min RSS s2_r3
1	1.16	1.25	1.11	1.17	1.32	1.44
10	1.11	1.14	1.22	0.98	1.21	1.28
100	1.25	1.23	1.29	1.05	1.46	1.23

3 Additional experimental results

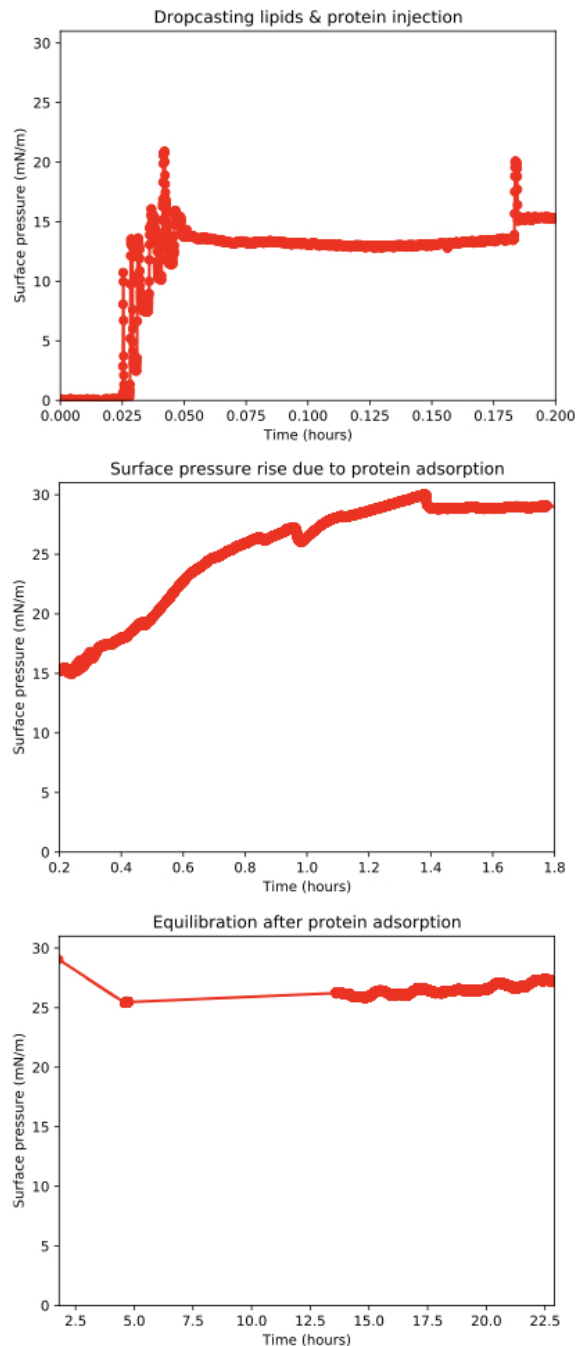


Figure S2: Surface-pressure as a function of incubation time. At $t = 0.025$ hr, the lipid monolayer was prepared by dropcasting a chloroform-based 1 mg/mL DPPG solution on the air-water interface. After equilibration for ~ 10 minutes, the high-concentration ($245 \mu\text{M}$) αS solution is injected, after which the VSGF experiments were started. The qualitative response of a fast initial surface-pressure rise and subsequent relaxation has been reproduced 4 times; in a similar fashion for LPR values of ~ 0.02 and ~ 16 .

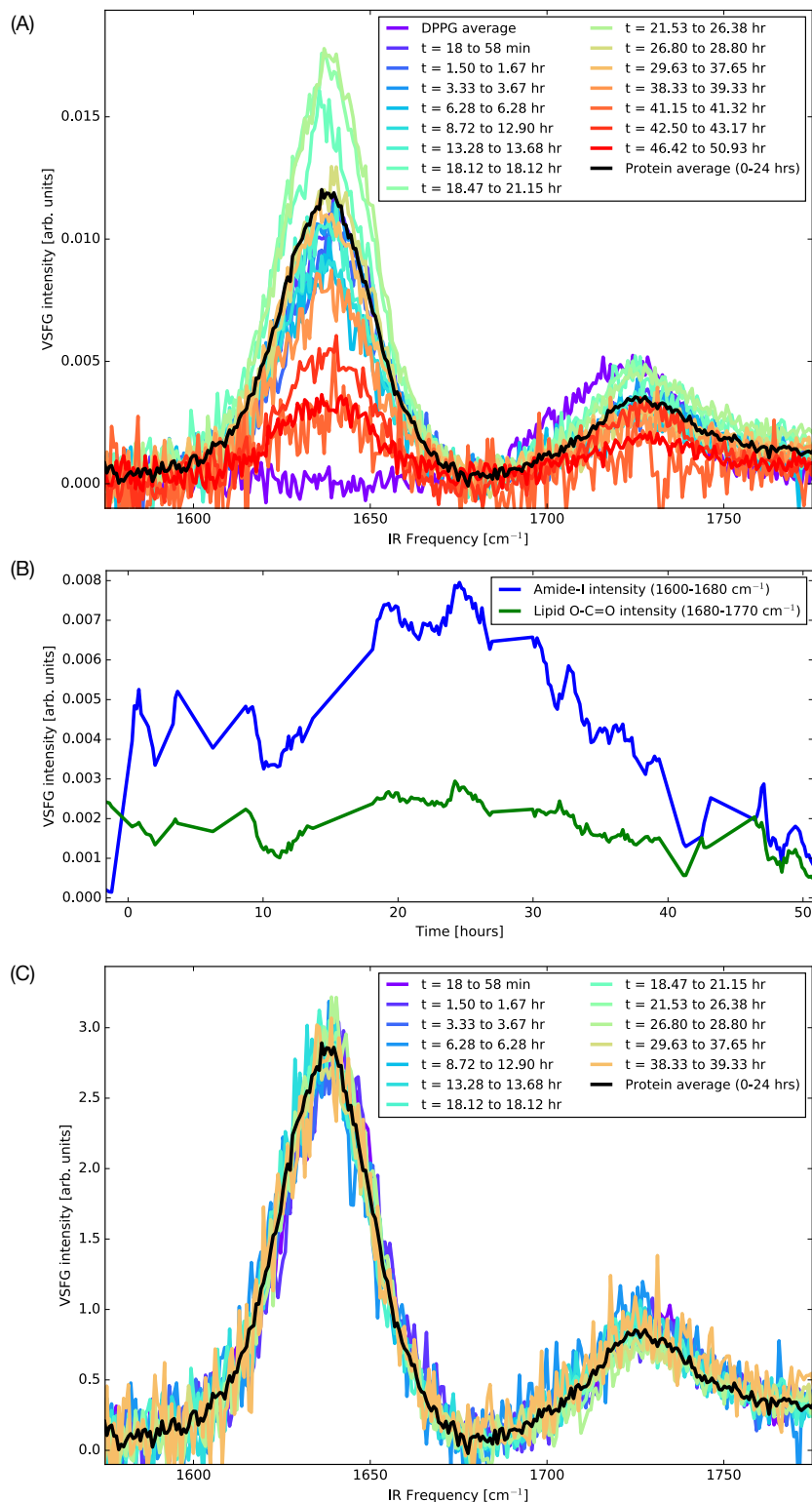
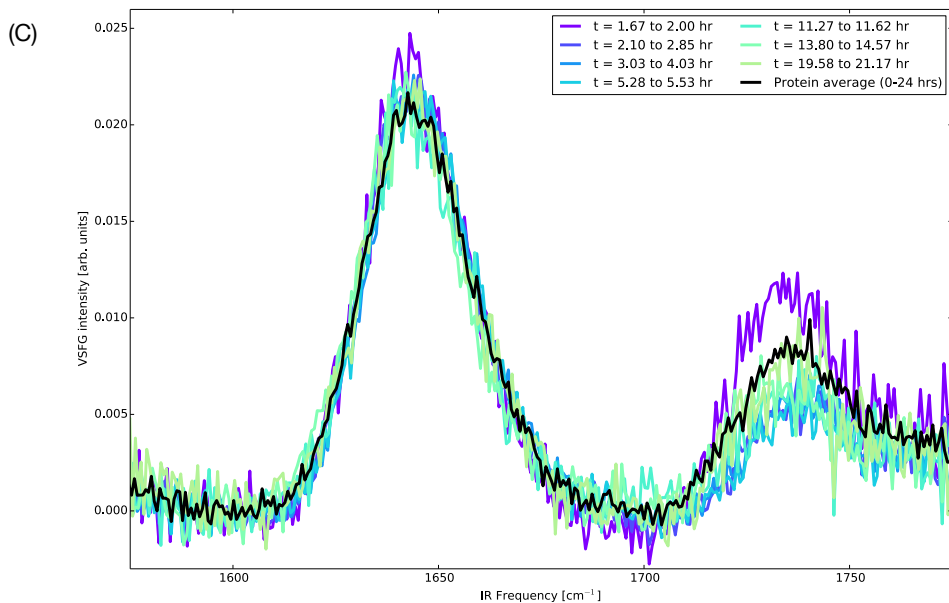
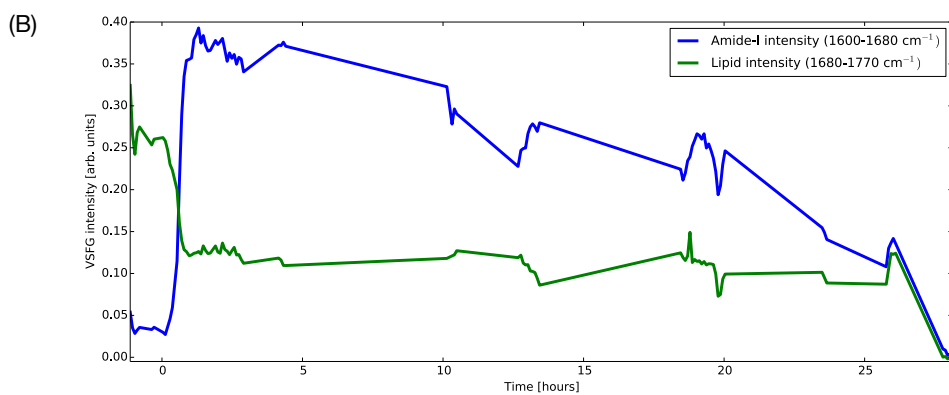
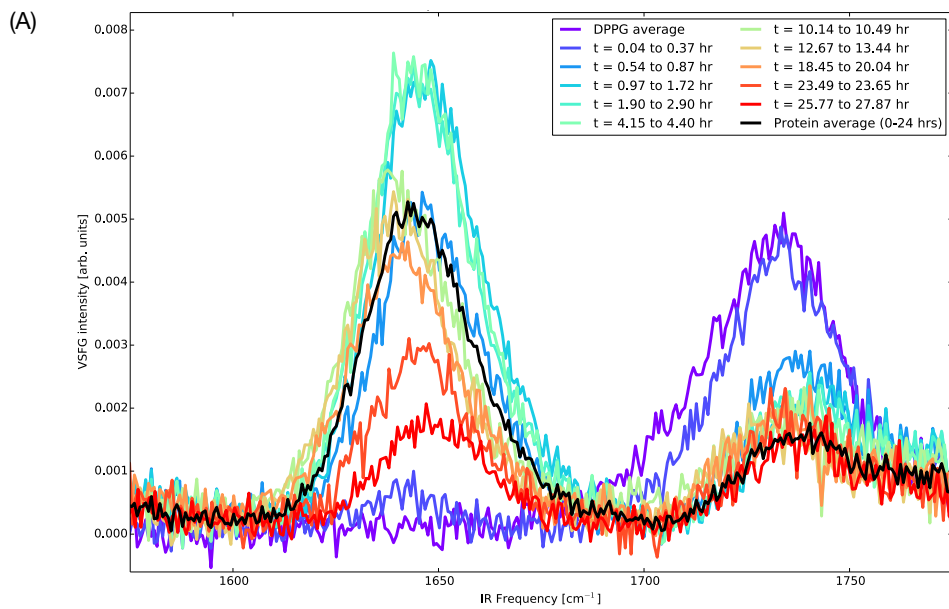


Figure S3: Intensity of the SSP polarization combination as a function of time for low LPR (this page) and high LPR (next page): (A) non-normalized, (B) amide-I and lipid-ester band intensity versus time, and (C) normalized spectra (the weak last 3 spectra from the time series are left out for clarity, but also those have an identical lineshape).



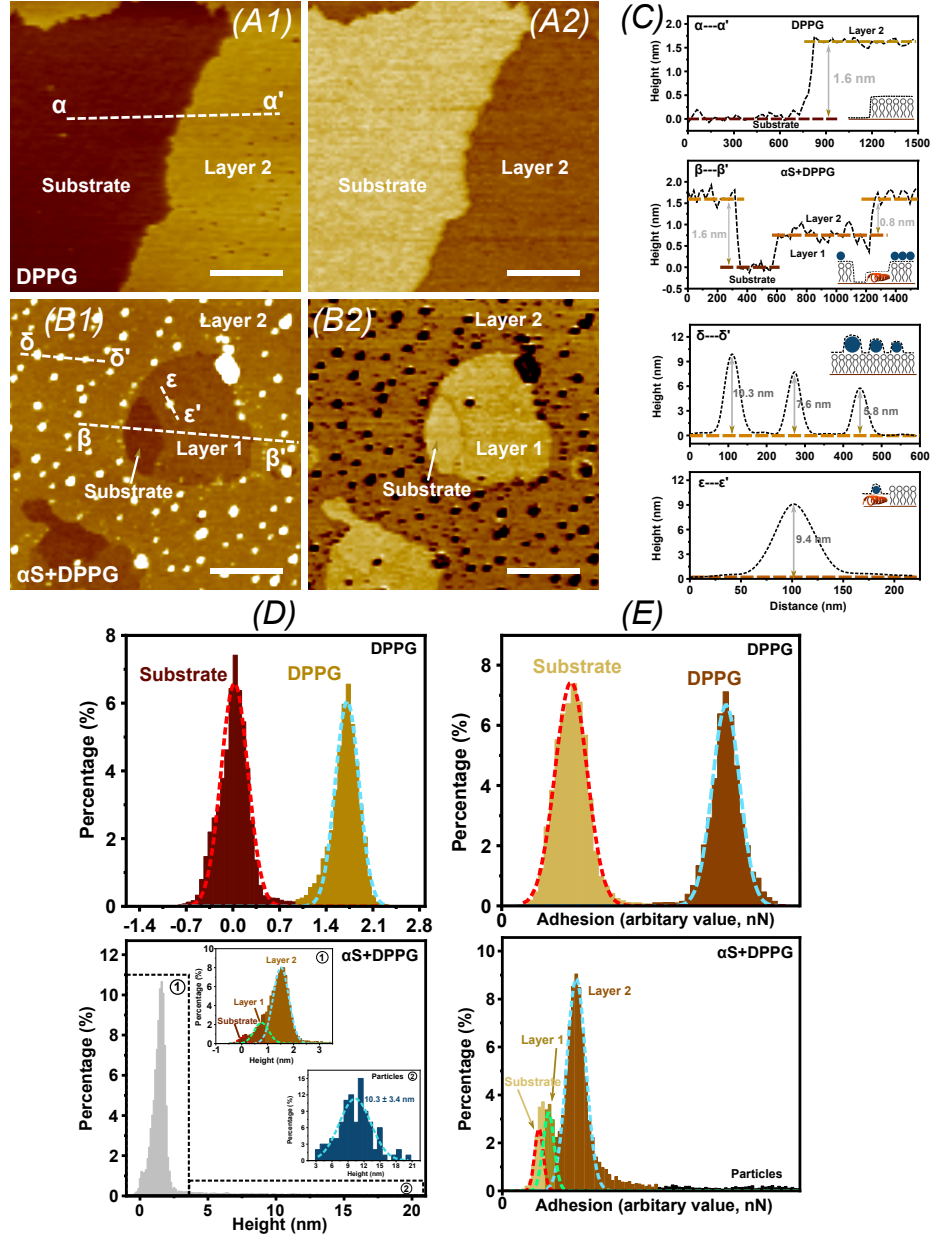


Figure S4: AFM images including defects, taken of the Langmuir-deposited interface (on freshly-cleaved mica) at the end of the VSFG experiment, which show that there are no fibrillar species present at the water-lipid interface. (A1,A2) Height and adhesion maps of a pure DPPG monolayer. Scale bars are 1 μ m. (B1,B2) Similar, but for a DPPG monolayer incubated for 24 hours with an α S solution at LPR = 16. As can be seen in the height profiles depicted in (C), the lipid monolayer is ~ 1.6 nm high (consistent with the lipid length found in the MD simulations), the monomeric protein layer is ~ 0.8 nm high (consistent with the diameter of an α -helix), while the oligomers (mostly to be found on the lipid layer, but there are also some found on the protein layer) are ~ 10 nm high. It is interesting to note that the largest oligomers are found on the edge of the areas where the lipids have been displaced by the proteins. (D,E) Height and adhesion histograms of the AFM image for DPPG (top) and α S-incubated DPPG (bottom) corroborate the layer assignment through the very different adhesion values for the lipid and the protein layers.

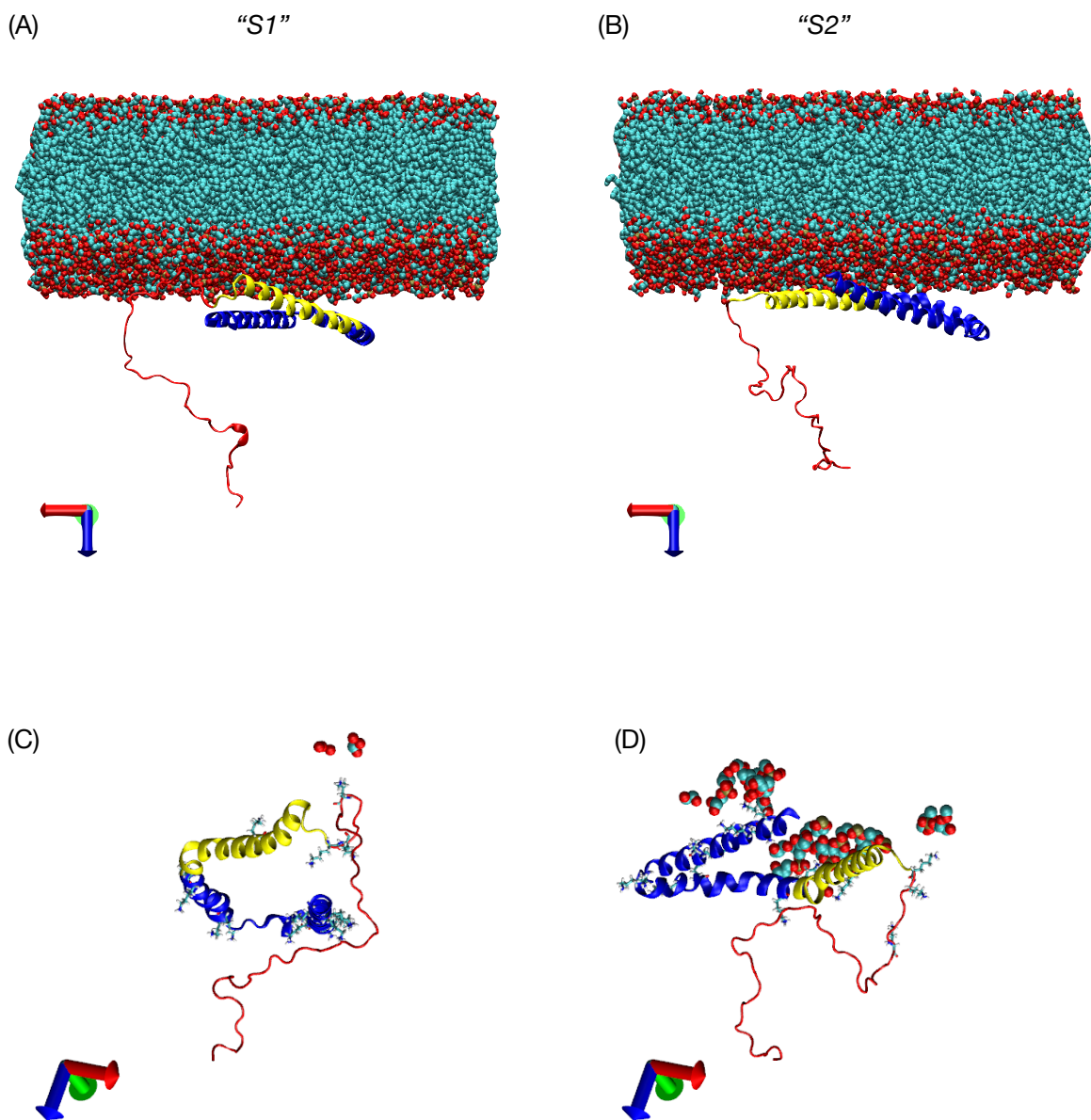


Figure S5: (A,B) Side view of the two starting structures prior to equilibration of the membrane. The two structures (based on PDB: 2KKW, model 1 and model 11, which have the C-terminus pointing in opposite direction with respect to the horseshoe plane) are topologically different, in the sense that the horseshoe structure interface with the membrane is flipped such that both sides of the helix are allowed to interact with the membrane in the two different starting conformations. (C,D) Depiction of the same starting structures, with the DPPG headgroup atoms within 10 Å of LYS residues shown as spheres. S2 has more LYS contact points (~ 6 vs. ~ 1 in S1). We see membrane detachment in one out of the three S1 trajectories, while three out of three S2 trajectories generally stay attached to the membrane (see Figure S13).

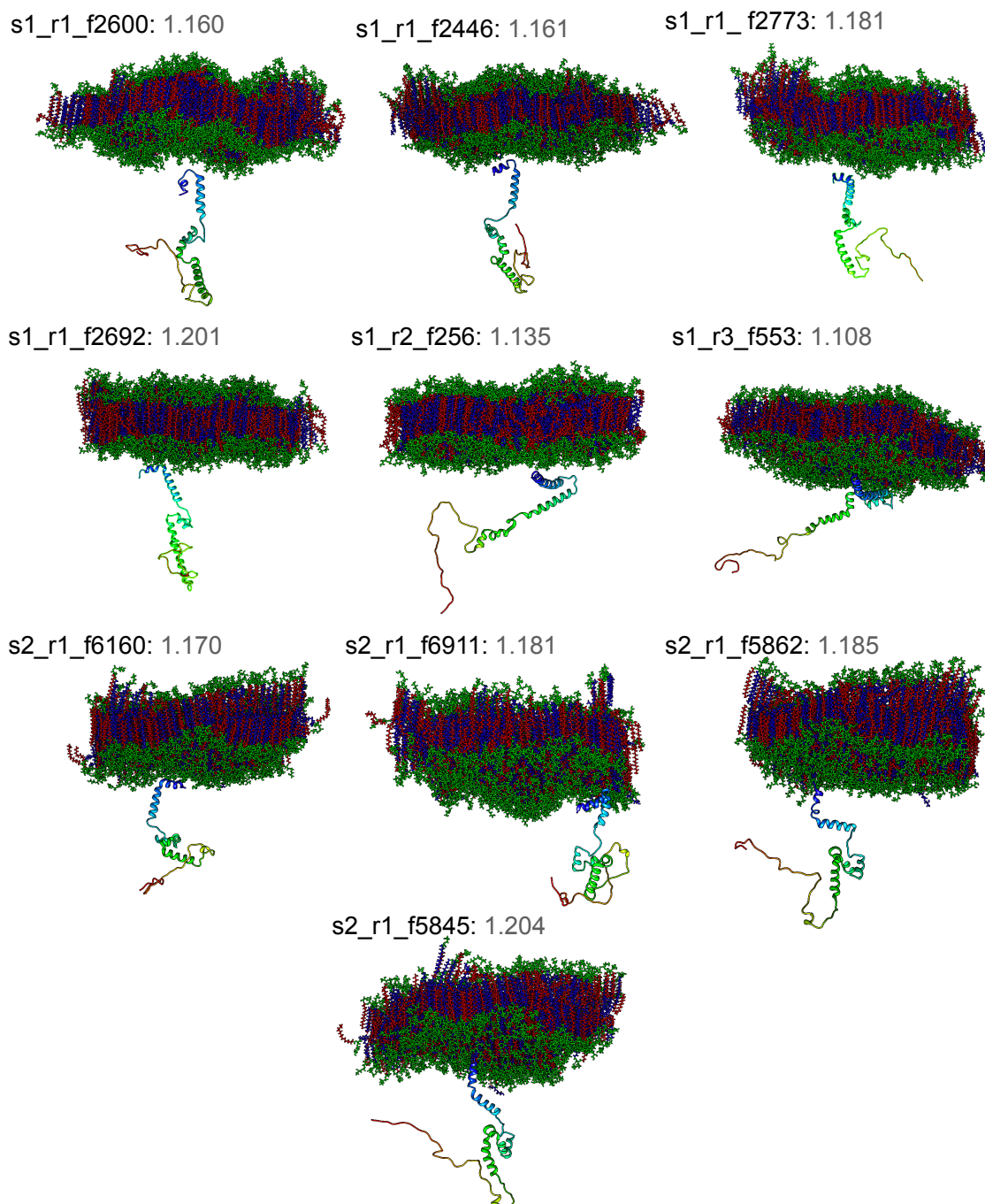
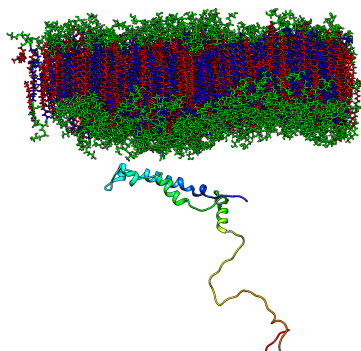
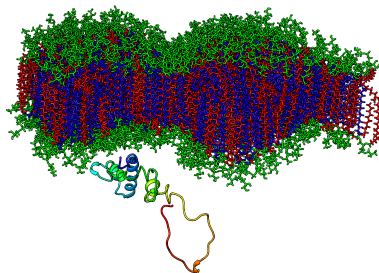


Figure S6: Overview of the 10 best-matching frames for the low-LPR (this page) and high-LPR (next page) experiments, as assessed by their RSS (the residual sum-of-squares between the calculated and experimental spectra), which is given in gray. As opposed to the Main Text figures, here, also the distal leaflets that are present in the simulations (see the MD description in the Methods sections) are shown. Rainbow coloring indicates sequence order, going from blue (N-terminus) to red (C-terminus). Interestingly, in the high-LPR case, the two structures that do not resemble the typical flat-lying helical structure very much (depicted on the lowest row on the next page) are also the two structures that contain an α S molecule that jumped to the next lipid leaflet (which is possible due to the periodic boundary constraints).

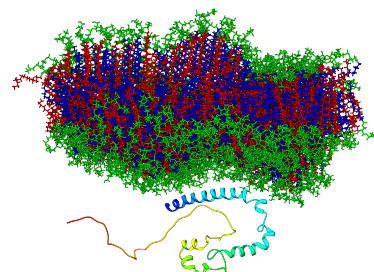
s1_r1_f5943: 1.260



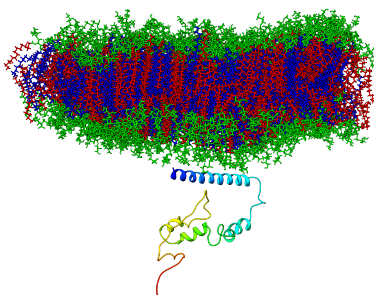
s1_r1_f6659: 1.155



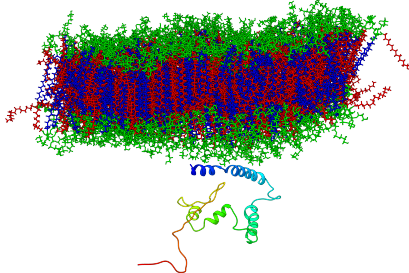
s2_r2_f1056: 1.422



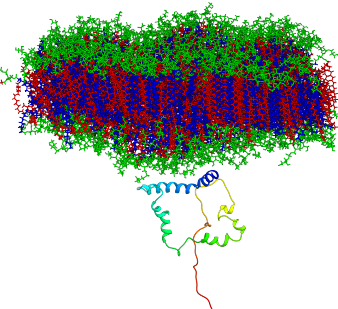
s2_r2_f1326: 1.279



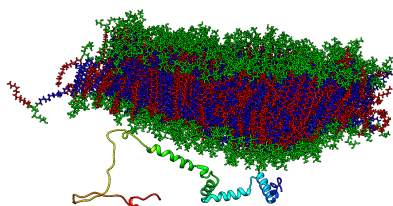
s2_r2_f1464: 1.563



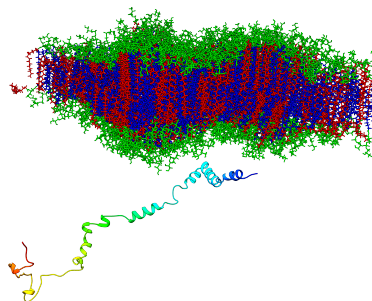
s2_r2_f2082: 1.428



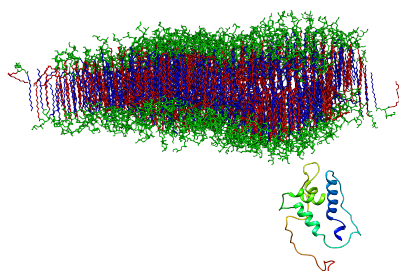
s2_r3_f1197: 1.558



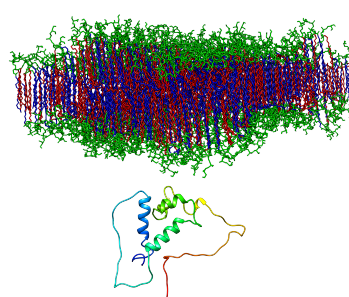
s2_r3_f2312: 1.252



s1_r3_f7655: 1.470



s1_r3_f7403: 1.500



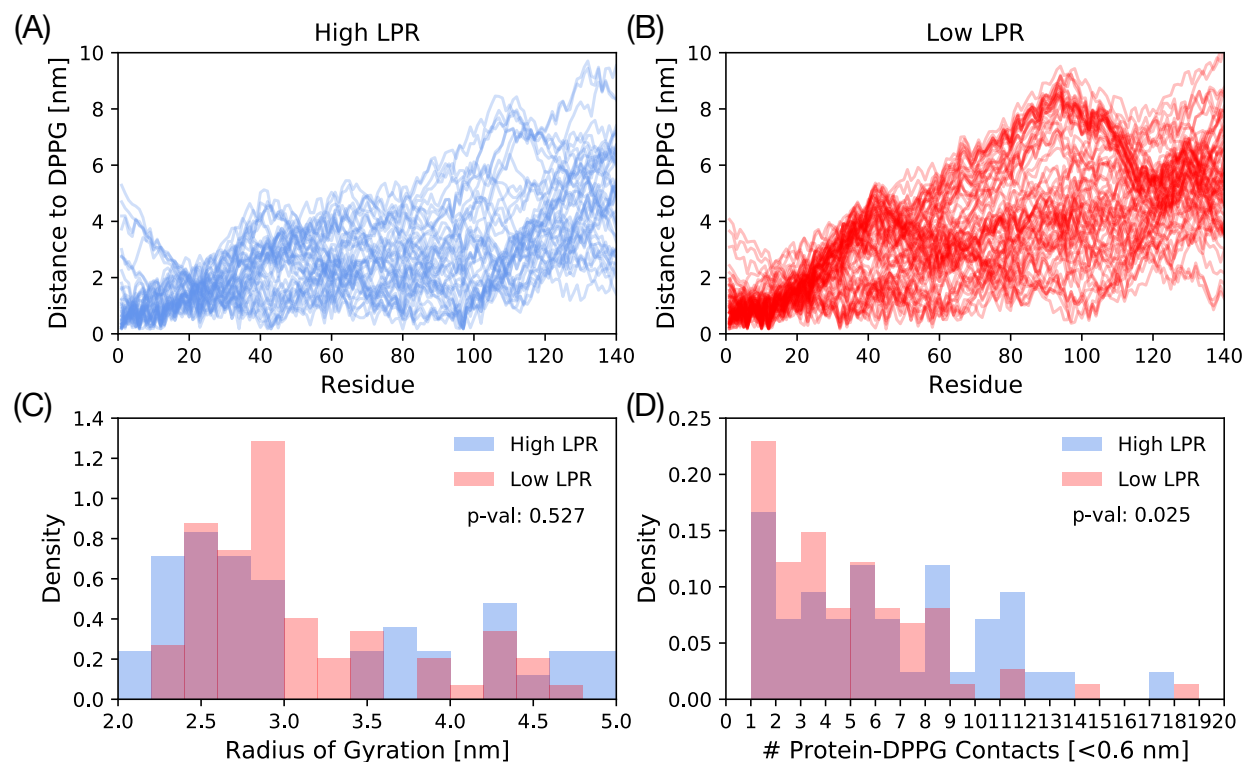


Figure S7: Individual traces for the distances to the nearest DPPG molecules for each of the frames within the high-LPR (A) and low-LPR (B) ensembles. (C) The radius of gyration is not significantly different ($p\text{-value} > 0.05$) as judged with a Mann-Whitney U test (implemented in Scipy v1.7.3, two-sided, exact mode). The number of protein-DPPG contacts, depicted in (D) is significantly different, indicating membrane interactions in a larger part of the protein at high LPR, as can be expected for the flat-lying structures within that ensemble.

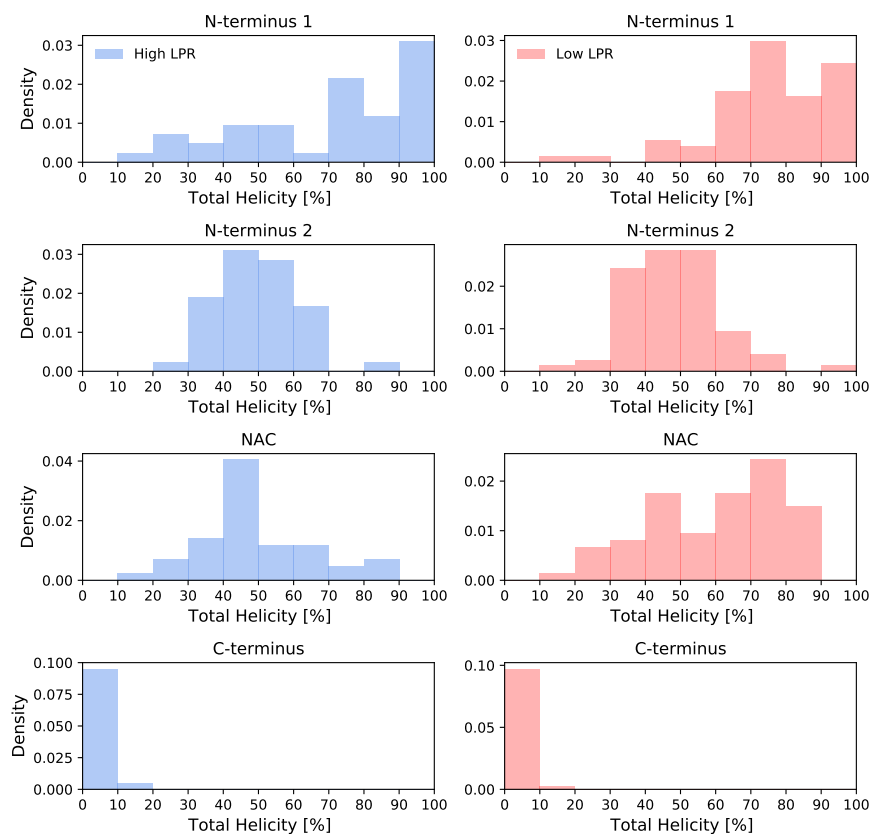


Figure S8: Helical contents of the high- and low-LPR ensembles, divided into protein regions: the N-terminus 1 (residue 1–30), the N-terminus 2 (residue 31–60), NAC region (residue 61–95) and C-terminus (residue 96–140).

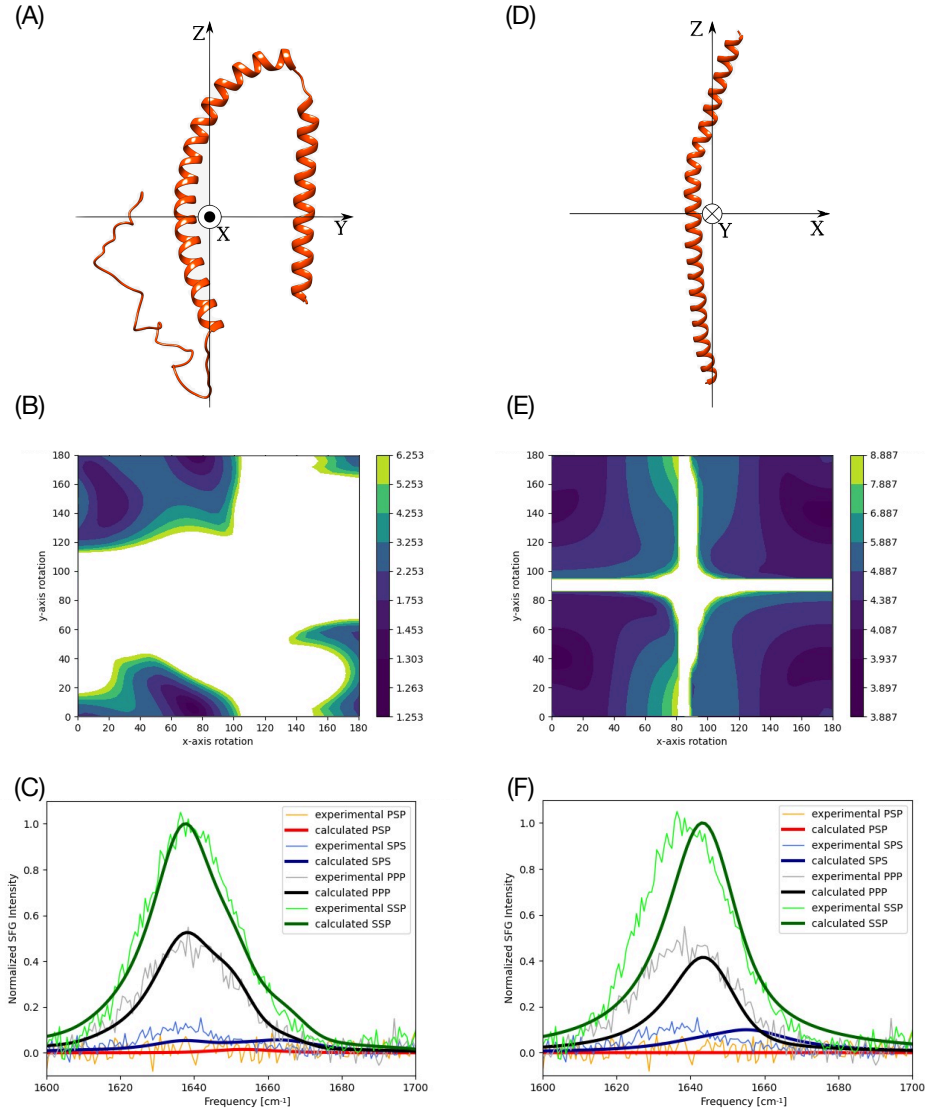
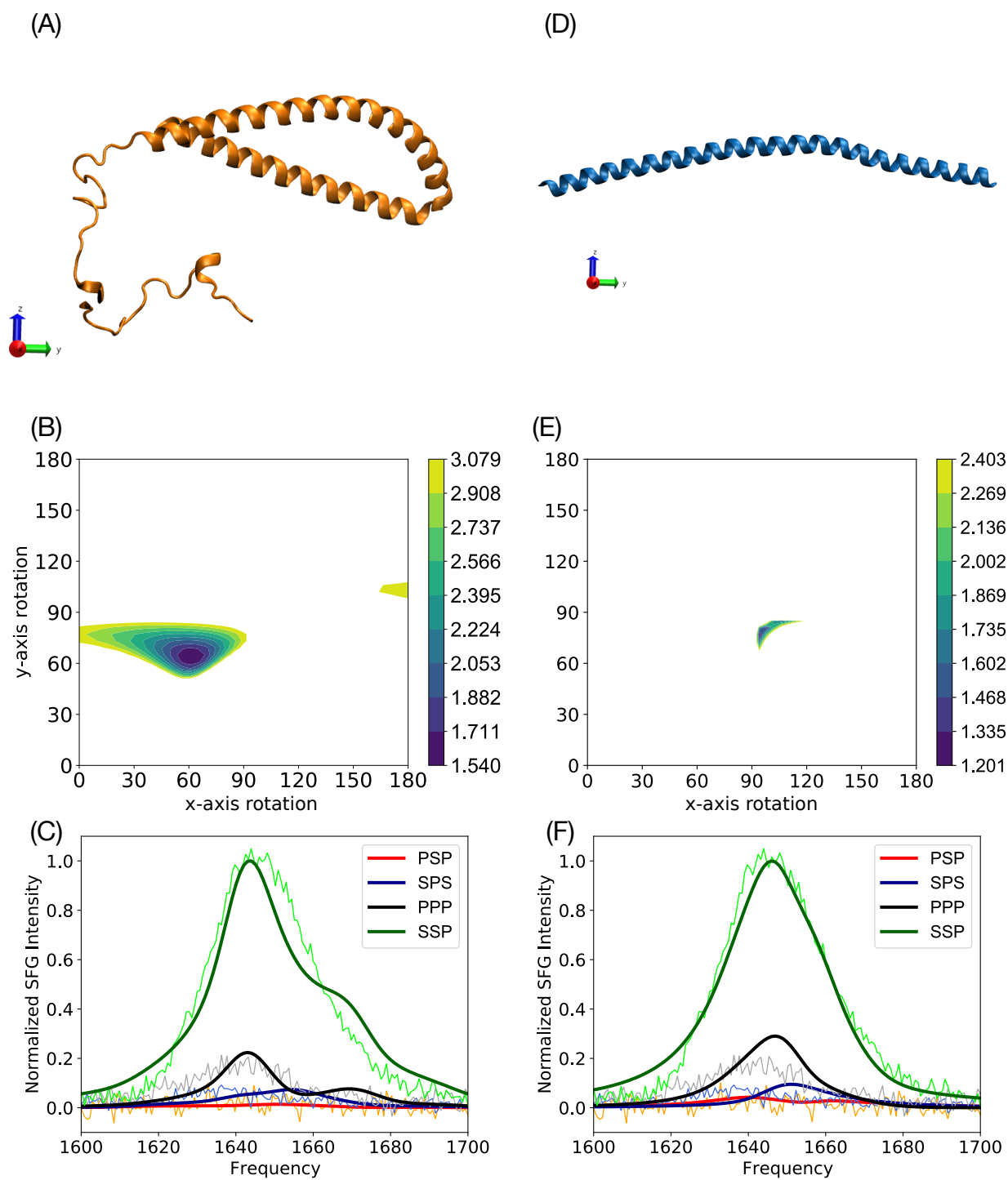


Figure S9: Structures and best-matching calculated VSFG spectra of two literature structures: low LPR (this page) and high LPR (next page). Left column: (A) model 26 of PDB 2KKW with (x,y)-axis rotations of $(0^\circ, 0^\circ)$, (B) the 2D-RSS plot as a function of the two rotation angles with a 4° resolution, and (C) the experimental and calculated spectra with a lowest RSS of 1.25 for an (x, y) rotation of $(73^\circ, 175^\circ)$. Right column: (D) model 1 of the structural ensemble provided the Langen group, rotated over the (x,y)-axes by $(0^\circ, 0^\circ)$, (E) the associated 2D-RSS plot as a function of the two rotation angles with a 4° resolution, and (F) the experimental and calculated spectra with a lowest RSS of 3.89 for an x, y rotation of $(175^\circ, 40^\circ)$. Same plots are shown on the next page, but for the match with the experimental high-LPR data, and with (A, B) the best orientations obtained for the best-performing model (#18) of PDB 2KKW (RSS = 1.540 for an (x, y) rotation of $(61^\circ, 65^\circ)$) and for model 1 of the Langen structure (RSS = 1.201 for an (x, y) rotation of $(94^\circ, 77^\circ)$), respectively. These results are discussed in detail in the ‘Spectral calculations’ section in the SI.



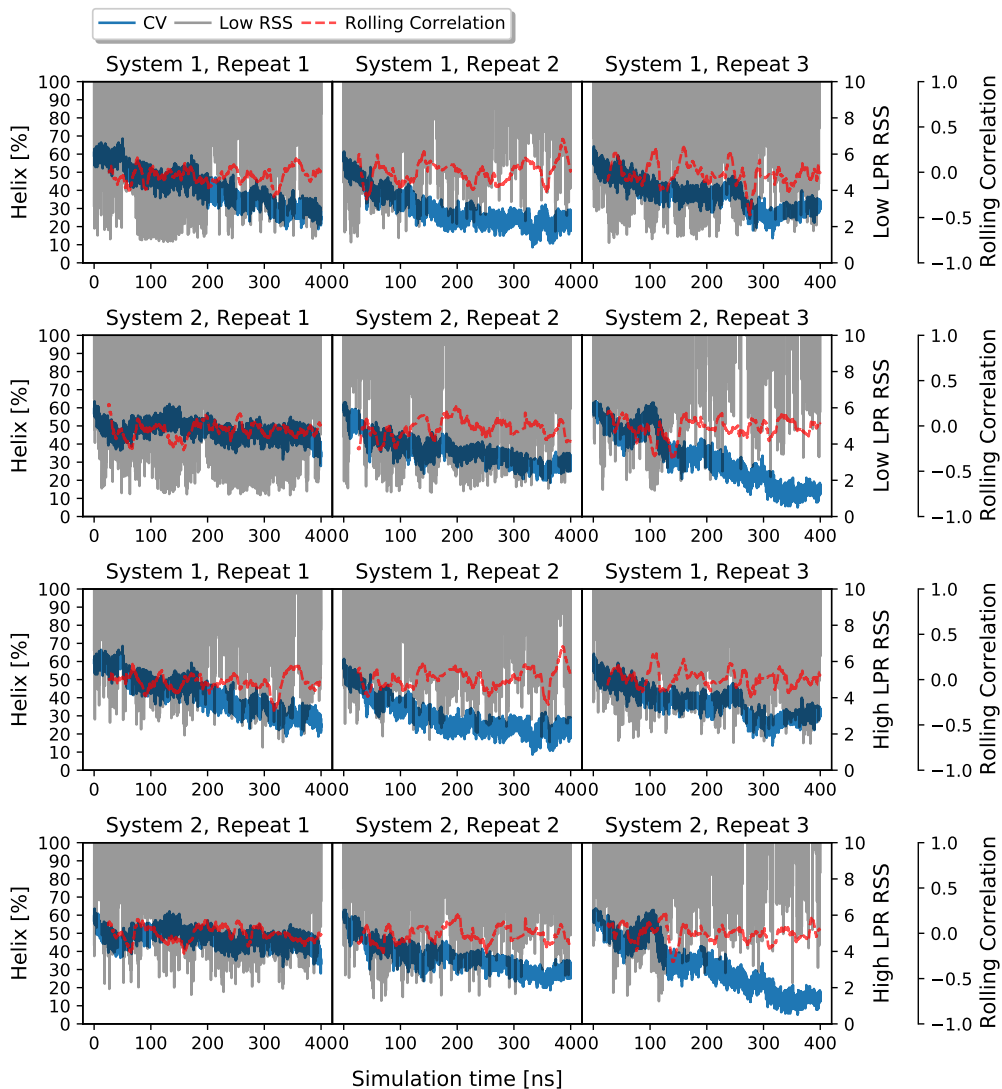
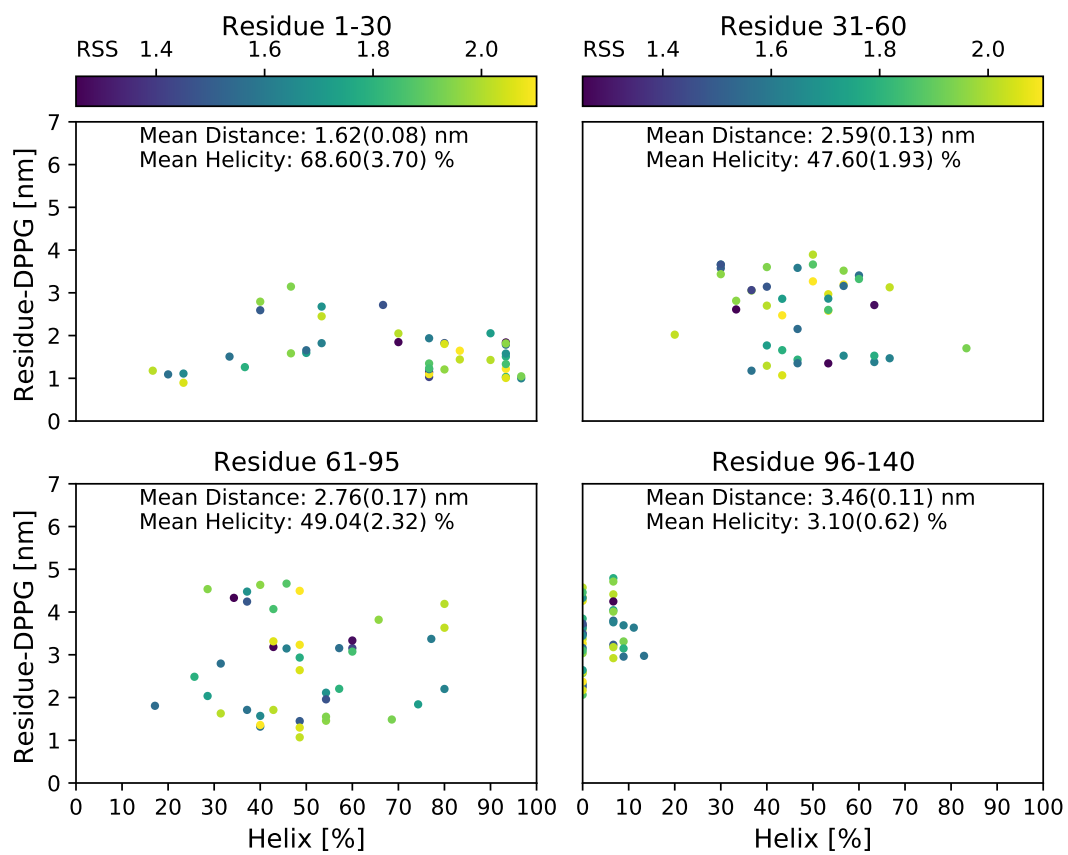


Figure S10: Low- and high-LPR RSS values plotted together with the fraction of residues found in a helix as a function simulation time for the 3 reproduction runs starting in the S1 conformation and in S2, along with the rolling correlation with a 25 ns window. The helical contents decreases throughout the trajectories, and in some of the trajectories (in particular in S1, repeat 2 and S2, repeat 3), very disordered structures are found at 400 ns with under $\sim 30\%$ of the residues adopting a helical structure. The helicity is calculated with the simplified DSSP scheme, MDTraj v1.9.4.



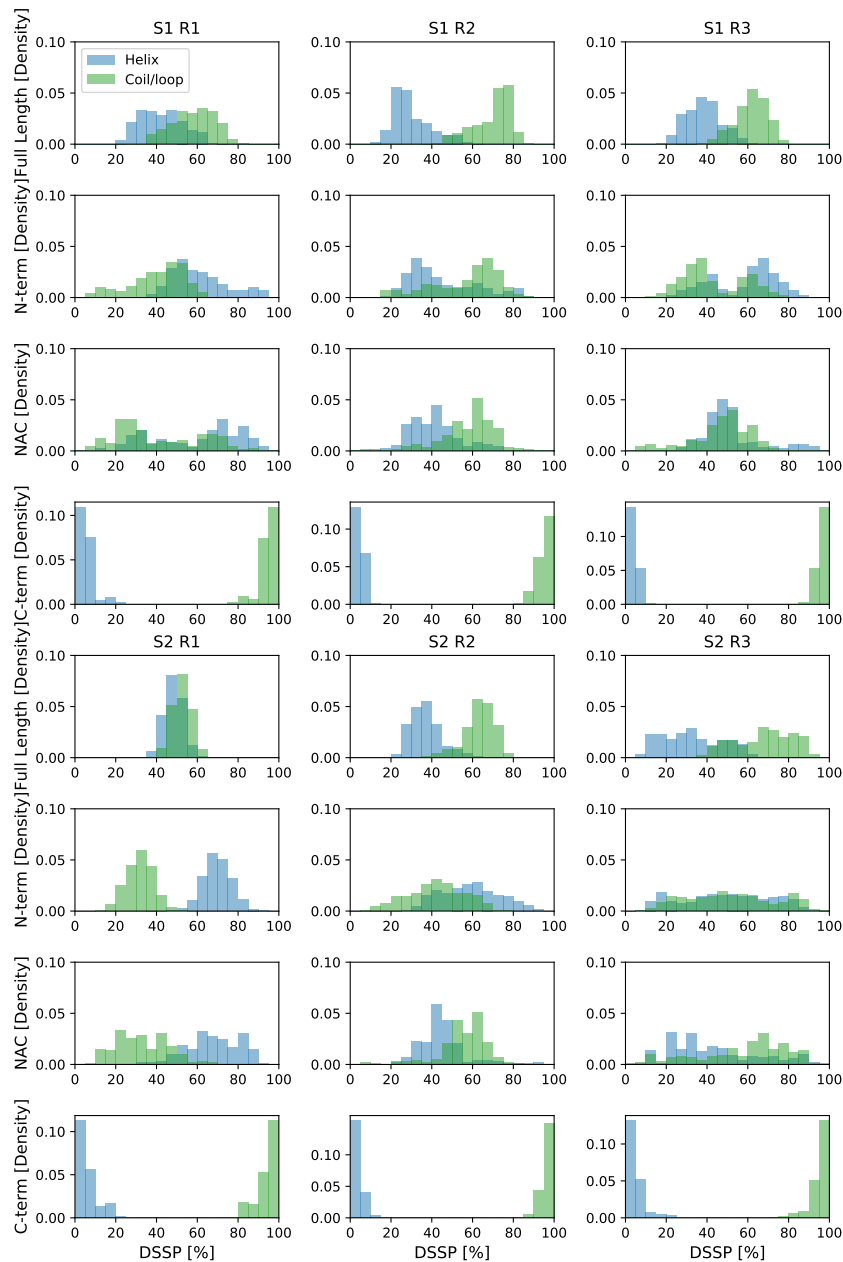


Figure S11: Secondary-structure contents during the 6 trajectories trajectories, specified for the full-length protein, the N-terminus (residue 1–60), NAC region (residue 61–95) and C-terminus (residue 96–140), with the top 4 rows = system S1 and the bottom 4 rows = system S2. Although the exact densities of secondary structure sometimes vary between the trajectories, the general trends are similar.

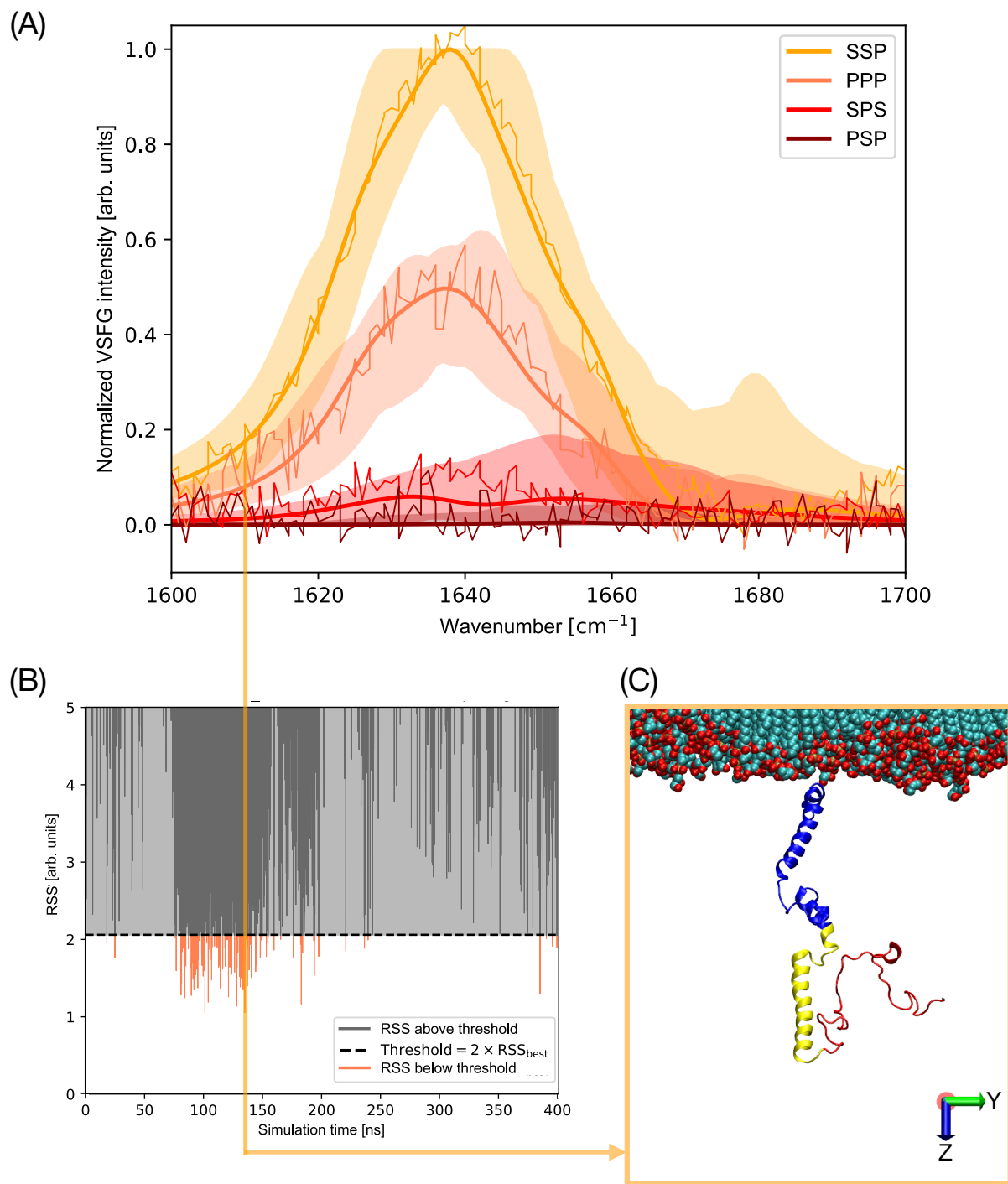


Figure S12: Modified, but equivalent figure to the low-LPR Figure 3 in the main text, but for a reproduction dataset where the 100 kDa filtration step was left out. The similarity between the two figures may corroborate the finding that the oligomers are VSFG silent, but the fact that similar kinetics are observed indicates also that there are few oligomers formed in the α S production process. Furthermore, it substantiates the robustness of the presented approach.

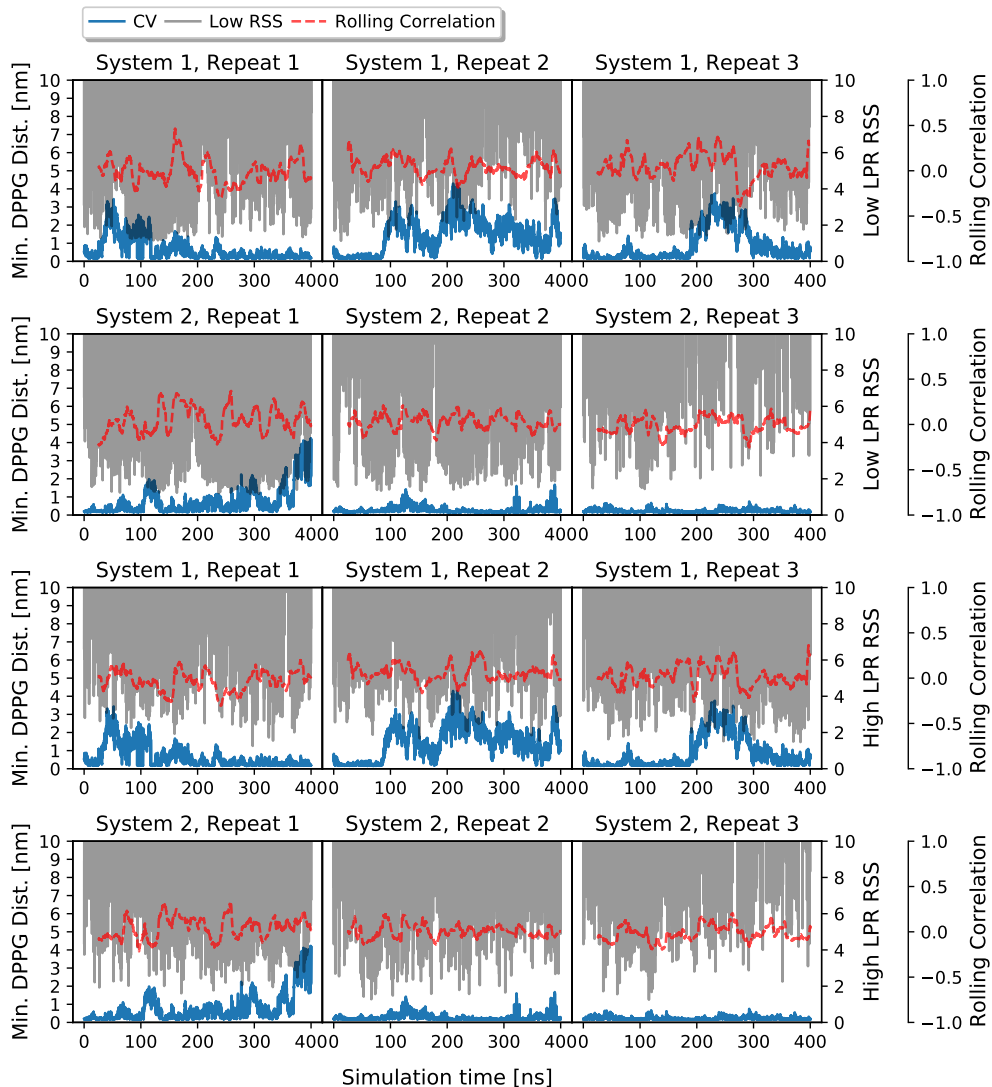


Figure S13: The ‘minimal distance to the DPPG monolayer’-collective variable (CV) for each of the 6 trajectories, indicating that for the S1 trajectories – started in an α S orientation with only one lysine contact – there is much more detachment (a much larger part of the CV parameter space is being probed) than for the S2 trajectories that have 6 lysine residues in close proximity of the DPPG layer at the start of the simulations.

References

- (S1) Mohammad-Beigi, H.; Shojaosadati, S. A.; Marvian, A. T.; Pedersen, J. N.; Klausen, L. H.; Christiansen, G.; Pedersen, J. S.; Dong, M.; Morshedi, D.; Otzen, D. E. Strong interactions with polyethylenimine-coated human serum albumin nanoparticles (PEI-HSA NPs) alter α -synuclein conformation and aggregation kinetics. *Nanoscale* **2015**, *7*, 19627–19640.
- (S2) Seki, T.; Chiang, K.-Y.; Yu, C.-C.; Yu, X.; Okuno, M.; Hunger, J.; Nagata, Y.; Bonn, M. The bending mode of water: A powerful probe for hydrogen bond structure of aqueous systems. *The journal of physical chemistry letters* **2020**, *11*, 8459–8469.
- (S3) Rodrigues, J. C.; Caseli, L. Incorporation of bacitracin in Langmuir films of phospholipids at the air-water interface. *Thin Solid Films* **2017**, *622*, 95–103.
- (S4) Nowotarska, S. W.; Nowotarski, K. J.; Friedman, M.; Situ, C. Effect of structure on the interactions between five natural antimicrobial compounds and phospholipids of bacterial cell membrane on model monolayers. *Molecules* **2014**, *19*, 7497–7515.
- (S5) Hellstrand, E.; Grey, M.; Ainalem, M.-L.; Ankner, J.; Forsyth, V. T.; Fragneto, G.; Haertlein, M.; Dauvergne, M.-T.; Nilsson, H.; Brundin, P.; Linse, S.; Nylander, T.; Sparr, E. Adsorption of α -Synuclein to Supported Lipid Bilayers: Positioning and Role of Electrostatics. *ACS Chemical Neuroscience* **2013**, *4*, 1339–1351, PMID: 23823878.
- (S6) Chakraborty, A.; Hui, E.; Waring, A. J.; Dhar, P. Combined effect of synthetic protein, Mini-B, and cholesterol on a model lung surfactant mixture at the air–water interface. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2016**, *1858*, 904–912.
- (S7) Fallah, M. A.; Gerding, H. R.; Scheibe, C.; Drescher, M.; Karreman, C.; Schildknecht, S.; Leist, M.; Hauser, K. Simultaneous IR-spectroscopic observation of α -synuclein, lipids, and solvent reveals an alternative membrane-induced oligomerization pathway. *ChemBioChem* **2017**, *18*, 2312–2316.

- (S8) Golbek, T. W.; Schmäser, L.; Rasmussen, M. H.; Poulsen, T. B.; Weidner, T. Lalicid acid antibiotic at a membrane surface probed by sum frequency generation spectroscopy. *Langmuir* **2020**, *36*, 3184–3192.
- (S9) Rao, J. N.; Jao, C. C.; Hegde, B. G.; Langen, R.; Ulmer, T. S. A combinatorial NMR and EPR approach for evaluating the structural ensemble of partially folded proteins. *Journal of the American Chemical Society* **2010**, *132*, 8657–8668.
- (S10) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of computational chemistry* **2008**, *29*, 1859–1865.
- (S11) Berendsen, H. J.; van der Spoel, D.; van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Computer physics communications* **1995**, *91*, 43–56.
- (S12) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences* **2018**, *115*, E4758–E4766.
- (S13) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *The journal of physical chemistry B* **2015**, *119*, 5113–5123.
- (S14) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B* **1998**, *102*, 3586–3616, PMID: 24889800.

- (S15) Dickson, C. J.; Madej, B. D.; Skjevik, Å. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. Lipid14: the amber lipid force field. *Journal of chemical theory and computation* **2014**, *10*, 865–879.
- (S16) AMBER17 lipids- distributed via AMBER20 MD Package, accessed October 2020. <https://ambermd.org/doc12/Amber20.pdf>.
- (S17) GMXlipid17.ff: Gromacs Port of the amber LIPID17 force field, October 2020, DOI: 10.5281/zenodo.3610470. <https://zenodo.org/record/3610470>.
- (S18) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N^*\log(N)$ method for Ewald sums in large systems. *The Journal of chemical physics* **1993**, *98*, 10089–10092.
- (S19) Berendsen, H. J.; Postma, J. v.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics* **1984**, *81*, 3684–3690.
- (S20) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of chemical physics* **2007**, *126*, 014101.
- (S21) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **1981**, *52*, 7182–7190.
- (S22) Roeters, S. J.; van Dijk, C. N.; Torres-Knoop, A.; Backus, E. H. G.; Campen, R. K.; Bonn, M.; Woutersen, S. Determining In Situ Protein Conformation and Orientation from the Amide-I Sum-Frequency Generation Spectrum: Theory and Experiment. *The Journal of Physical Chemistry A* **2013**, *117*, 6311–6322, PMID: 23566310.
- (S23) Baiz, C. R.; Błasiak, B.; Bredenbeck, J.; Cho, M.; Choi, J.-H.; Corcelli, S. A.; Dijkstra, A. G.; Feng, C.-J.; Garrett-Roe, S.; Ge, N.-H.; Hanson-Heine, M. W. D.; Hirst, J. D.; Jansen, T. L. C.; Kwac, K.; Kubarych, K. J.; Londergan, C. H.; Maekawa, H.; Reppert, M.; Saito, S.; Roy, S.; Skinner, J. L.; Stock, G.; Straub, J. E.;

- Thielges, M. C.; Tominaga, K.; Tokmakoff, A.; Torii, H.; Wang, L.; Webb, L. J.; Zanni, M. T. Vibrational Spectroscopic Map, Vibrational Spectroscopy, and Intermolecular Interaction. *Chemical Reviews* **2020**, *120*, 7152–7218, PMID: 32598850.
- (S24) Zhuang, X.; Miranda, P. B.; Kim, D.; Shen, Y. R. Mapping molecular orientation and conformation at interfaces by surface nonlinear optics. *Phys. Rev. B* **1999**, *59*, 12632–12640.
- (S25) Hosseinpour, S.; Roeters, S. J.; Bonn, M.; Peukert, W.; Woutersen, S.; Weidner, T. Structure and dynamics of interfacial peptides and proteins from vibrational sum-frequency generation spectroscopy. *Chemical reviews* **2020**, *120*, 3420–3465.
- (S26) Roeters, S. J.; Golbek, T. W.; Bregnhøj, M.; Drace, T.; Alamdari, S.; Roseboom, W.; Kramer, G.; Šantl-Temkiv, T.; Finster, K.; Pfaendtner, J.; Woutersen, S.; Boesen, T.; Weidner, T. Ice-nucleating proteins are activated by low temperatures to control the structure of interfacial water. *Nature communications* **2021**, *12*, 1–9.
- (S27) Hale, G. M.; Querry, M. R. Optical Constants of Water in the 200-nm to 200- μ m Wavelength Region. *Appl. Opt.* **1973**, *12*, 555–563.
- (S28) Ham, S.; Kim, J.-H.; Lee, H.; Cho, M. Correlation between electronic and molecular structure distortions and vibrational properties. II. Amide I modes of NMA–n D 2 O complexes. *The Journal of chemical physics* **2003**, *118*, 3491–3498.
- (S29) Lu, H.; Schäfer, A.; Lutz, H.; Roeters, S. J.; Lieberwirth, I.; Muñoz-Espí, R.; Hood, M. A.; Bonn, M.; Weidner, T. Peptide-controlled assembly of macroscopic calcium oxalate nanosheets. *The journal of physical chemistry letters* **2019**, *10*, 2170–2174.
- (S30) Weidner, T.; Castner, D. G. SFG analysis of surface bound proteins: a route towards structure determination. *Physical Chemistry Chemical Physics* **2013**, *15*, 12516–12524.

- (S31) Barth, A. Infrared spectroscopy of proteins. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **2007**, *1767*, 1073–1101.
- (S32) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Research on Biomolecules* **1983**, *22*, 2577–2637.
- (S33) Zarbiv, Y.; Simhi-Haham, D.; Israeli, E.; Elhadi, S. A.; Grigoletto, J.; Sharon, R. Lysine residues at the first and second KTKEGV repeats mediate α -Synuclein binding to membrane phospholipids. *Neurobiology of disease* **2014**, *70*, 90–98.
- (S34) Jao, C. C.; Hegde, B. G.; Chen, J.; Haworth, I. S.; Langen, R. Structure of membrane-bound α -synuclein from site-directed spin labeling and computational refinement. *Proceedings of the National Academy of Sciences* **2008**, *105*, 19666–19671.