Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Electronic Supplementary Information (ESI) Reconciling Membrane Protein Simulations with Experimental DEER Spectroscopy Data

Shriyaa Mittal[a], Soumajit Dutta[b], and Diwakar Shukla[*,a,b,c,d,e]

[*]Correspondence:
[a]Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL USA. E-mail: diwakar@illinois.edu; Tel: +1 217-300-0021
[b]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL USA
[c]National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, IL USA
[d]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL USA
[e]NIH Center for Macromolecular Modeling and Bioinformatics, University of Illinois at Urbana-Champaign, Urbana, IL USA

LIST OF TABLES

LIST OF FIGURES

METHODS

MD Simulations

All simulations were setup up using CHARMM-GUI (5–9), built with a rectangular box and a minimum water height of 10 Å above and below the membrane. System-specific details are provided below. All simulations were run using NAMD 2.13 MD package (10) and the CHARMM 36 force field (11–14) on the Blue Waters petascale computing facility. We used the NAMD inputs generated by CHARMM-GUI for minimization and equilibration in six consecutive steps followed by production runs in the NPT ensemble. The constant temperature was maintained by employing Langevin dynamics with a damping coefficient of 1 ps^{-1}. The Langevin piston method was employed to maintain a constant pressure of 1.0 atm with a piston period of 50 fs. Nonbonded interactions were smoothly switched off at 8–10 Å and long-range electrostatic interactions were calculated using the particle mesh Ewald (PME) method. For all simulation steps, bond distances involving hydrogen atoms were fixed using the SHAKE algorithm. Minimization was done for 10 000 steps, total equilibration and production run time for individual simulations are noted in Table S1. Production simulations were run at 303.15 K, trajectory parameters were determined every 2 fs, and coordinates were saved every 100 ps. Length of individual simulation datasets vary between ~18-32 μs simulation time (Table S1). All trajectory analysis was done using MDTraj 1.7 (1) except where otherwise noted. Analysis methods and workflows are explained in the supplementary material.

Determining a micelle size for membrane protein simulations

Previous work on simulating protein-micelle complexes (15) posits the use of the number of detergents more than the aggregation number of a detergent-only micelle that is 135-145 for the n-Dodecyl-β-D-Maltoside (BDDM) detergent (16, 17). Moreover, the BDDM micelle size was determined to be 72 kDa (18); with a 510.621 g/mol molecular weight of a BDDM molecule (19) this yields ~141 detergent molecules in the micelle. To test the stability of the protein-micelle complexes, we took a single structure of the $PepT_{So}$ protein from our previous simulations (20) and embedded it in 150, 180, and 200 BDDM detergent molecules. The three simulation setups with 145 668, 145 847, and 146 061 atoms respectively comprised of protein, detergents, waters, and 0.15 M KCl ions. Simulation setup with 150 detergent molecules was minimized for 10 000 steps and the other two were minimized for 20 000 steps. We ran each of these simulations for 60 ns each post-equilibration and only the last 50 ns were used for analysis (Table S2) to assess the protein's structure and dynamics in all three micelle sizes. RMSD of the protein converges to values between 0.28-0.32 nm within 50 ns, and these values are lower when only the transmembrane region of the protein is included (Fig. S2A). We do not expect to sample any conformational change in the protein's structure in such short trajectories.

We then evaluated the extent of sphericity of the micelle, measured by calculating its eccentricity where a perfectly spherical object has eccentricity 0. We find that in all three cases, micelles in our simulations are spherical with an average eccentricity of 0.23 (±0.02) for micelle with 150 and 200 detergents and 0.22 (±0.02) for micelle formed by 180 detergent molecules (Fig. S2C). The radii of the micelles do not show much variation, indicating that the micelles do not distort (Fig. S2D). As expected, micelles with more detergents have a larger average radius - 4.4 nm, 4.58 nm, and 4.68 nm for 150, 180, and 200 detergent micelles, respectively. Fig. S2E shows a radial distribution of distances between BDDM detergent molecule headgroups. Since the distribution is the same for all three micelle sizes, we conclude that detergent packing is similar in all three micelles.

Our preliminary simulations indicated that protein dynamics and shape of detergent micelle do not vary with the number of detergents in the micelle. We chose 150 detergent micelle for the rest of our simulations to keep the system sizes smaller and conserve computational resources. We were also able to confirm that the simulation setup is stable for all three micelle size choices.

Setting up LeuT simulations in bilayer and detergent micelle

We compiled 28 LeuT crystal structures among which all but two have no mutations in the proteins sequence. Structures 3TT1, an Outward Facing (OF) structure with two mutations, and 3TT3, an Inward Facing (IF) structure with four mutations (21), were modeled on the wild-type LeuT sequence using Modeller (22) interface in Chimera (23). A recent study by Gotfryd *et al.* reported a LeuT IF-occluded conformation without any mutations (24) but the study and the structure were released after the simulations in this work had been performed. Based on these PDBs, we identified 36 unique structural models for the LeuT protein from residue Arg5 to Ala513 (509 residues in all). Most of these 36 structures were missing residues either on EL2, EL3, or EL6, and the size of the largest missing region in any structure was six residues. These missing regions were modeled to yield 72 LeuT structures as

a starting point for our simulations. These structures were aligned in VMD (25) using orient and a linear algebra Tcl package, La. During setup in CHARMM-GUI, LeuT structures were capped with ACE and CT3 residues. For protein-bilayer complexes, the structures were embedded in a POPE bilayer of 150 lipid molecules equally distributed in the upper and lower leaflet using the Insertion method. For protein-micelle complexes, protein structures were embedded in 150 BDMM detergent molecules. We only added three Cl^- ions to neutralize the system. Since we are only interested in the equilibrium conformational changes of *apo* protein, we did not want to introduce Na^+ ions that are known to play an important role in the transport mechanism of LeuT. Ion binding and substrate transport are coupled and ions can be considered as a substrate.

Setting up $PepT_{So}$ simulations in bilayer and detergent micelle

We used 42 structures extracted from our previous simulations of $PepT_{So}$ (20) from residue Pro8 to Tyr512 (505 residues in all). During setup in CHARMM-GUI, the protein was capped with ACP and CT3 residues. For protein-bilayer complexes, the structures were embedded in a heterogeneous POPE/POPG (3:1 ratio) bilayer of 200 lipid molecules equally distributed in the upper and lower leaflet. For protein-micelle complexes, protein structures were embedded in 150 BDMM detergent molecules. We added 0.15 M NaCl ions in addition to neutralizing the system.

Setting up $PepT_{So}$ Simulations with MTSSL probes on a single residue pair

The same method as for $PepT_{So}$ in detergent micelle was followed, and residues Asn174 and Ser466 were mutated to MTSSL (1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl methanethiosulfonate) (26) probes. This corresponds to one of the extracellular residue pairs chosen by Fowler *et al.* for DEER experiments (2).

Details for system size and simulation time are provided in Table S1. For all simulations described above we examine convergence by randomly sampling 25%, 50%, and 75% of the trajectories and graphing experimental residue-pair distance distributions shown in Fig. S3-S7. We choose to look at these residue-pair distance distributions, as a check for convergence, because these will be the main focus of most results in this work. We see that for all systems, error bars are small even with 25% simulation data, and they decrease as we include a larger portion of the data. We conclude that multiple trajectories sample each region of the conformational ensemble and no single trajectory shifts the distance distributions completely.

Restrained-ensemble molecular dynamics (reMD) simulations for $PepT_{So}$

We used 42 different protein conformations as a starting point for reMD simulations in vacuum which means the protein was not surrounded by lipids, water, or ions. CHARMM-GUI's default 25 spin-label copies were attached to each labeled protein residue. Experimental distance distributions from Fowler *et al.* were provided as target histograms (2). We also used default values for force constants, bin widths, and Gaussian natural spread (9). Simulations were run using a special version of NAMD 2 (10, 27). For system reMD (1 dist), we attached MTSSL probes on residues Asn174 and Ser466, and restrained this distance. For system reMD (2 dist), MTSSL probes were placed on four residues and two distances were restrained, Asn174-Ser466 and Arg201-Glu364. These residue pairs are on the opposite side of the protein. For system reMD (8 dist), MTSSL probes were places on 12 residues, and eight experimentally studied residue pairs were restrained. Details for system size and simulation time are provided in Table S3. We used an integration timestep of 1 fs and saved trajectory coordinates at a frequency of 50 ps. Since reMD simulations are biased simulations, where the distance between the probe molecules is restrained to a targeted distribution using harmonic forces we use these simulations as an opportunity to explore the protein and the MTSSL probe dynamics when the experiment and experiment distance distributions show a perfect match. This is also why reMD simulations in vacuum are efficient and sufficient to sample the spin probe dynamics.

Experimental DEER distances

Experimental DEER distances and distance distributions were extracted from previous experiments published in ref. (2) and (3) using Plot Digitizer Java program.
For $PepT_{So}$, eight DEER distributions are available:

- Five intracellular distances (86-432,141-432,141-438,141-500,201-364)
- Three extracellular distances (47-330,174-401,174-466)

For LeuT protein, we have examined 24 distance distributions because these distributions have data available for Apo system in ref. (3).

- 17 intracellular distances (185-271,79-277,184-277,7-86,12-86,12-377,71-193,193-377,12-371,71-89,71-184,71-377,79-377,71-425,71-455,277-425,277-455)
- 7 extracellular distances (309-480,123-240,208-240,37-123,37-208,123-306,208-306)

$PepT_{So}$ experimental distributions were fitted to multiple Gaussian distributions to get an equal-sized-bin distribution for KL divergence calculations and restrained MD simulations. Comparisons shown in Fig. S23.

List of LeuT structural models

2A65 (28), 2Q6H (29), 2Q72 (29), 2QB4 (29), 2QEI (29), 3F3A (30), 3F3C (30), 3F3D (30), 3F3E (30), 3F48 (30), 3F4I (30), 3F4J (30), 3GJD (31), 3GWU (32), 3GWV (32), 3GWW (32), 3TT1 (chains A & B) (21), 3TT3 (chain A) (21), 3USG (33), 3USI (chains A & B) (33), 3USJ (chains A & B) (33), 3USK (chains A, B, C, & D) (33), 3USL (33), 3USM (33), 3USO (chains A & B) (33), 3USP (33), 5JAE (chains A & B) (34), 5JAF (34).

MtsslWizard Implementation

MtsslWizard is a software package to find out the all the possible rotamers of mtssl probes that can be added to a residue of a specific protein structure (4). Here, our goal is to find the distance distributions of ON atoms of two mtssl probes attached in two different residues considering all possible rotamers of mtssl probes for entire protein conformational ensemble. Different conformations of proteins may lead to distinct rotamer distribution. To represent structures from entire ensemble obtained from the simulation, we clustered the conformational ensemble into 100 clusters. 500 samples were obtained from these 100 clusters based on probability of the populations of each cluster. Rotamer calculations were performed on each 500 structures. Distance distributions from entire ensemble are compared against the experimental distribution.

Rotamer library calculation

For rotamer library calculations shown in Fig. S24 and S25, both MD simulation datasets of $PepT_{So}$ and LeuT proteins in bilayer and micelle setting were clustered into 500 states. Clustering was performed using k-means algorithm with eight experimentally characterized distances ($PepT_{So}$) and two experimentally characterized distances 185-271 and 309-480 (for LeuT). To represent the MD ensemble, 10 000 frames were selected from each dataset proportional to the frame counts of states. *RotamerConvolveMD* package (35, 36) within MDAnalysis (37, 38) was used for rotamer library calculation for experimentally characterized residue pairs.

Unbiased simulation of frames from reMD simulations

To visualize the effect of restraint on the protein structure, reMD simulation data with one, two, and eight restraints were clustered into five macrostates. One structure was selected from each cluster center. Each selected protein structure was then placed in a bilayer setting with MTSSL probes added at the same residue positions as in reMD simulation. 50 ns of production run was generated after minimizing and equilibrating each structure. Time traces from 50 ns production runs are shown in Fig. S26.

Data Analysis

Micelle radius. First, we compute the radius of gyration (R_g) of the micelle using *compute_rg* in MDTraj 1.7 (1), which is related to the micelle radius (R) as, $R = \sqrt{\frac{5}{3}} R_g$ (39). This formula hold when the micelle is assumed to be spherical in shape.

Eccentricity. The shape of the micelle and the protein-micelle complex is determined using the ratio between moments of inertia I_1, I_2, and I_3 S2. Eccentricity is calculated as $1 - I_{min}/I_{avg}$ (39, 40). The moments are inertia are defined as the eigenvalues of a moment of inertia tensor calculated using *compute_inertia_tensor* in MDTraj 1.7 (1).

Distance distributions. All inter-residue distance distributions are estimated as the distance between the closest heavy atoms between the two residues unless otherwise mentioned.

Inter-helix distances Transmembrane helix ends for proteins are defined based on OPM database (41) numbering for 14 helices in $PepT_{So}$ and 12 in LeuT. We determine inter-helix distances among all helices on intracellular and extracellular side of the proteins. For $PepT_{So}$ these are $2 \times (14)(13)/2 = 182$ distances and for LeuT these are $2 \times (12)(11)/2 = 132$ distances.

Kullback-Leibler (KL) divergence. KL divergence (also called relative entropy) is a measure of how one probability distribution is different from a second, reference probability distribution. KL divergence for two distributions P and Q is 0 if and only if P and Q are exactly equal. For two discrete probability distributions P and Q, defined on the same probability space, X, the KL divergence of Q from P is defined to be,

$$KL(P|Q) = -\sum_{x \epsilon X} P(x) log(\frac{Q(x)}{P(x)}) \tag{1}$$

KL divergence is an asymmetric measure by definition, and wherever possible we have used this measure both ways to validate our conclusions regarding similarity and difference among probability distribution. We used $scipy.stats.entropy$ routine to calculate KL divergence values. Another useful measure of divergence between probability distribution we use is Symmetrised Divergence, which is symmetric and non-negative defined as,

$$\text{Divergence} = KL(P|Q) + KL(Q|P) \tag{2}$$

When calculating frequencies used for the KL divergence we corrected for the presence of frequencies of zero by adding a very small value to the probability distribution.

Helical content. The helical content of all TM helices is calculated as defined in the NAMD 2.11 manual (10). The python implementation is taken from https://github.com/amoffett/alpha_helical_content as used in ref. (42). The individual helices in this work are determined based on the OPM database web server (41) for $PepT_{So}$ PBD 4UVM (2) and LeuT PDB 2A65 (28). Specifically, TM1 of LeuT refers to residues in TM1a only, which are residues 15 to 25 whereas the helix ends at residue 35.

Table 1: List of MD simulations.

Complex	Components[#1]	# Trajectories	# Atoms	Equilibration run (ps)	Simulation time (for analysis[#2], μs)
LeuT-bilayer	150 POPE, Cl^-	72	56,707 - 66,784	675	32.18
LeuT-micelle	150 BDDM, Cl^-	72	107,521 - 145,589	450	28.73
$PepT_{So}$-bilayer	150 POPE, 50 POPG, NaCl	42	66,045 - 74,700	675	27.3
$PepT_{So}$-micelle	150 BDDM, NaCl	42	130,506 - 195,681	750	20.42
$PepT_{So}$-micelle-MTSSL probes	150 BDDM, NaCl	42	128404 - 181800	750	18.78

#1: All systems contain protein and TIP3P water.
#2: For all trajectories, we eliminate the first 10 ns of the production run from analysis.

Table 2: Geometry of protein-micelle complexes with varied micelle sizes.

	Complex PepTSo w/ detergents	Micelle Radius (nm)	I1 : I2 : I3	Eccentricity
Micelle	150 detergents	4.4 ± 0.02	1.46 : 1 : 1.24	0.23 ± 0.02
	180 detergents	4.58 ± 0.03	1.42 : 1 : 1.25	0.22 ± 0.02
	200 detergents	4.68 ± 0.02	1.53 : 1 : 1.16	0.22 ± 0.02
Protein+Micelle	150 detergents	-	1.26 : 1 : 1.16	0.16 ± 0.02
	180 detergents	-	1.23 : 1 : 1.18	0.16 ± 0.01
	200 detergents	-	1.38 : 1 : 1.13	0.17 ± 0.02

Table 3: List of reMD simulations.

System	# Trajectories	# Atoms	Equilibration run (ps)	Production run (ns)	Simulation time (μs)
reMD (1 dist)	42	9795	25	~95	3.98
reMD (2 dist)	42	11645	25 (2 setups required 50 ps)	~95	3.88
reMD (8 dist)	42	19049	25 (8 setups required longer)	~65	2.67

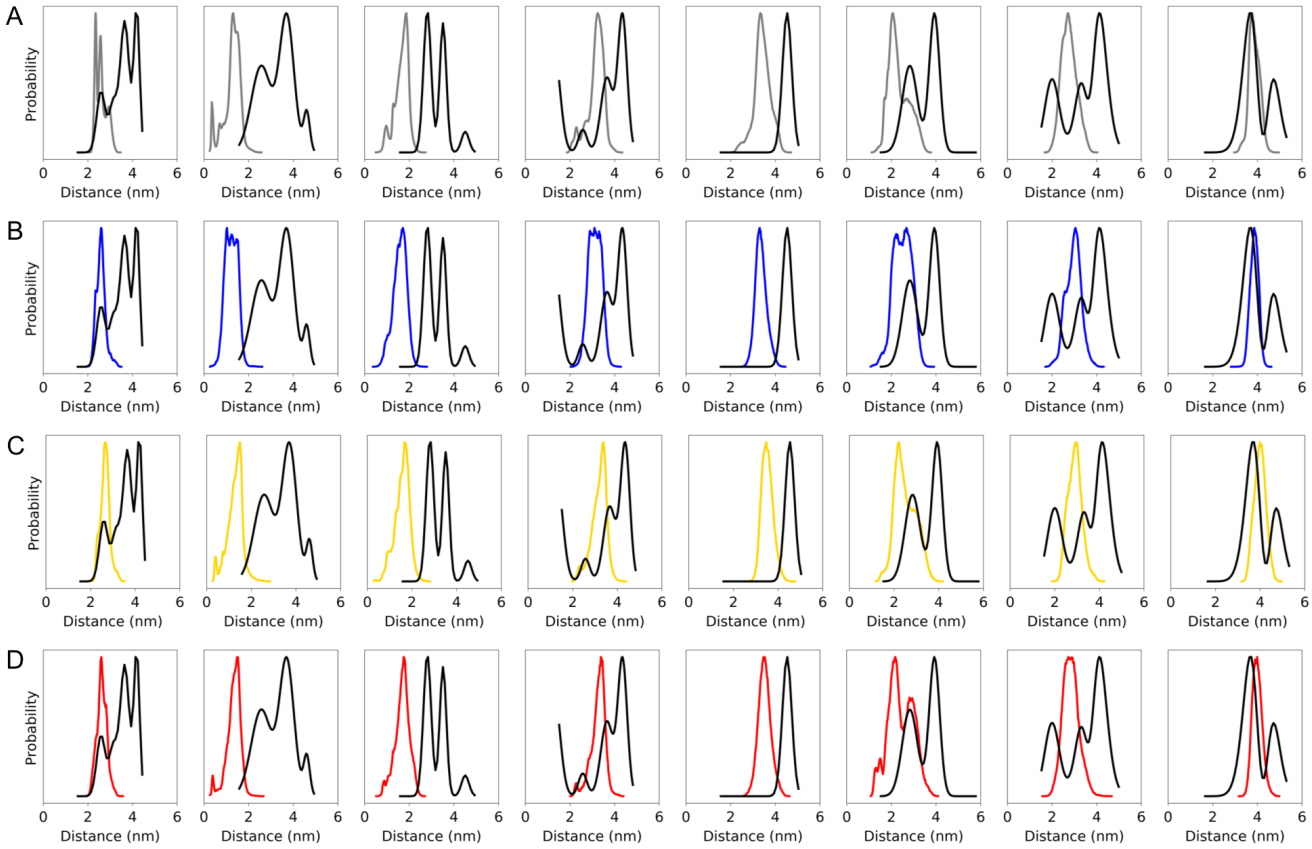

Figure 1: Experimentally characterized residue-pair distance distributions as observed in our MD simulations in (A) POPC bilayer (simulations previously performed (20)), (B) POPE/POPG (3:1 ratio) bilayer, (C) BDDM micelle, and (D) BDDM micelle with MTSSL labeled residue pair. Black lines show experimental DEER distance distributions obtained from Fowler *et al.* as discussed above (2). The eight DEER distance distributions shown correspond to distance between residue-pairs 86-432, 141-432, 141-438, 141-500, 201-364, 47-330, 174-401, and 174-466) respectively.

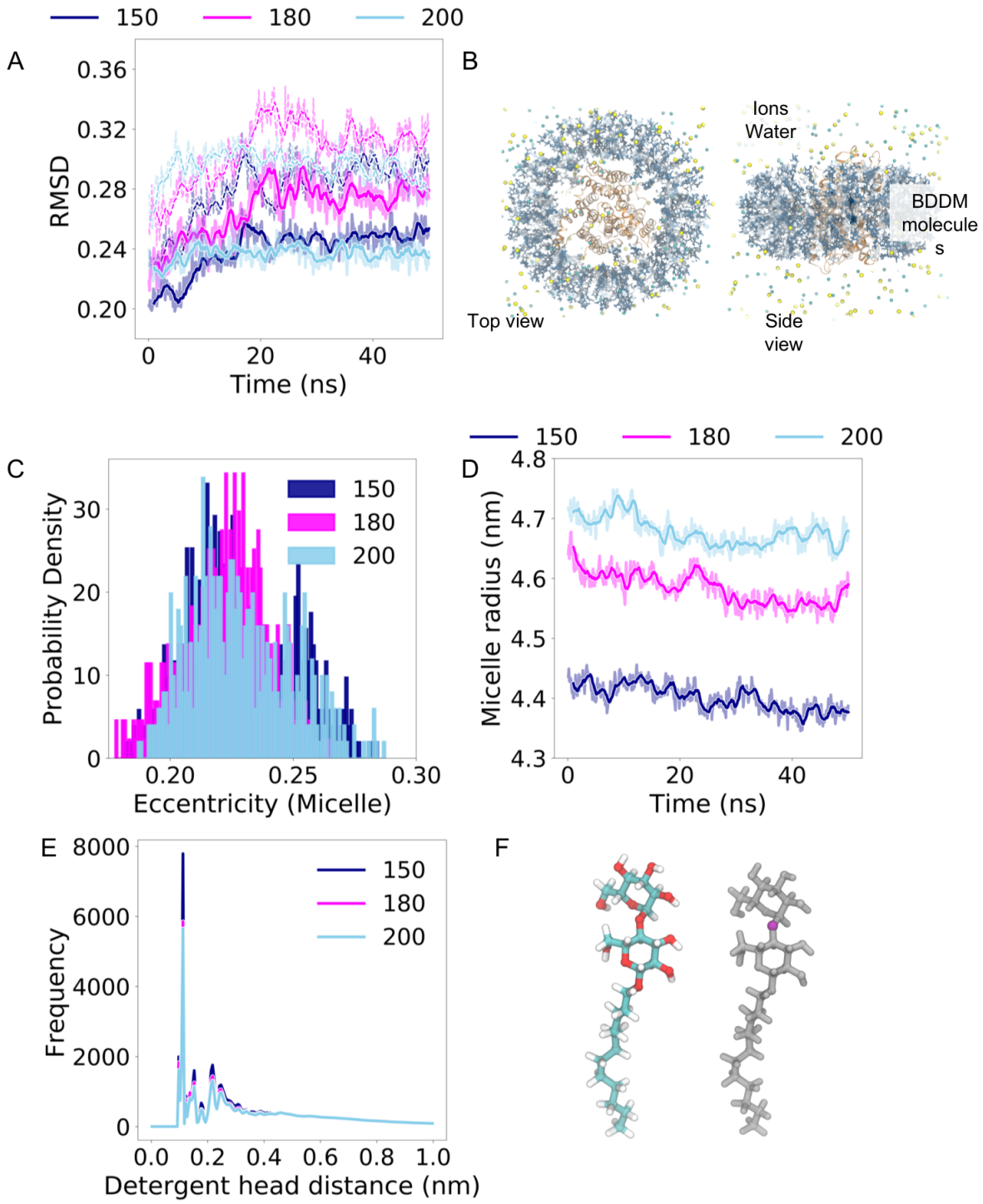

Figure 2: (A) RMSD of protein with respect to the starting frame is shown with time. The dotted lines show RMSD of the full protein while the bold lines show RMSD of the transmembrane region of the protein. Shaded regions show instantaneous values while the lines show a running time average RMSD over a 1 ns time window. (B) An example protein-micelle setup top and side view including BDDM detergent molecules and ions. (C) Probability distribution of micelle eccentricity values. (D) Micelle radius with time is shown. Shaded regions show instantaneous values while the lines show a running time average radius over a 1 ns time window. (E) Radial distribution of distances between BDDM detergent molecule headgroups. Headgroup distances are estimated using the distance among oxygen atoms highlighted in magenta in (F). Colors indicate three micelle sizes, micelle with 150 (blue), 180 (magenta), 200 (sky blue) detergent molecules.

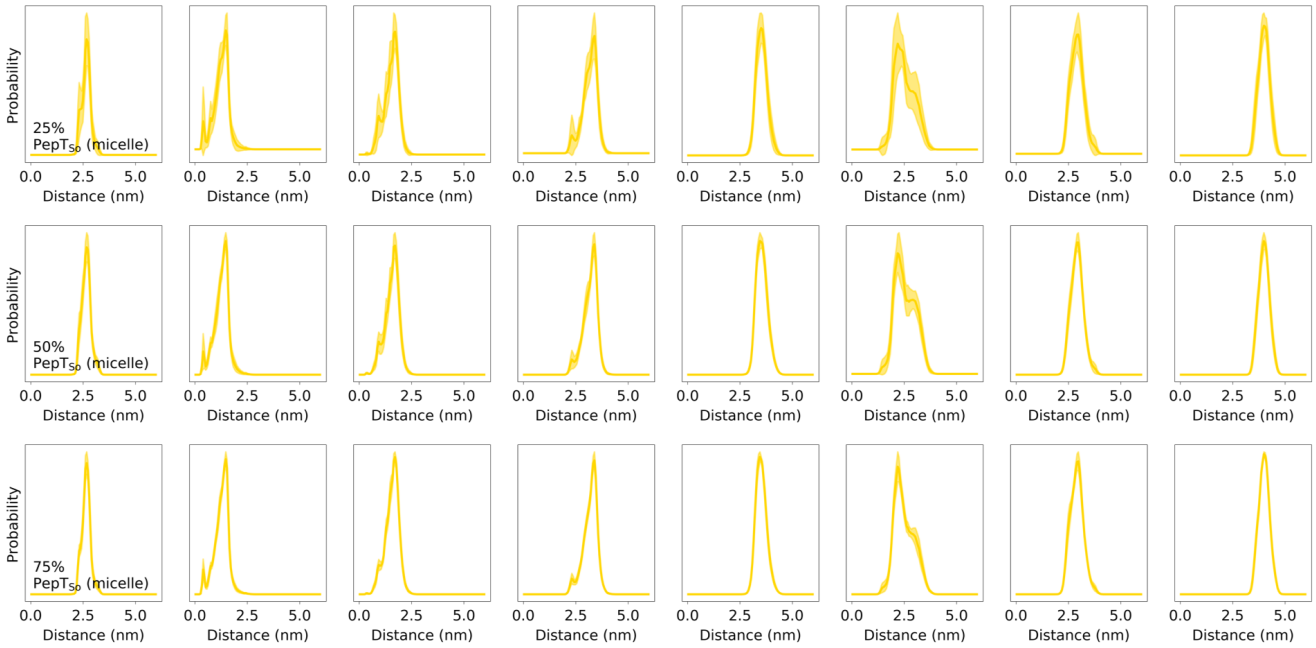

Figure 3: Residue-pair distance distributions for $PepT_{So}$ simulations in BDDM micelle averaged over 25%, 50%, and 75% of the collected trajectories. Filled regions show error bars in the distance distribution as obtained from 10 iterations where a subset of the trajectories is selected randomly. The eight DEER distance distributions shown correspond to distance between residue-pairs 86-432, 141-432, 141-438, 141-500, 201-364, 47-330, 174-401, and 174-466 respectively.

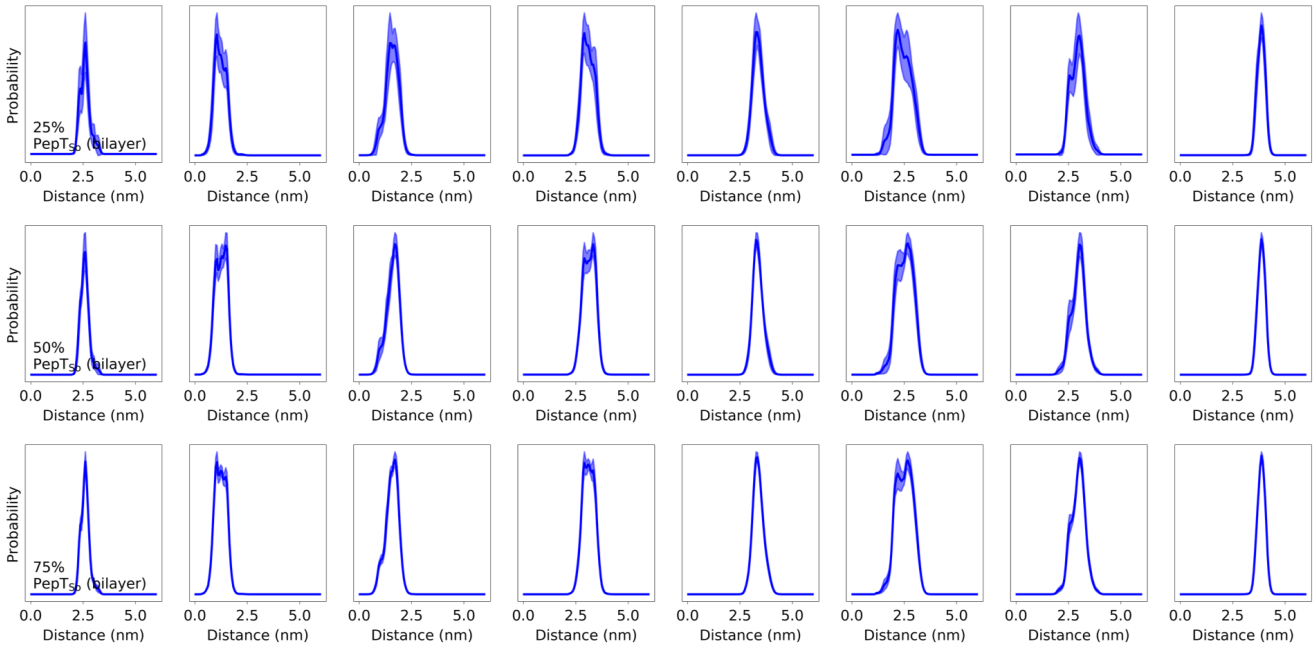

Figure 4: Residue-pair distance distributions for $PepT_{So}$ simulations in bilayer averaged over 25%, 50%, and 75% of the collected trajectories. Filled regions show error bars in the distance distribution as obtained from 10 iterations where a subset of the trajectories is selected randomly. The eight DEER distance distributions shown correspond to distance between residue-pairs 86-432, 141-432, 141-438, 141-500, 201-364, 47-330, 174-401, and 174-466 respectively.

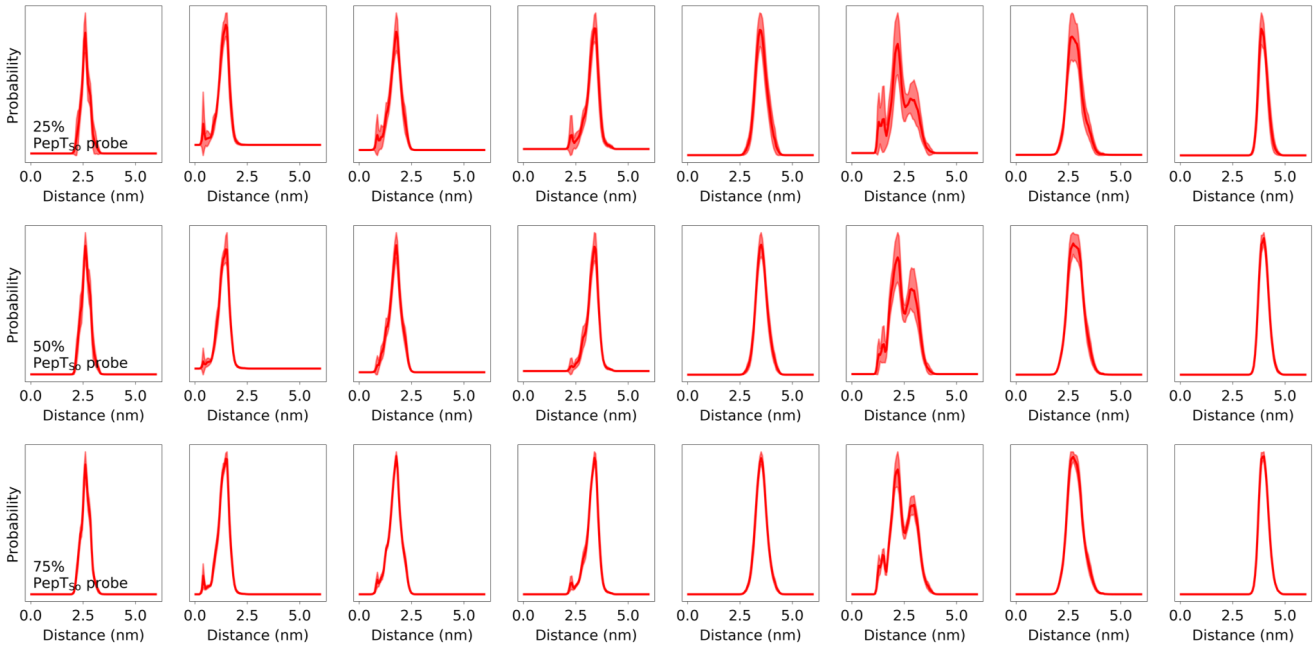

Figure 5: Residue-pair distance distributions for PepT$_{So}$ simulations in BDDM micelle with a residue pair labeled residue pair averaged over 25%, 50%, and 75% of the collected trajectories. Filled regions show error bars in the distance distribution as obtained from 10 iterations where a subset of the trajectories is selected randomly. The eight DEER distance distributions shown correspond to distance between residue-pairs 86-432, 141-432, 141-438, 141-500, 201-364, 47-330, 174-401, and 174-466 respectively.

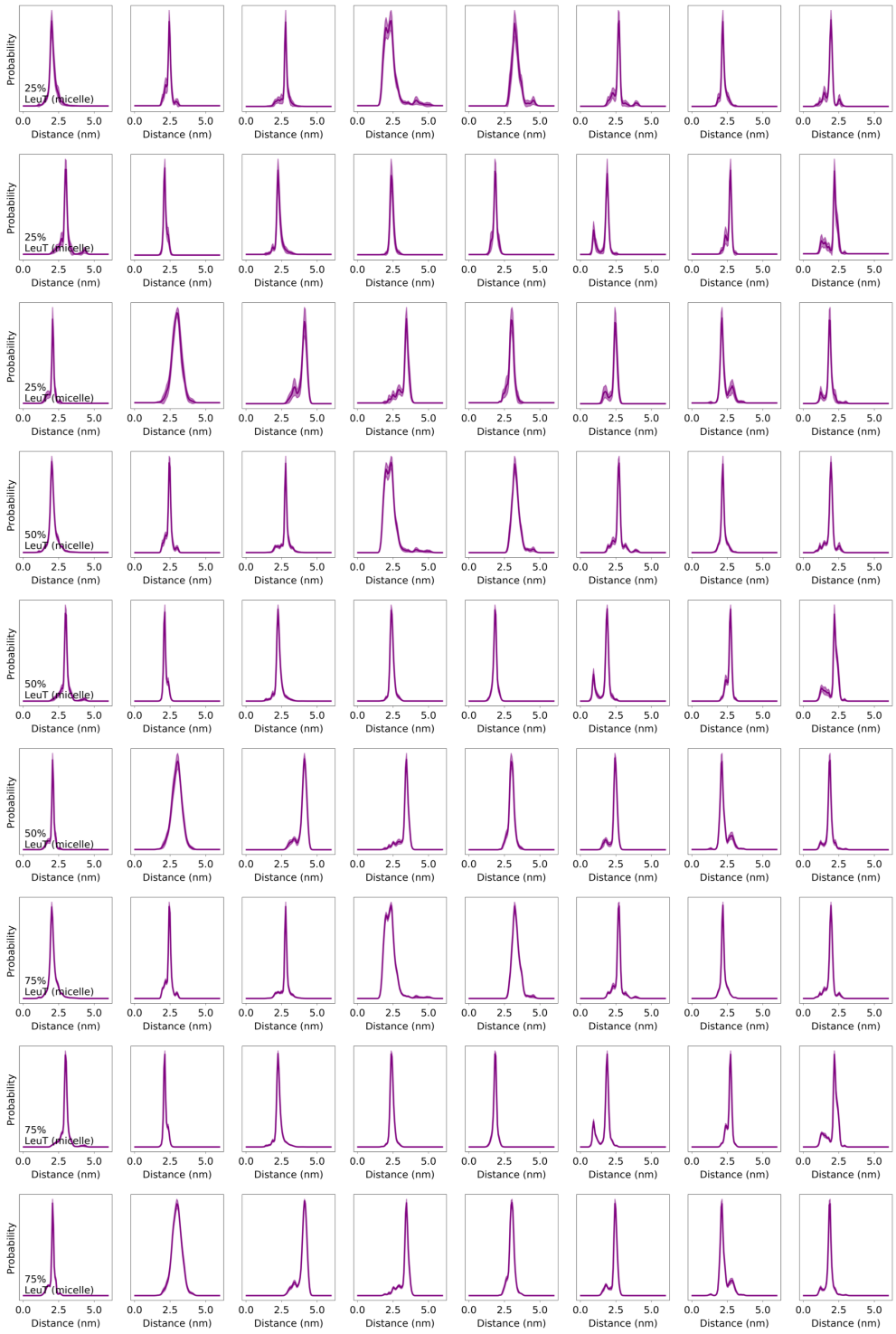

Figure 6: Residue-pair distance distributions for LeuT simulations in BDDM micelle averaged over 25%, 50%, and 75% of the collected trajectories. Filled regions show error bars in the distance distribution as obtained from 10 iterations where a subset of the trajectories is selected randomly. In all 24 DEER distance distributions are shown, where rows 1, 4, and 7 show distances distributions for residue-pairs 185-271, 79-277, 184-277, 7-86, 12-86, 12-377, 71-193, and 193-377; rows 2, 5, and 8 show distance distributions for residue-pairs 12-371, 71-89, 71-184, 71-377, 79-377, 71-425, 71-455, and 277-425; and rows 3, 6, and 9 show distance distributions for residue-pairs 277-455, 309-480, 123-240, 208-240, 37-123, 37-208, 123-306, and 208-306.

Figure 7: Residue-pair distance distributions for LeuT simulations in bilayer averaged over 25%, 50%, and 75% of the collected trajectories. Filled regions show error bars in the distance distribution as obtained from 10 iterations where a subset of the trajectories is selected randomly. In all 24 DEER distance distributions are shown, where rows 1, 4, and 7 show distances distributions for residue-pairs 185-271, 79-277, 184-277, 7-86, 12-86, 12-377, 71-193, and 193-377; rows 2, 5, and 8 show distance distributions for residue-pairs 12-371, 71-89, 71-184, 71-377, 79-377, 71-425, 71-455, and 277-425; and rows 3, 6, and 9 show distance distributions for residue-pairs 277-455, 309-480, 123-240, 208-240, 37-123, 37-208, 123-306, and 208-306.

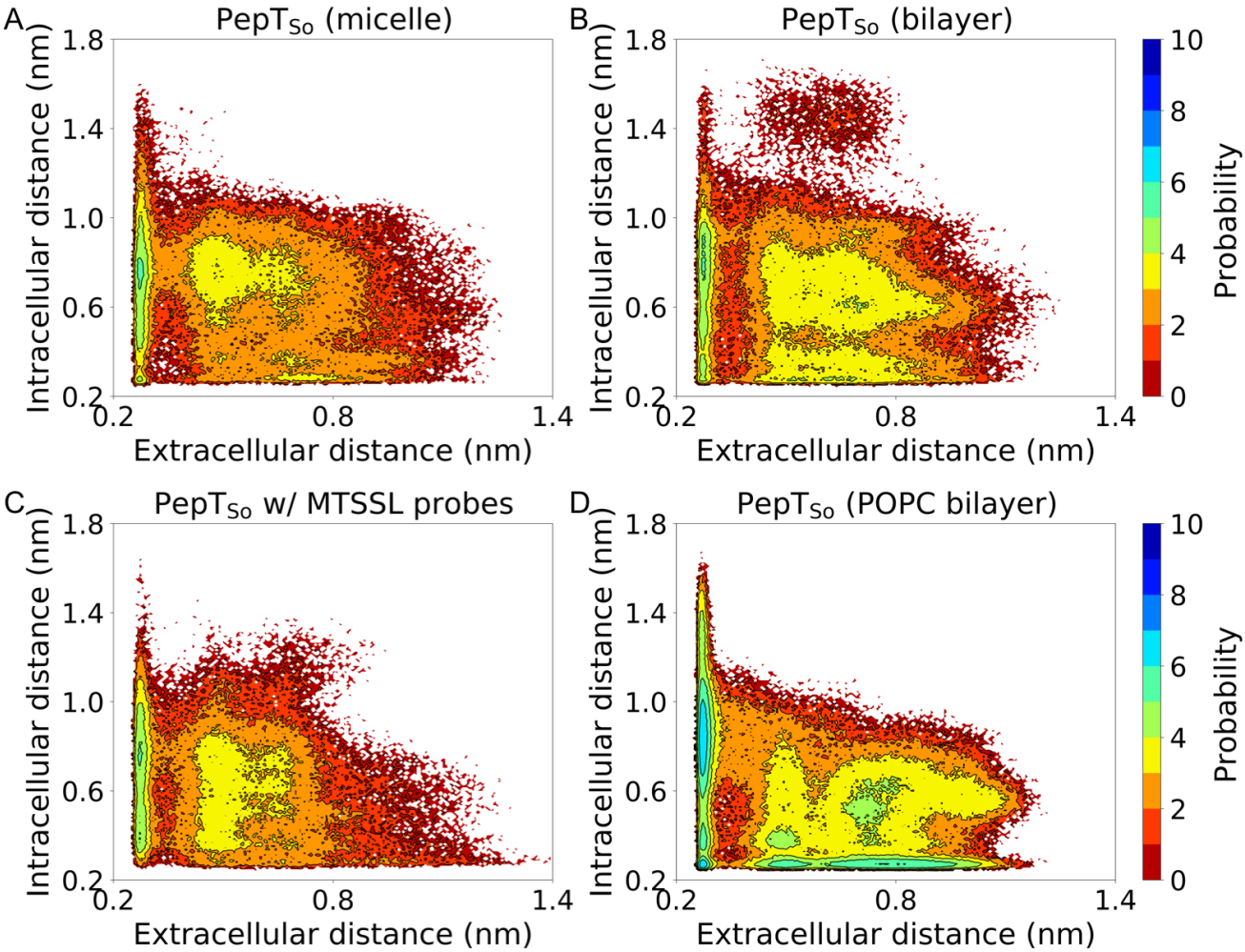

Figure 8: The conformational landscapes of $PepT_{So}$ protein are generated by projecting all simulation data on the chosen extracellular and intracellular side distances measured between Arg32-Asp316 and Ser131-Tyr431, respectively. (A) Conformational landscape for $PepT_{So}$ MD simulations in BDDM micelle. (B) Conformational landscape for $PepT_{So}$ MD simulations in POPE/POPG (3:1 ratio) bilayer. (C) Conformational landscape for $PepT_{So}$ MD simulations in BDDM micelle with an MTSSL labeled residue pair. (D) Conformational landscape from our previous simulations in a POPC bilayer and using an AMBER FF14SB force field (20).

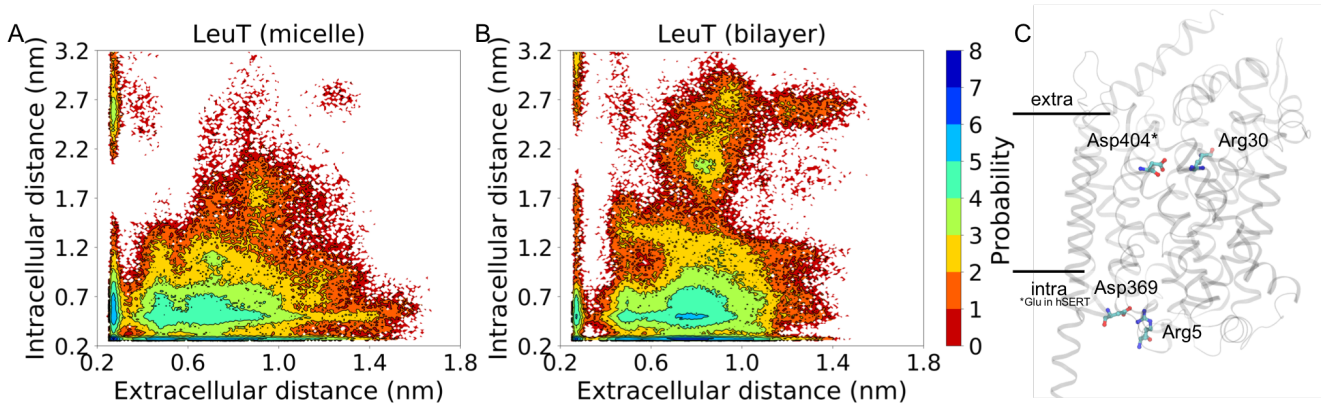

Figure 9: The conformational landscapes of LeuT protein are generated by projecting all simulation data on the chosen extracellular and intracellular side distances measured between Arg30-Asp404 and Arg5-Asp369, respectively. (A) Conformational landscape for LeuT MD simulations in BDDM micelle. (B) Conformational landscape for LeuT MD simulations in a bilayer. (C) Gating residues used to determine extracellular and intracellular distances are shown on a cartoon representation of a three-dimensional LeuT structure.

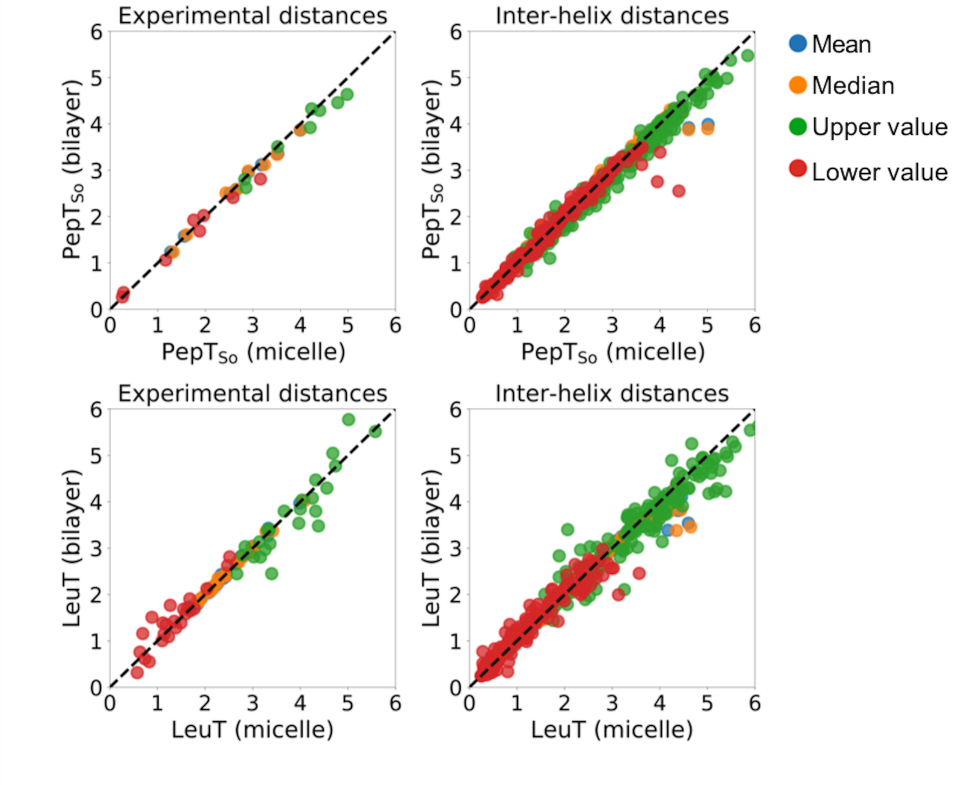

Figure 10: Comparing mean (blue), median (orange), upper value (green), and lower value (red) of distance distributions of experimental residue-pair distances and all inter-helix residue-pair distances. Markers below the black dotted line indicate larger values observed in micelle environment. Markers above the black dotted line indicate larger values observed in bilayer environment. Markers along the black dotted line indicate similar observations in micelle and bilayer simulations.

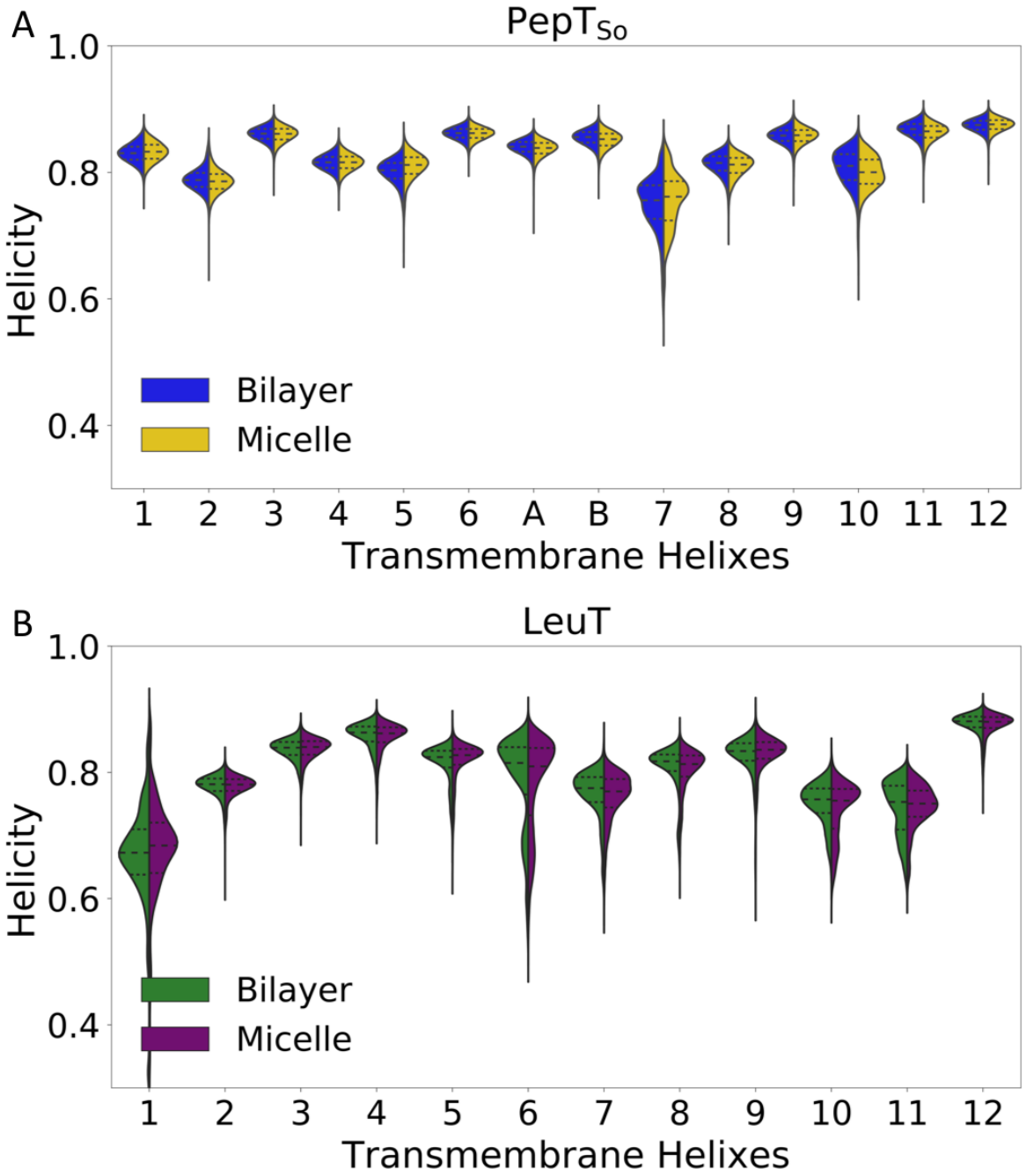

Figure 11: (A) Violin plot shows alpha-helical content for 14 TM helices as observed from MD simulations of $PepT_{So}$ protein in micelle (yellow, right) and bilayer (blue, left). (B) Violin plot shows alpha-helical content for 12 TM helices as observed from MD simulations of LeuT protein in micelle (purple, right) and bilayer (green, left).

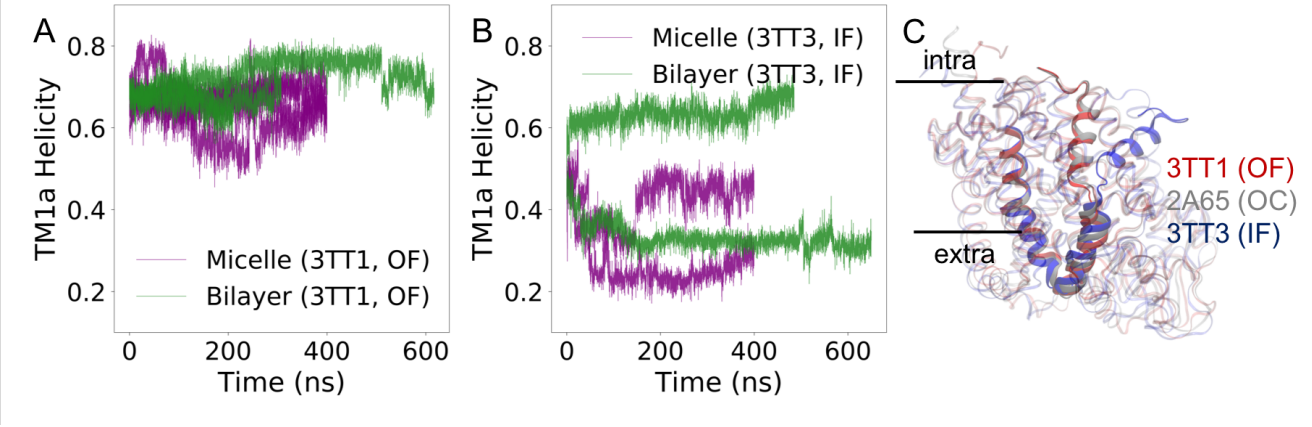

Figure 12: (A) TM1a alpha-helical content of trajectories started from OF structure of LeuT in micelle (purple) and bilayer (green). (B) TM1a alpha-helical content of trajectories started from the IF structure of LeuT in micelle (purple) and bilayer (green). (C) Superposed structures of LeuT's OF, OC, and IF structures.

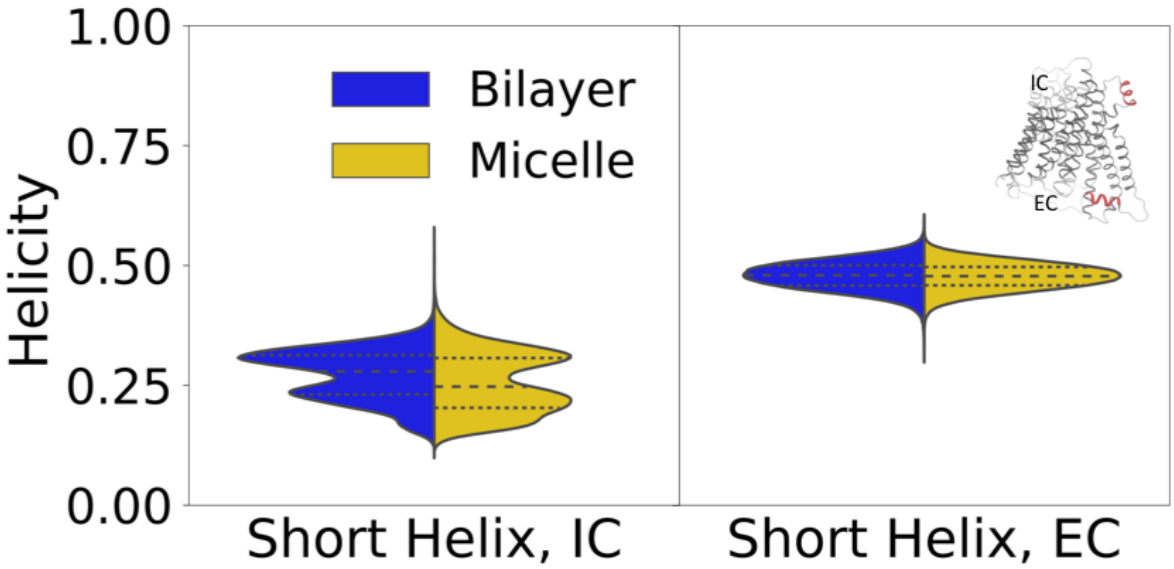

Figure 13: Violin plot shows alpha-helical content of a short helix on the intracellular (IC) side and another of the extracellular (EC) side of $PepT_{So}$ protein in micelle (yellow, right) and bilayer (blue, left). Inset shows two short helices in red on the $PepT_{So}$ protein structure in grey.

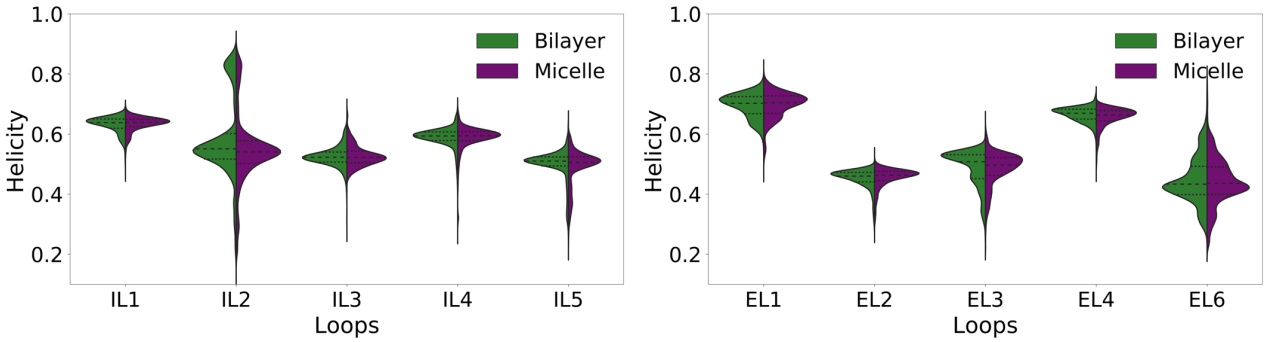

Figure 14: Violin plot shows alpha-helical content of intracellular loops (ILs) and extracellular loops (ELs) in LeuT protein in micelle (purple, right) and bilayer (green, left). Loop EL5 is only four residues long and too short to determine its alpha-helical content.

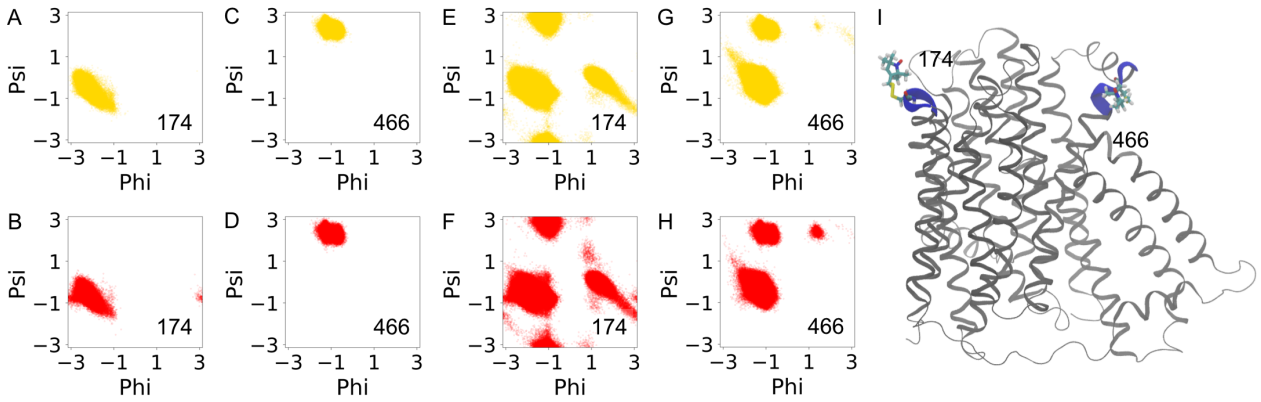

Figure 15: (A-D) Ramachandran plots for residues 174 and 466. Yellow and red colors indicate residue dihedral angle distribution in micelle and bilayer MD simulations, respectively. (E-F) Ramachandran plots for regions surrounding residues 174 and 466. (I) Residues 174 and 466 are shown on a cartoon representation of a three-dimensional $PepT_{So}$ structure.

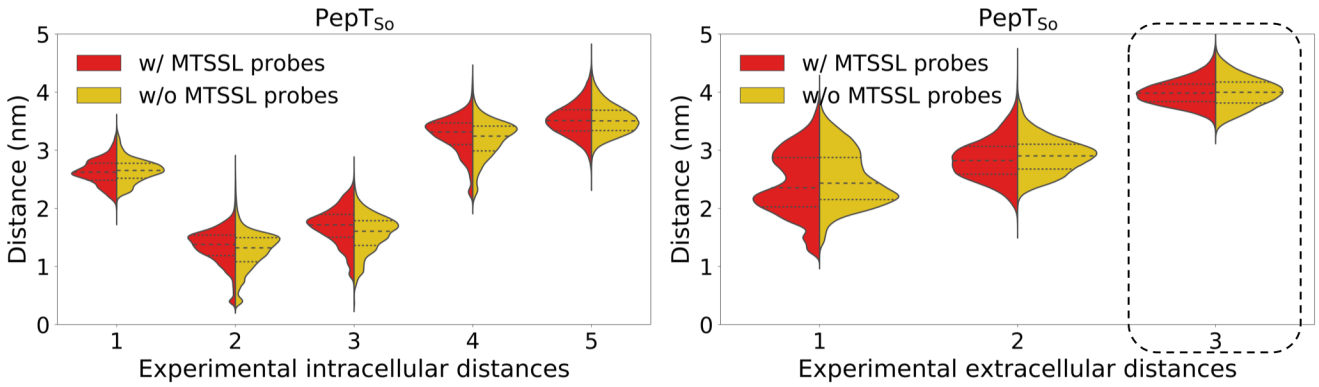

Figure 16: (A) Violin plot shows distance distributions for 5 intracellular residue-pair distances and 3 extracellular residue-pair distances measured by Fowler *et al.* as observed from MD simulations of PepT$_{So}$ protein in micelle without MTSSL probes (yellow, right) and with an MTSSL probe labeled residue pair (red, left). The black dotted outlined residue pair is the labeled residue pair.

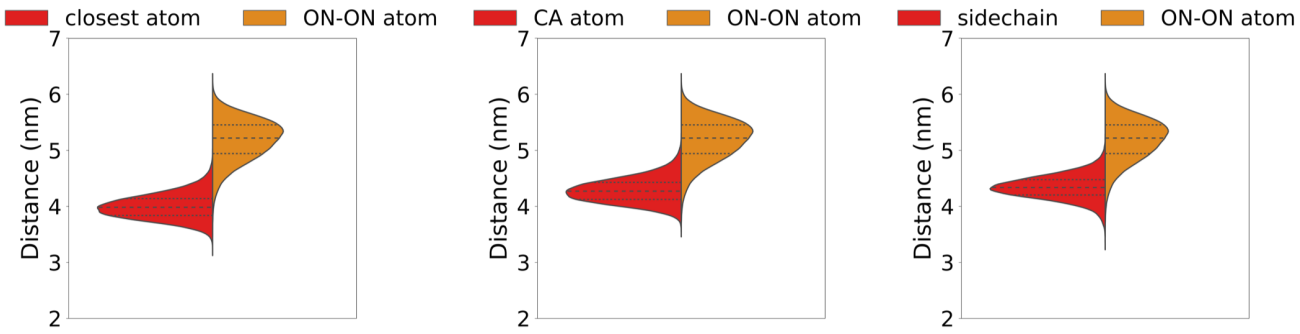

Figure 17: Violin plots compare distance distributions for simulations with MTSSL probes as measured between the ON atom with the closest heavy atom, C_α atom, and the closest sidechain atom of the labeled residues. ON-ON atom distance distributions are shown in orange and the backbone atom distance distributions are shown in red.

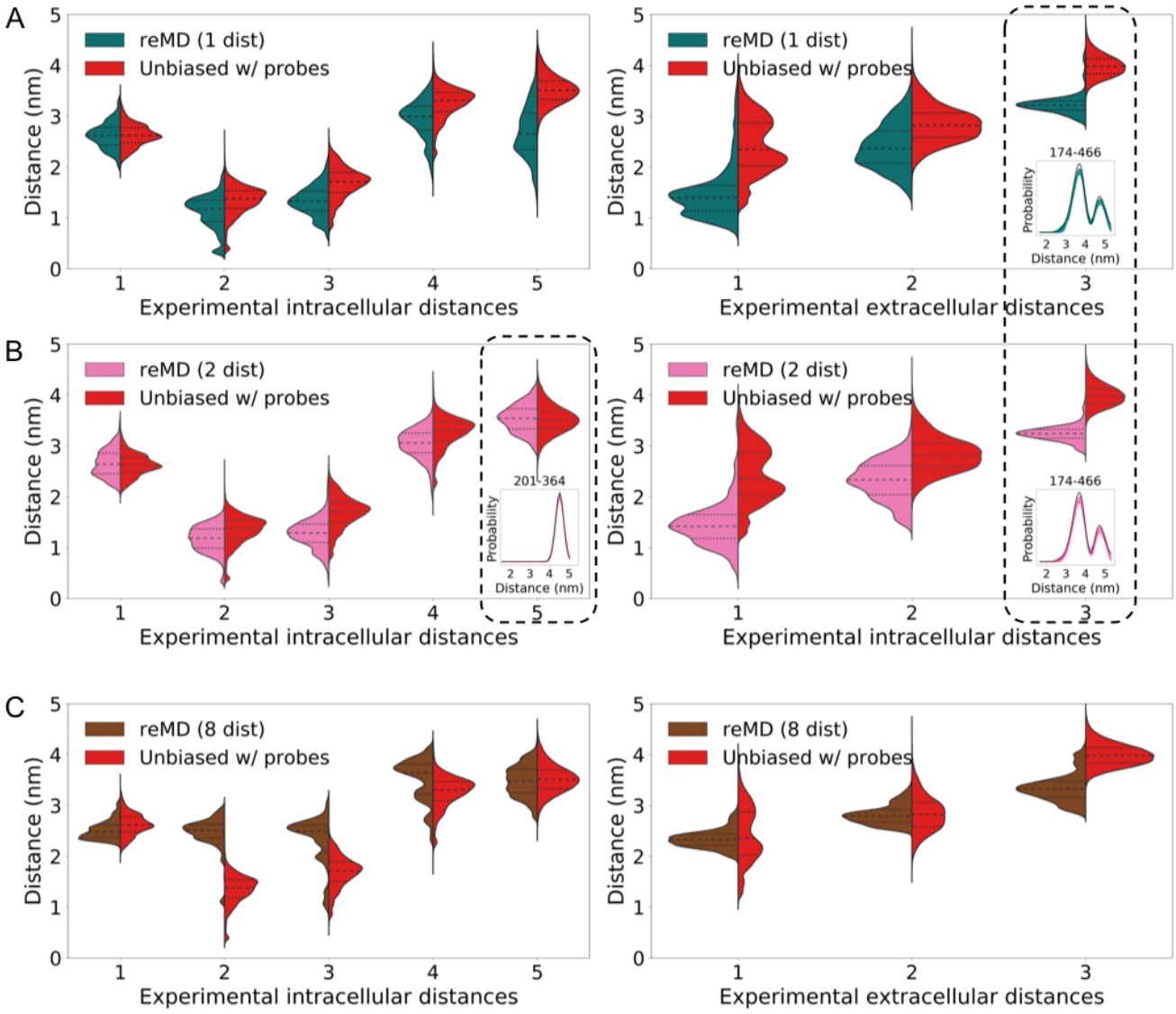

Figure 18: (A) Violin plot shows distance distributions for 5 intracellular residue-pair distances and 3 extracellular residue-pair distances as observed from (A) reMD (1 dist) simulations where residue pair 174-466 is restrained, teal violin plots, (B) reMD (2 dist) where residue pairs 174-466 and 201-364 are restrained, pink violin plots, and (C) reMD (8 dist) where all 8 residue pairs are restrained, brown violin plots. Yellow violin plots correspond to unbiased simulations of $PepT_{So}$ protein in micelle with MTSSL molecules on residues 174 and 466. Black dotted outlined residues pairs in (A) and (B) are restrained pairs and probe distances are shown to match with experimental DEER distance distributions.

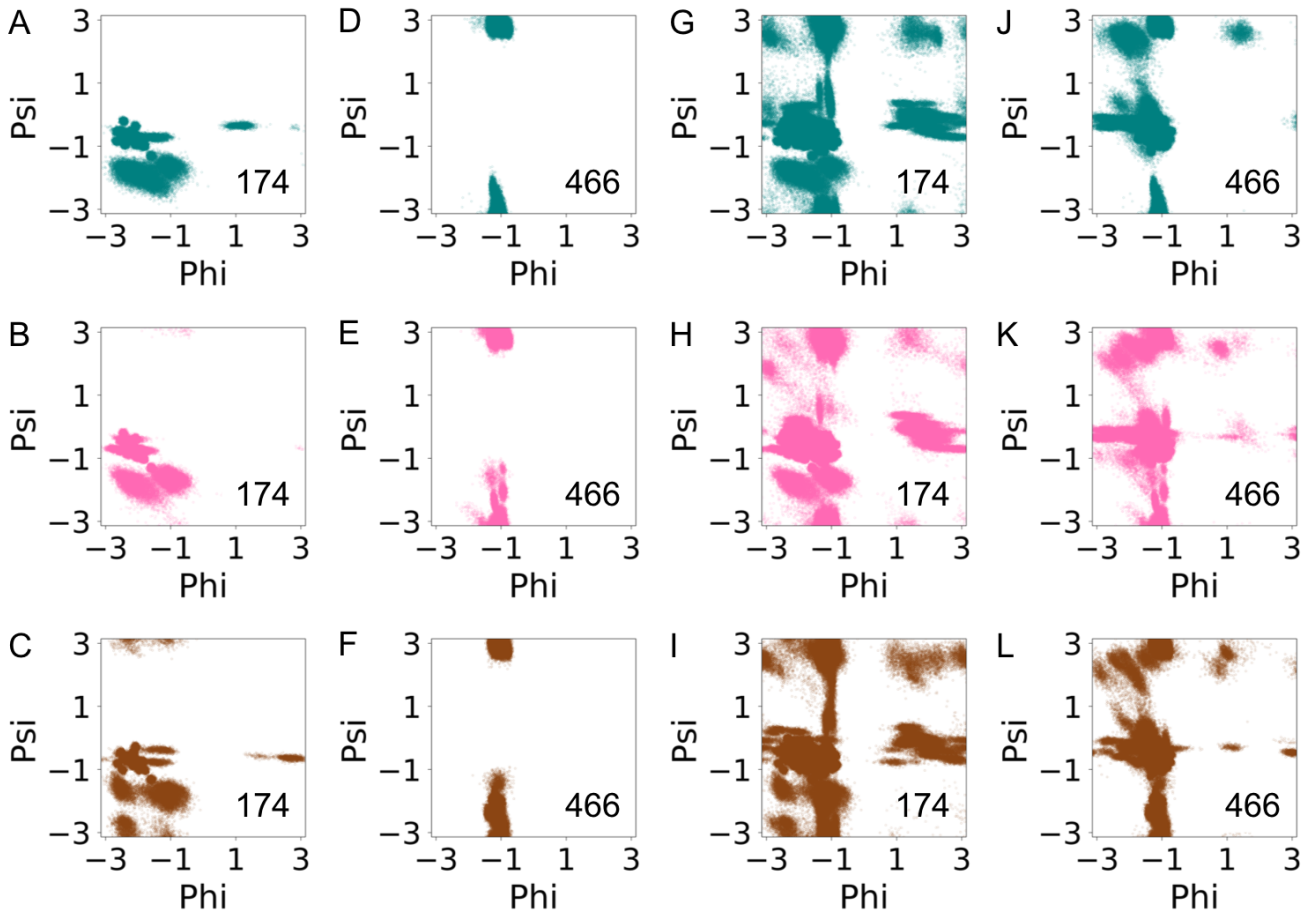

Figure 19: (A-F) Ramachandran plots for residues 174 and 466. Teal, pink, and brown colors indicate residue dihedral angle distribution in reMD (1 dist), reMD (2 dist), and reMD (8 dist) MD simulations, respectively. (G-L) Ramachandran plots for regions surrounding residues 174 and 466.

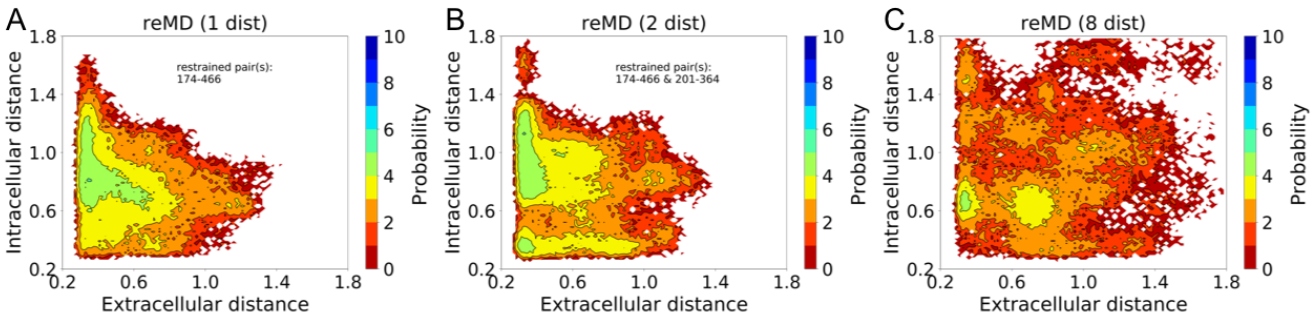

Figure 20: Conformational landscape for $PepT_{So}$ (A) reMD (1 dist), (B) reMD (2 dist), and (C) reMD (8 dist) simulations. The conformational landscapes are generated using the same residue pairs as in Figure S8.

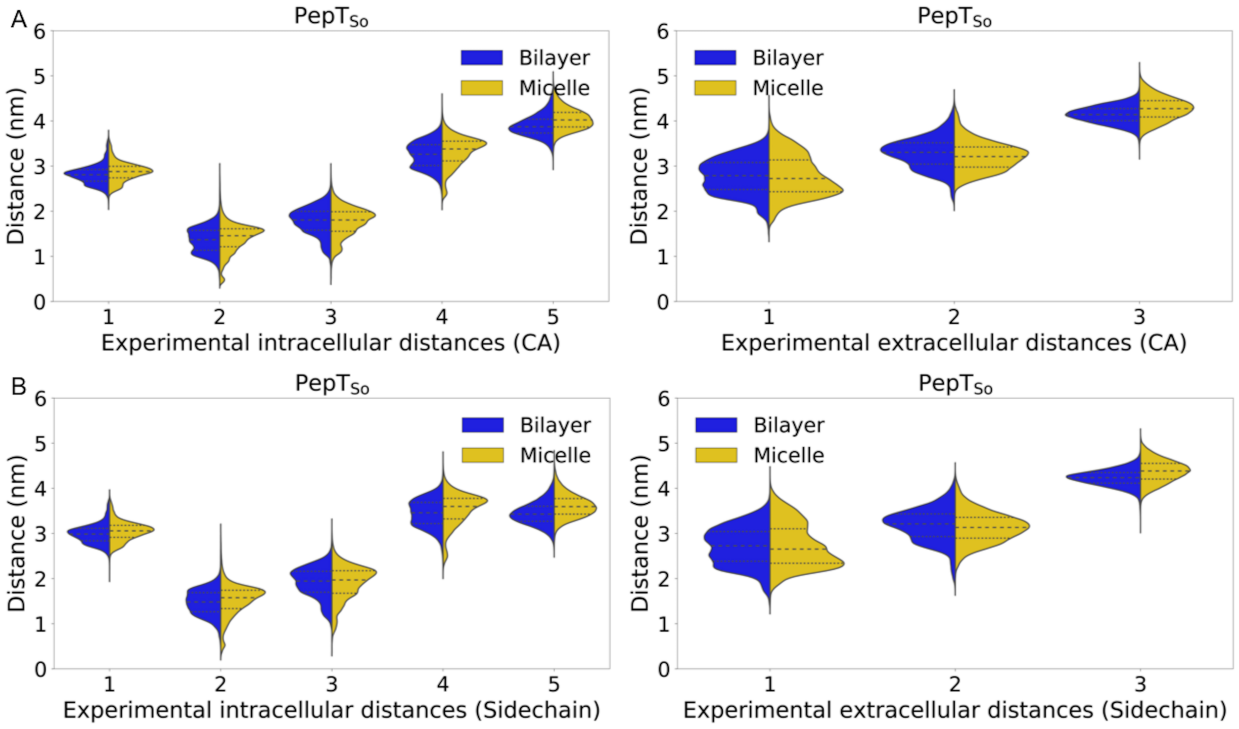

Figure 21: Violin plot shows distance distributions for 5 intracellular residue-pair distances and 3 extracellular residue-pair distances measured by Fowler *et al.* as observed from MD simulations of $PepT_{So}$ protein in micelle (yellow, right) and bilayer (blue, left) (2). (A) Residue pair backbone distances as measured between C_α atom of residues. (B) Residue pair sidechain distances i.e. closest distance between any two non-hydrogen atoms in residue sidechains.

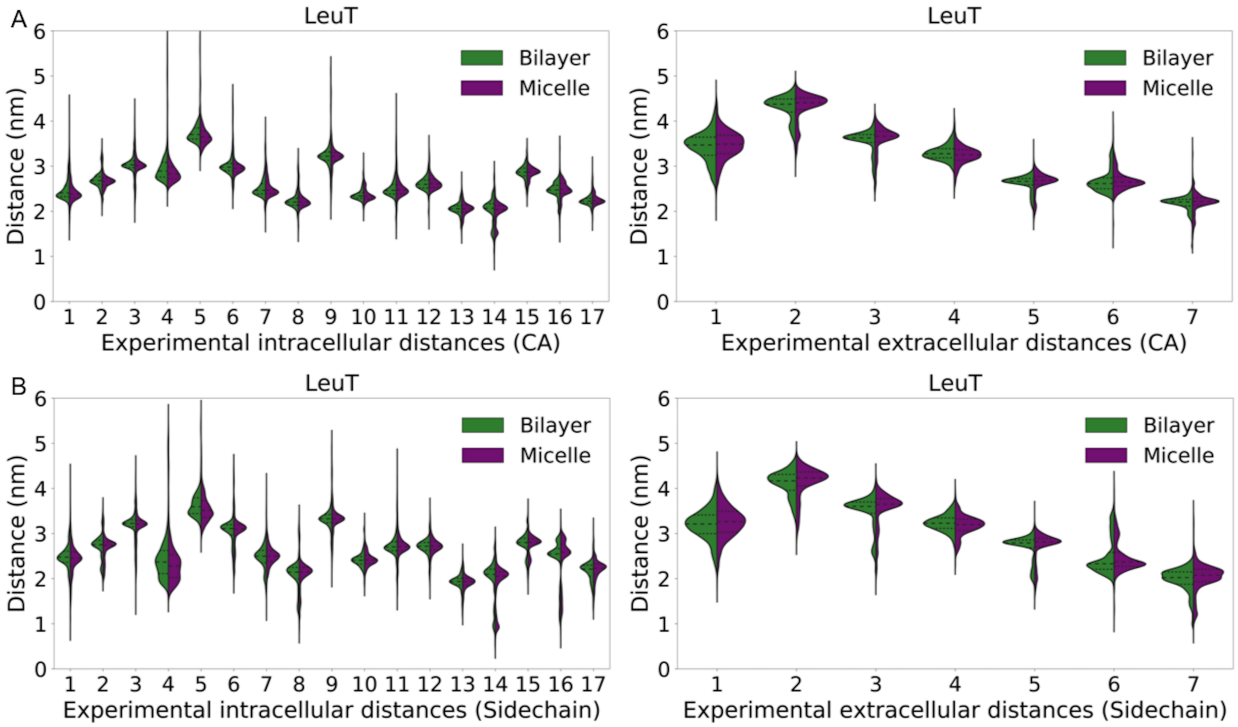

Figure 22: Violin plot shows distance distributions for 17 intracellular residue-pair distances and 7 extracellular residue-pair distances measured by Kazmier *et al.* as observed from MD simulations of LeuT protein in micelle (purple, right) and bilayer (green, left) (3). (A) Residue pair backbone distances as measured between C_α atom of residues. (B) Residue pair sidechain distances i.e. closest distance between any two non-hydrogen atoms in residue sidechains.

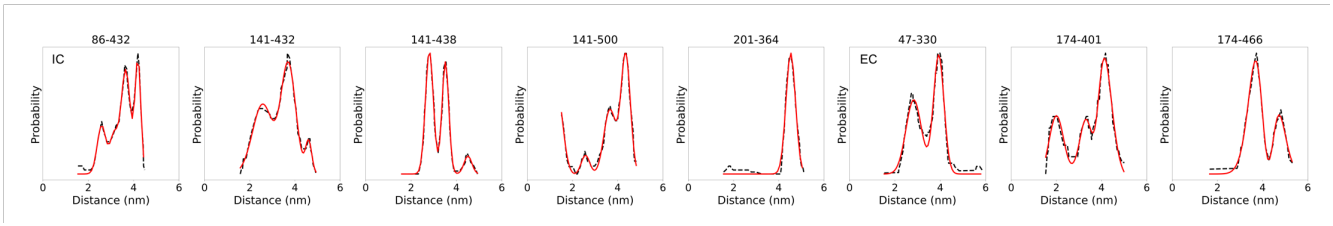

Figure 23: Black dotted lines indicate experimental distributions obtained by tracking data from Fowler *et al.* and red lines indicate multiple Gaussian fitted to the experimental traces (2).

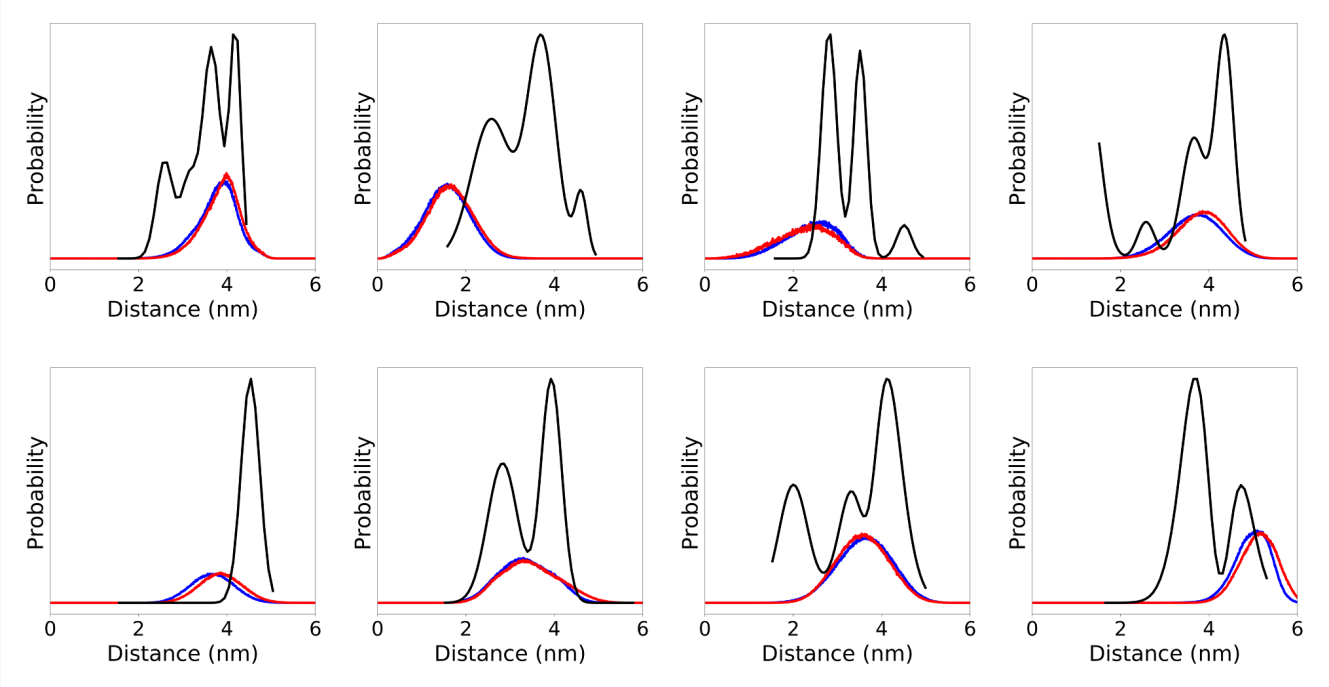

Figure 24: Experimentally characterized residue-pair distance distributions obtained using rotamer library analysis of MD data in POPE/POPG (3:1 ratio) bilayer (color: blue) and BDDM micelle (color: red). Black lines show experimental DEER distance distributions obtained from Fowler *et al.* as discussed above (2). The eight DEER distance distributions shown correspond to distance between residue-pairs 86-432, 141-432, 141-438, 141-500, 201-364, 47-330, 174-401, and 174-466) respectively.

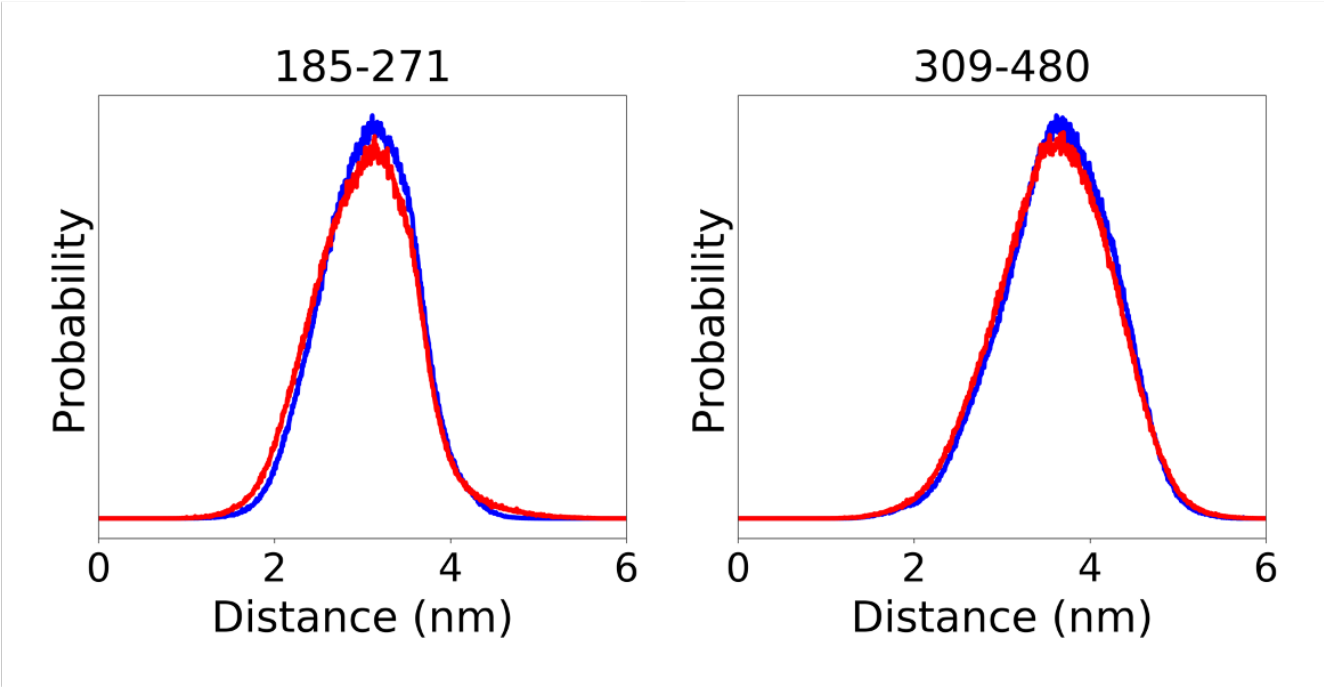

Figure 25: Experimentally characterized residue-pair distance distributions obtained using rotamer library analysis of MD data in POPE/POPG (3:1 ratio) bilayer (color: blue) and BDDM micelle (color: red).

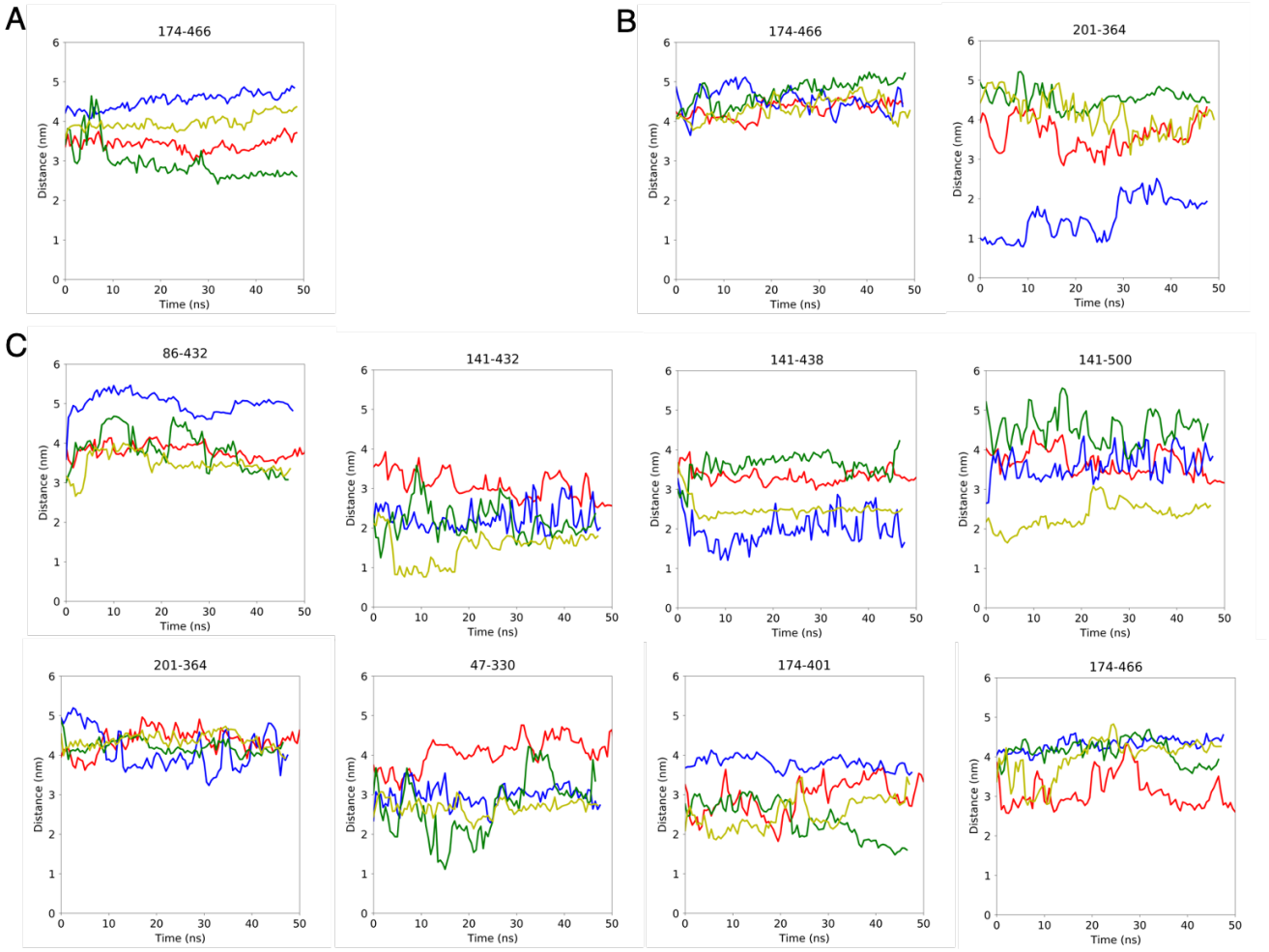

Figure 26: Distance variations with time for unbiased simulation of frames from reMD simulation with probes on (A) one, (B) two, and (C) eight residue pairs.

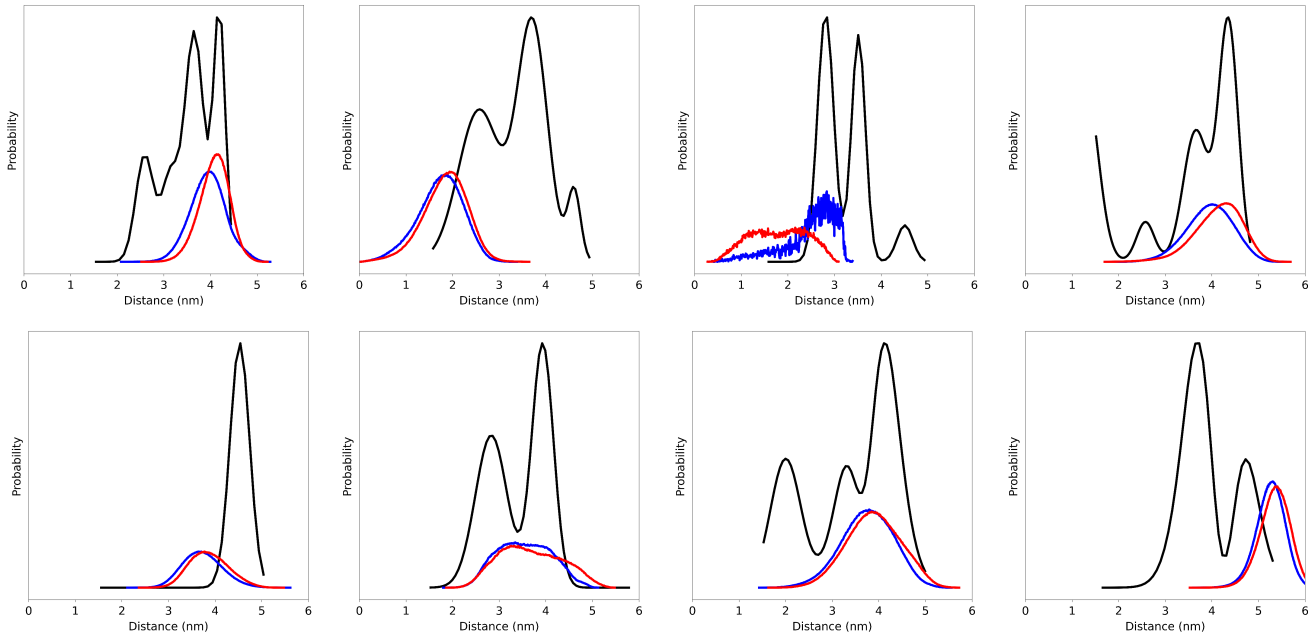

Figure 27: Experimentally characterized residue-pair distance distributions obtained using MtsslWizard analysis of MD data in POPE/POPG (3:1 ratio) bilayer (color: blue) and BDDM micelle (color: red) (4). Black lines show experimental DEER distance distributions obtained from Fowler *et al.* (2) . The eight DEER distance distributions shown correspond to distance between residue-pairs 86-432, 141-432, 141-438, 141-500, 201-364, 47-330, 174-401, and 174-466) respectively.

Table 4: Comparison of means and medians of extracellular and intracellular distance distributions of $PepT_{So}$ in micelle and bilayer (POPE/POPG) compositions. These distances were calculated using "Closest-heavy" distance scheme in MDTraj 1.7 (1). The standard errors in the mean calculations are calculated using bootstrapping procedure. 100 bootstrapped samples were selected where each sample consists of 80% of the total simulation. The residue pairs selected for the distance measurements are obtained from Fowler et al. (2)

Residue Position	Number	Mean ± Std. Err (Bilayer)	Mean ± Std. Err (Micelle)	Median (Bilayer)	Median (Micelle)
Extracellular	1	2.50 ± 0.00	2.51 ± 0.00	2.50	2.43
Extracellular	2	2.96 ± 0.00	2.90 ± 0.00	2.98	2.90
Extracellular	3	3.87 ± 0.00	4.00 ± 0.00	3.87	4.00
Intracellular	1	2.58 ± 0.00	2.65 ± 0.00	2.60	2.65
Intracellular	2	1.24 ± 0.00	1.27 ± 0.00	1.24	1.32
Intracellular	3	1.58 ± 0.00	1.56 ± 0.00	1.61	1.61
Intracellular	4	3.12 ± 0.00	3.18 ± 0.00	3.12	3.24
Intracellular	5	3.35 ± 0.00	3.52 ± 0.00	3.34	3.50

Table 5: Comparison of means and medians of extracellular and intracellular distance distributions of LeuT in micelle and bilayer (POPE/POPG) compositions. These distances were calculated using "Closest-heavy" distance scheme in MDTraj 1.7 (1). The standard errors in the mean calculations are calculated using bootstrapping procedure. 100 bootstrapped samples were selected where each sample consists of 80% of the total simulation. The residue pairs selected for the distance measurements are obtained from Kazmier et al. (3)

Residue Position	Number	Mean ± Std. Err (Bilayer)	Mean ± Std. Err (Micelle)	Median (Bilayer)	Median (Micelle)
Extracellular	1	2.91 ± 0.00	2.96 ± 0.00	2.92	2.97
Extracellular	2	3.97 ± 0.00	3.99 ± 0.00	4.02	4.08
Extracellular	3	3.26 ± 0.00	3.31 ± 0.00	3.37	3.42
Extracellular	4	3.04 ± 0.00	2.99 ± 0.00	3.03	3.01
Extracellular	5	2.34 ± 0.00	2.37 ± 0.00	2.42	2.44
Extracellular	6	2.25 ± 0.00	2.27 ± 0.00	2.12	2.16
Extracellular	7	1.80 ± 0.00	1.84 ± 0.00	1.84	1.86
Intracellular	1	2.13 ± 0.00	2.10 ± 0.00	2.07	2.05
Intracellular	2	2.46 ± 0.00	2.43 ± 0.00	2.47	2.45
Intracellular	3	2.82 ± 0.00	2.75 ± 0.00	2.80	2.78
Intracellular	4	2.43 ± 0.00	2.37 ± 0.00	2.35	2.27
Intracellular	5	3.43 ± 0.00	3.33 ± 0.00	3.36	3.28
Intracellular	6	2.70 ± 0.00	2.68 ± 0.00	2.70	2.69
Intracellular	7	2.19 ± 0.00	2.22 ± 0.00	2.18	2.19
Intracellular	8	1.90 ± 0.00	1.91 ± 0.00	1.94	1.93
Intracellular	9	3.01 ± 0.00	2.98 ± 0.00	2.99	2.99
Intracellular	10	2.16 ± 0.00	2.16 ± 0.00	2.14	2.13
Intracellular	11	2.32 ± 0.00	2.31 ± 0.00	2.27	2.28
Intracellular	12	2.39 ± 0.00	2.43 ± 0.00	2.40	2.42
Intracellular	13	1.83 ± 0.00	1.84 ± 0.00	1.84	1.85
Intracellular	14	1.74 ± 0.00	1.70 ± 0.00	1.87	1.85
Intracellular	15	2.66 ± 0.00	2.65 ± 0.00	2.69	2.69
Intracellular	16	2.06 ± 0.00	2.10 ± 0.00	2.20	2.20
Intracellular	17	2.03 ± 0.00	2.05 ± 0.00	2.06	2.08

REFERENCES

1. R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane and V. S. Pande, *Biophysical Journal*, 2015, **109**, 1528–1532.

2. P. W. Fowler, M. Orwick-Rydmark, S. Radestock, N. Solcan, P. M. Dijkman, J. A. Lyons, J. Kwok, M. Caffrey, A. Watts, L. R. Forrest and S. Newstead, *Structure*, 2015, **23**, 290–301.

3. K. Kazmier, S. Sharma, M. Quick, S. M. Islam, B. Roux, H. Weinstein, J. A. Javitch and H. S. Mchaourab, *Nature Structural & Molecular Biology*, 2014, **21**, 472.

4. G. Hagelueken, R. Ward, J. H. Naismith and O. Schiemann, *Applied Magnetic Resonance*, 2012, **42**, 377–391.

5. B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *Journal of Computational Chemistry*, 2009, **30**, 1545–1614.

6. J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, C. L. Brooks, A. D. MacKerell, J. B. Klauda and W. Im, *Journal of Chemical Theory and Computation*, 2015, **12**, 405–413.

7. E. L. Wu, X. Cheng, S. Jo, H. Rui, K. C. Song, E. M. Dávila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R. M. Venable, J. B. Klauda and W. Im, *Journal of Computational Chemistry*, 2014, **35**, 1997–2004.

8. X. Cheng, S. Jo, H. S. Lee, J. B. Klauda and W. Im, *Journal of Chemical Information and Modeling*, 2013, **53**, 2171–2180.

9. Y. Qi, J. Lee, X. Cheng, R. Shen, S. M. Islam, B. Roux and W. Im, *Journal of Computational Chemistry*, 2019, **41**, 415–420.

10. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *Journal of Computational Chemistry*, 2005, **26**, 1781–1802.

11. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *The Journal of Physical Chemistry B*, 1998, **102**, 3586–3616.

12. K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell, *Journal of Computational Chemistry*, 2009, **31**, 671–690.

13. J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell and R. W. Pastor, *The Journal of Physical Chemistry B*, 2010, **114**, 7830–7843.

14. R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig and A. D. MacKerell, *Journal of Chemical Theory and Computation*, 2012, **8**, 3257–3273.

15. X. Cheng, J.-K. Kim, Y. Kim, J. U. Bowie and W. Im, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2016, **1858**, 1566–1572.

16. J. Lipfert, L. Columbus, V. B. Chu, S. A. Lesley and S. Doniach, *The Journal of Physical Chemistry B*, 2007, **111**, 12427–12438.

17. L. Columbus, J. Lipfert, K. Jambunathan, D. A. Fox, A. Y. L. Sim, S. Doniach and S. A. Lesley, *Journal of the American Chemical Society*, 2009, **131**, 7320–7326.

18. P. Strop and A. T. Brunger, *Protein Science*, 2005, **14**, 2207–2211.

19. *PubChem Database*.

20. B. Selvam, S. Mittal and D. Shukla, *ACS Central Science*, 2018, **4**, 1146–1154.

21. H. Krishnamurthy and E. Gouaux, *Nature*, 2012, **481**, 469–474.

22. N. Eswar, B. Webb, M. A. Marti-Renom, M. Madhusudhan, D. Eramian, M. yi Shen, U. Pieper and A. Sali, *Current Protocols in Bioinformatics*, 2006, **15**, 5.6.1–5.6.30.

23. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *Journal of Computational Chemistry*, 2004, **25**, 1605–1612.

24. K. Gotfryd, T. Boesen, J. S. Mortensen, G. Khelashvili, M. Quick, D. S. Terry, J. W. Missel, M. V. LeVine, P. Gourdon, S. C. Blanchard, J. A. Javitch, H. Weinstein, C. J. Loland, P. Nissen and U. Gether, *Nature Communications*, 2020, **11**, 10.1038/s41467–020–14735–w.

25. W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics*, 1996, **14**, 33–38.

26. S. Jo, X. Cheng, S. M. Islam, L. Huang, H. Rui, A. Zhu, H. S. Lee, Y. Qi, W. Han, K. Vanommeslaeghe, A. D. MacKerell, B. Roux and W. Im, in *Advances in Protein Chemistry and Structural Biology*, Elsevier, 2014, pp. 235–265.

27. R. Shen, W. Han, G. Fiorin, S. M. Islam, K. Schulten and B. Roux, *PLOS Computational Biology*, 2015, **11**, e1004368.

28. A. Yamashita, S. K. Singh, T. Kawate, Y. Jin and E. Gouaux, *Nature*, 2005, **437**, 215–223.

29. S. K. Singh, A. Yamashita and E. Gouaux, *Nature*, 2007, **448**, 952–956.

30. S. K. Singh, C. L. Piscitelli, A. Yamashita and E. Gouaux, *Science*, 2008, **322**, 1655–1661.

31. M. Quick, A.-M. L. Winther, L. Shi, P. Nissen, H. Weinstein and J. A. Javitch, *Proceedings of the National Academy of Sciences*, 2009, **106**, 5563–5568.

32. Z. Zhou, J. Zhen, N. K. Karpowich, C. J. Law, M. E. A. Reith and D.-N. Wang, *Nature Structural & Molecular Biology*, 2009, **16**, 652–657.

33. H. Wang, J. Elferich and E. Gouaux, *Nature Structural & Molecular Biology*, 2012, **19**, 212–219.

34. L. Malinauskaite, S. Said, C. Sahin, J. Grouleff, A. Shahsavar, H. Bjerregaard, P. Noer, K. Severinsen, T. Boesen, B. Schiøtt, S. Sinning and P. Nissen, *Nature Communications*, 2016, **7**, 10.1038/ncomms11673.

35. Y. Polyhach, E. Bordignon and G. Jeschke, *Physical Chemistry Chemical Physics*, 2011, **13**, 2356–2366.

36. L. S. Stelzl, P. W. Fowler, M. S. Sansom and O. Beckstein, *Journal of Molecular Biology*, 2014, **426**, 735–751.

37. N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, *Journal of Computational Chemistry*, 2011, **32**, 2319–2327.

38. R. Gowers, M. Linke, J. Barnoud, T. Reddy, M. Melo, S. Seyler, J. Domański, D. Dotson, S. Buchoux, I. Kenney and O. Beckstein, Proceedings of the 15th Python in Science Conference, 2016.

39. C. D. Bruce, M. L. Berkowitz, L. Perera and M. D. E. Forbes, *The Journal of Physical Chemistry B*, 2002, **106**, 3788–3793.

40. P. J. Bond and M. S. Sansom, *Journal of Molecular Biology*, 2003, **329**, 1035–1053.

41. M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg and A. L. Lomize, *Nucleic Acids Research*, 2011, **40**, D370–D376.

42. A. S. Moffett, K. W. Bender, S. C. Huber and D. Shukla, *Biophysical Journal*, 2017, **113**, 2354–2363.