PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d2cp02890e

Reconciling membrane protein simulations with experimental DEER spectroscopy data†

Shriyaa Mittal,[a] Soumajit Dutta[b] and Diwakar Shukla *[abcde]

Spectroscopy experiments are crucial to study membrane proteins for which traditional structure determination methods still prove challenging. Double electron-electron resonance (DEER) spectroscopy experiments provide protein residue-pair distance distributions that are indicative of their conformational heterogeneity. Atomistic molecular dynamics (MD) simulations are another tool that have been proven to be vital to study the structural dynamics of membrane proteins such as to identify inward-open, occluded, and outward-open conformations of transporter membrane proteins, among other partially open or closed states of the protein. Yet, studies have reported that there is no direct consensus between the distributional data from DEER experiments and MD simulations, which has challenged validation of structures obtained from long-timescale simulations and using simulations to design experiments. Current coping strategies for comparisons rely on heuristics, such as mapping the nearest matching peaks between two ensembles or biased simulations. Here we examine the differences in residue-pair distance distributions arising due to the choice of membranes around the protein and covalent modification of a pair of residues to nitroxide spin labels in DEER experiments. Through comparing MD simulations of two proteins, $PepT_{So}$ and LeuT–both of which have been characterized using DEER experiments previously–we show that the proteins' dynamics are similar despite the choice of the detergent micelle as a membrane mimetic in DEER experiments. On the other hand, covalently modified residues show slight local differences in their dynamics and a huge divergence when the oxygen atom pair distances between spin labeled residues are measured rather than protein backbone distances. Given the computational expense associated with pairwise MTSSL labeled MD simulations, we examine the use of biased simulations to explore the conformational dynamics of the spin labels only to reveal that such simulations alter the underlying protein dynamics. Our study identifies the main cause for the mismatch between DEER experiments and MD simulations and will accelerate the development of potential mitigation strategies to improve the match.

Received 25th June 2022,
Accepted 23rd January 2023

DOI: 10.1039/d2cp02890e

rsc.li/pccp

Introduction

DEER spectroscopy, also known as pulsed electron–electron double resonance (PELDOR), has made incredible progress in the study of biomolecules such as cytoplasmic and membrane proteins and nucleic acids,[1–4] including experiments *in vitro* and *in vivo*.[5–11] In DEER experiments, a spin probe is covalently attached to two residues on the biomolecules. Distances between these two spin probes can be determined by measuring the dipolar coupling between an electron pair and one unpaired electron on each of the spin probes. The interaction between electrons is measured in the time domain and then mathematically transformed into distance distributions. Methodological developments have made it possible to obtain distance distributions up to 10 nm in cytoplasmic proteins and 8 nm in membrane proteins,[3,12–14] and up to 16 nm with sparse spin-labeling that can avoid the deleterious impact of multiple spin labels in close proximity.[15] Conformational ensembles obtained from these distance distributions help to elucidate important conformational changes and metastable states in biomolecules. This information can be further used as a restraint in integrative structural modeling of large proteins.[16] Recently, DEER has been used to capture the sequence of ligand-induced conformational changes in the protein on the angstrom lengthscale and the

[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

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d2cp02890e

sub-milisecond timescale.[17] This will provide not only structural details of biomolecules but also a mechanistic understanding of biomolecular function. However, the current status of the DEER technology has space for improvement. (1) It is still hard to attach probes in residues deeply embedded inside a protein as it might impair protein functionality.[18] (2) As discussed by Hett *et al.*, time resolved DEER is not able to follow processes that occur below microseconds due to limitations of the Microsecond freeze Hyper-Quench device.[17] (3) Limitations with respect to the distance range, flexibility and data quality for DEER measurement were also discussed by Schiemann *et al.*[1] Therefore, other computational and experimental techniques can be used to further validate or add more information to study protein structure and dynamics.

One such computational technique to study time-resolved protein conformational changes is molecular dynamics (MD) simulation. Given the advance in computational resources, there are numerous extensive MD simulation studies of membrane proteins including GPCRs, transporters, ion channels, integrins, and transmembrane receptor kinases.[19–21] The information obtained from DEER experiments, residue-pair distance distributions, can be directly compared to dynamics information from MD simulations in order to characterize the structural consequences of the obtained distance distributions. Yet, there is often no direct consensus between the distributional data from DEER experiments and MD simulations, which has challenged the validation of structures obtained from long-timescale simulations. Several methods have been introduced to reconcile the experimentally characterized distance distributions with simulations such as restrained ensemble MD (reMD)[22,23] and ensemble-biased metadynamics (EBMetaD)[24] simulations, both methods employ the experimentally obtained distance distribution to bias a simulation ensemble. Another method to syncretize unbiased MD simulations with experiments is labeling a residue with a spin probe whose conformational orientations are sampled using a spin probe rotamer library[25,26] or molecular dynamics dummy spin-label (MDDS) simulations,[27,28] where an artificial dummy atom is used to consider spin label motions. These methods are independent of any experimental data bias and relatively computationally inexpensive since no additional simulations are required, but are unable to consider the protein's conformational dynamics.

Typically, we observe mismatches in terms of relative peak heights when there are multiple peaks in the distance distributions, peak positions, and lower and higher extremes of the distance values. Commonly we observe that experimental distributions exhibit larger distance values, which are not sampled in any of the MD simulation ensembles. These differences can be visualized in Fig. S1A (ESI†) where we compare distance distributions from our previous simulations on a peptide transporter protein with experimental DEER distributional data. Most potential for mismatch between experiments and simulation distance distributions stems from differences between experimental conditions and standard simulation protocols. Since membrane proteins are embedded in lipid bilayers under physiological conditions, simulations are typically performed in lipid bilayers. These lipid bilayers can be homogeneous or heterogeneous with different types of lipid molecules.[29] Bilayer mimetics such as nanodiscs,[30] lipodisq nanoparticles,[31] bicelles,[32] liposomes,[33] micelles[34] are more amenable to biophysical experiments and have been used for DEER spectroscopy studies of membrane proteins. Specifically, detergent micelles are most commonly used and a widely used detergent is n-dodecyl-β-D-maltoside (BDDM).

Another significant basis for a mismatch between observed peaks in experiments and simulations is the use of spin probes in DEER experiments, which is absent in wild-type protein simulations. Using site-directed spin labeling (SDSL), two nitroxide spin labels are attached to two cysteine mutated residues. These spin labels can be of different types such as 1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl)methanethiosulfonate (MTSSL), iodoacetamide-PROXYL, unnatural amino acids *p*-acetyl-l-phenylalanine and 2,2,6,6-tetramethylpiperidine-1-oxyl-4-amino-4-carboxylic acid, and a spin-labeled lysine. DEER experimental measurements among two spin labels are a proxy to explain the protein's residue-pair distances. Relying on cysteine modifications and the addition of flexible spin probe molecules pose a possibility of modifying the observed protein's dynamics from DEER experiments. For example, the MTSSL spin probe has five linker dihedrals attributing large rotational flexibility to the protein residue.[26] Recently metal cations such as Gd^{3+}, Cu^{2+} and Mn^{2+} based spin labels that are more rigid have been used[35–37] but their applications in the study of membrane proteins are limited.[38]

Based on the above discussed modifications in DEER experiments as compared to physiological conditions, we propose five potential impacts on a protein, its dynamics, and hence the observed DEER experimental observables. Since DEER experiments are typically performed with proteins embedded in bilayer mimetics, such as detergent micelles rather than lipid bilayers, membrane diffusion, packing flexibility and interactions can (1) allow for shifts in DEER distributions and peaks and (2) alter the secondary structure and accessibility of various helices and loops in the protein. Previous studies that draw comparisons between micelle and bilayer environments on membrane proteins have been limited to either small peptides such as single transmembrane helices or are based on ns-timescale simulations that do not provide a realistic picture of a protein's conformational dynamics. (3) Since DEER measurements require a covalent modification at at least two sites of the protein, we evaluate whether this modulates the underlying free energy landscape of the protein by biasing it to adopt only a subset of the available conformations. In addition to modulating the protein overall dynamics, we examine the impact that the MTSSL probes have locally on the modified residues, their neighboring residues, and their structural properties. (4) We also examine how accurately do the distance distributions obtained from the dipolar coupling of MTSSL spin nitroxide probes provide an approximation of protein dynamics where there is no MTSSL probe. (5) Multiple flexible bonds of nitroxide spin probes[26] such as MTSSL spin probes may have different timescales compared to those from the wild-type residue which will equilibrate at a different timescale from that of the protein

changing the experimentally observed dipolar couplings. We evaluate these perturbations and their impacts in this work to discern which among these is the main cause for the mismatch between experiments and simulations.

Here, we directly compare the biophysical effect of different experimental and simulation conditions by performing MD simulations under conditions similar to experimental conditions. To evaluate the effect of membrane environments on the protein structure and dynamics, we compared long-timescale simulations of two proteins in a BDDM micelle and a more typical lipid bilayer. Specifically, we performed simulations of two proteins, $PepT_{So}$ and LeuT, which are biologically important representative proteins of two different membrane protein families, major facilitator superfamily (MFS) and neurotransmitter: sodium symporter (NSS), respectively. Residue pairs in both proteins have been previously characterized using DEER experiments.[14,34,39,40] LeuT has many three-dimensional structures determined through X-ray crystallography and has been investigated using computational simulations. Recently, two crystal structures of $PepT_{So}$ were resolved[34,41] and we had examined this protein using MD simulations in our previous work.[20] We followed our micelle and bilayer simulations by introducing nitroxide spin labels MTSSL on a pair of residues in $PepT_{So}$ to examine the perturbations caused by the probe's site-specific mutations during DEER spectroscopy experiments. We then performed restrained ensemble molecular dynamics (reMD) simulations to evaluate the spin pair equilibration and its impact on the protein's conformational landscape and residue-pair distance distributions.

Results

Residue-pair distances from proteins in micelles resemble trends in bilayer-embedded proteins

$PepT_{So}$ is a proton-coupled bacterial symporter for which, recently, researchers characterized eight inter-residue distance distributions using DEER.[34] There are two known crystal structures for this protein found in the bacteria *Shewanella oneidensis*, 2XUT,[41] and 4UVM,[34] both in the inward-facing conformations of the protein. $PepT_{So}$ belongs to the proton-dependent oligopeptide transporter (POT) family and the major facilitator superfamily (MFS) whose members have a wide variety of functions and are found in many different organisms including humans. All MFS transporters share a common structural fold consisting of 12 transmembrane helices;[43] however, POT family transporters within MFS often have two additional helices. Like most POT family transporters, $PepT_{So}$ has 14 transmembrane helices.

LeuT, a leucine transporter, has many high-resolution crystal structures and has been extensively characterized using DEER experiments.[14,39,40] LeuT belongs to the neurotransmitter: sodium symporter (NSS) family whose other members include dopamine, noradrenaline, GABA, glycine, and serotonin transporters. LeuT was the first structure resolved using X-ray crystallography from the NSS family and consists of 12 transmembrane helices. Although many structures have been resolved since then, only one structure is inward-facing as a quadruple mutant (3TT1[44]).

$PepT_{So}$ and LeuT are model proteins from two different families of membrane proteins. While LeuT has been studied using computational simulations with both unbiased and biased protocols, there are only a few short-timescale computational studies focused on $PepT_{So}$. Our previous work sampled the conformational dynamics of $PepT_{So}$ using long-timescale 54 ms MD simulations and analyzed its equilibrium dynamics using Markov state model based analysis.[20] These simulations were carried out in a POPC bilayer using the AMBER FF14SB force field. To compare the dynamics of $PepT_{So}$ protein in detergent micelles and bilayers and solely capture the effect of the membrane environment, we replicated our simulations in a POPE/POPG bilayer using the CHARMM 36 force field. Simulations from our previous work[20] provide a benchmark for sufficient conformational sampling since we were able to sample IF, OC, OF, and multiple other intermediate states (Fig. S8, ESI†). Here, we compare our atomistic molecular dynamics simulations of $PepT_{So}$ and LeuT in BDDM micelles and lipid bilayers. Length of individual simulation data sets vary between $\sim$20–32 μs simulation time (Table S1, ESI†).

As shown in Fig. S8A and B (ESI†), we projected our $PepT_{So}$ simulation data sets on gating residue pairs, Ser131-Tyr431 on the intracellular side and Arg32-Asp316 on the extracellular distance. We compared the sampled regions with our previous simulations (Fig. S8D, ESI†) to conclude that all physical conformations of the protein have been well sampled. Similarly, as shown in Fig. S9 (ESI†) we projected our LeuT simulation data sets on one residue pair each on the intracellular and extracellular side of the protein, Arg5-Asp369 and Arg30-Asp404, respectively. These residue pairs are based on gating residues identified in hSERT.[45] This human serotonin transporter has a typical LeuT-fold and shares 35.5% sequence similarity with LeuT protein. Among the gating residues, Asp404 from LeuT is homologous to Glu493 in hSERT and the other three residues are arginines.

Upon comparing simulated and experimental distance distributions from our micelle and bilayer simulations (Fig. S1B and C, ESI†), we see that distance distributions obtained from micelle simulations are no better at matching with experiments. However, by comparing distance distributions shown in Fig. 1, we examined the impact of the choice of membranes on the protein's dynamics. For $PepT_{So}$ protein, five out of a total of eight distance distributions show a higher median value (middle horizontal line on violin plots in Fig. 1A) in the micelle as compared to the bilayer. For instance, residue pair 174-466 shows a single peak in the distributions for both the micelle and bilayer, but the data has a median value of 3.87 nm mean in the bilayer whereas this value is 4.00 nm in the micelle. On the other hand, two distance distributions for residue pairs 47-330 and 174-401 show lower median values in the micelle than in the bilayer. One distance distribution for residue pair 141-438 shows about the same value of 1.6 nm in both the micelle and bilayer. In Fig. S10 (ESI†), we show that most of the mean or median values lie along the black dotted line, indicating that they are similar for micelles and bilayers. Mean and median values for all inter-helix distances also fall along the

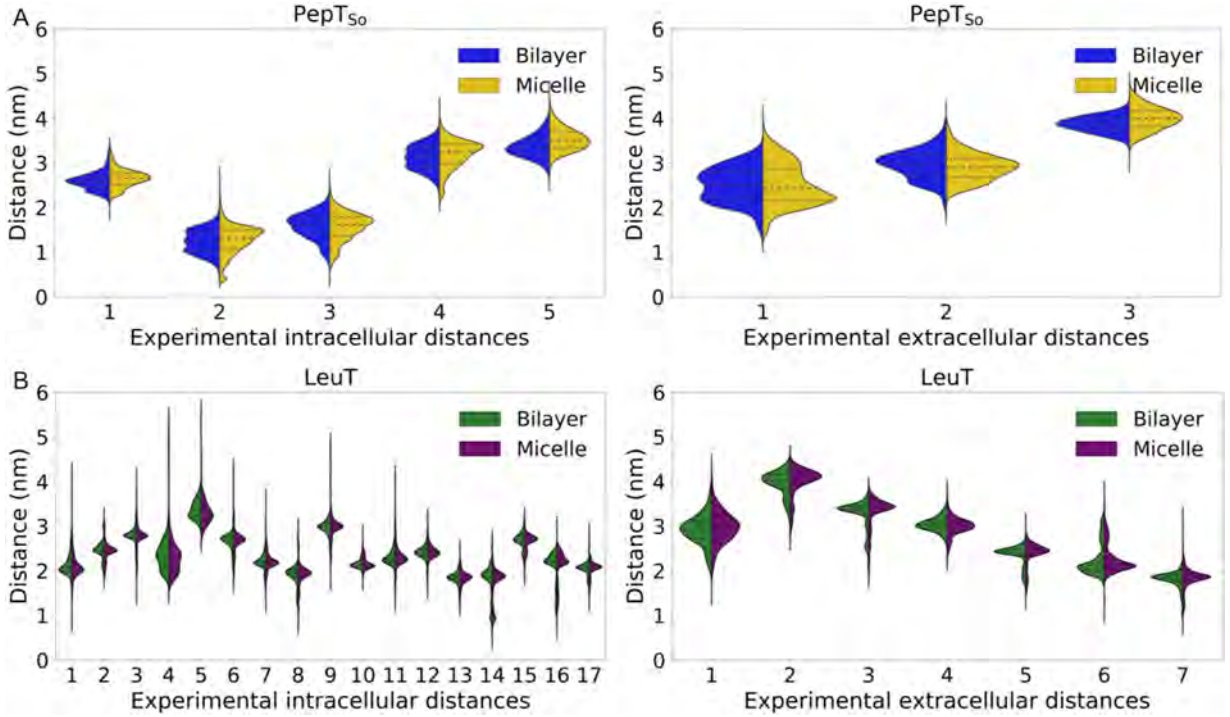

Fig. 1 (A) Violin plot shows distance distributions for five intracellular residue-pair distances and three extracellular residue-pair distances measured by Fowler *et al.* as observed from MD simulations of $PepT_{So}$ protein in micelles (yellow, right) and bilayers (blue, left).[34] (B) Violin plot shows distance distributions for 17 intracellular residue-pair distances and seven extracellular residue-pair distances measured by Kazmier *et al.* as observed from MD simulations of LeuT protein in micelles (purple, right) and bilayers (green, left).[40] Distances calculated here are between the residue pair non-hydrogen atoms that are closest. These distances were calculated using "closest-heavy" distance scheme in MDTraj.[42] Distance distributions among C_α atoms and sidechain atoms are shown in Fig. S21 and S22 (ESI†).

dotted line indicating that the differences are minimal (Table S4, ESI†).

Based on visual inspection, not only the positions of the peaks but the number of peaks in the distance distributions can differ such as three peaks in bilayers *versus* two in micelles for residue pairs 141-432 and 141-500. Interestingly, these two distance distributions show new peaks in micelles where little or no data is seen in those regions in the bilayer. For LeuT, although the distance distributions differ, the variation is much less (Fig. 1B and Fig. S10, Table S5, ESI†), for example, none of the 24 experimental distance distributions show a peak in bilayers which is not there in micelles or *vice versa*.

For $PepT_{So}$, five of the experimental residue-pair distance distributions also show slightly broader distributions. For inter-helix distributions (Fig. S10, ESI†) we see that few upper values and lower values lie below the dotted line, meaning that the distributions move towards larger values in micelles. Does this mean that micellar environments shift the distributions to larger values? This is unlikely because for LeuT, we see values that are both above and below the black dotted line in experimental distances as well as inter-helix distances.

We conclude that the reason for the mismatch between DEER experimental observables and MD simulations distance distributions is not due to the use of detergent micelles in MD simulations, since there are no dramatic or homogeneous shifts in the distance distributions from our simulations. However, this observation may be limited to two proteins in the current study as the effect of the membrane on protein dynamics has been established.[46]

Proteins in micelles and bilayers show structural similarity

For both proteins, we have measured and shown the helicity of transmembrane helices in Fig. S11 (ESI†). Values closer to 1 indicate helical nature and decreasing values show loss of helicity. TMs 7 and 10 exhibit a wider range of helicity in $PepT_{So}$ which indicates their dynamic nature. Selvam *et al.* reported that one of the extracellular gating residues is on TM7 and one of the intracellular gating residues is on TM10.[20] Given that the median of TM7 helicity is 0.76 in both micelles and bilayers, lowest among all other transmembrane helices, none of the helices lose their entire alpha-helical nature. Moreover, broader distributions for TMs 7 and 10 are seen in both micellar and bilayer environments.

TMs 1 and 6 in LeuT show wide helicity ranges in both the environments. Readers must note that TM1 here indicates residues of TM1a, the first half of the TM1 helix. TM1a is of particular interest in LeuT and other NSS family transporters[44,47,48] because in IF structures this region is away from the bundle as shown in Fig. S12C (ESI†). Low values of TM1 helicity arise from IF trajectories and other trajectories that transition to IF like states. Our simulation ensemble includes two independent trajectories based on the IF structure 3TT3.[44] LeuT TM1a dynamics show a

significant distinction between OF and IF states in our MD simulation trajectories (Fig. S12, ESI†); TM1a helicity drops to 20–30% in IF trajectories whereas this is 50–80% in OF trajectories. Due to the dynamic nature of this region, it follows that one of the gating residues on both the intracellular and extracellular side of the protein is also positioned on TM1. This distinct behavior of TM1a is also seen in Fig. S9A and B (ESI†), where LeuT is open on both extracellular and intracellular sides. Other studies on transporter proteins using extensive MD simulations[45,49,50] have also reported observation of this hourglass-like state of the transporter. Terry *et al.* have reported evidence for this conformation in LeuT which is due to a weaker coupling between the extracellular and intracellular side of LeuT.[51] We suggest that this weaker coupling allows LeuT to explore a large range of intracellular gating distance while the extracellular side of the protein is also open.

TM regions of $PepT_{So}$ and LeuT show structural similarity in both the micelle and bilayer, but could the choice of the membrane milieu affect the intracellular and extracellular flanking regions of our proteins? We compared the distributions for these regions such as the helicity of two short helices in $PepT_{So}$ one on each side. For LeuT, we compared the helical content of the loop regions which connect the TM helices. Fig. S13 and S14 (ESI†) show that distance distributions are similar and not impacted by the choice of the membrane environment. In this work, we do not consider the molecular-level differences in protein residue interactions with lipids or detergents and differences in membrane curvature that could lead to variation in the stability of loop regions.

Fig. 2 strikingly shows that TM helicity median and mean values lie along the black-dotted line, and in most cases, lower and upper values also do not deviate much in the micelle and bilayer. In general, helicity values or distributions are not different which means that micelles do not impact the structure of the protein.

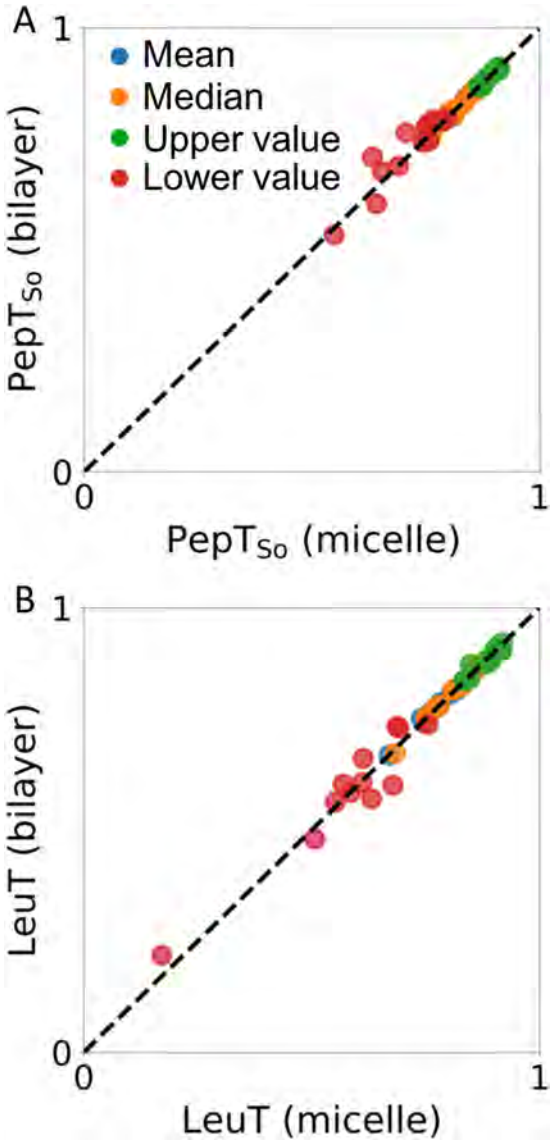

Fig. 2 Comparing the mean (blue), median (orange), upper value (green), and lower value (red) of alpha-helical content of (A) 14 TM helices in $PepT_{So}$, and (B) 12 TM helices in LeuT. Markers below the black dotted line indicate larger values observed in the micelle environment. Markers above the black dotted line indicate larger values observed in the bilayer environment. Markers along the black dotted line indicate similar observations in micelle and bilayer simulations. The size of the dots does not indicate any measured quantity.

Covalent modification due to MTSSL probes cause small local structural perturbations in the protein

We calculated Kullback–Leibler (KL) divergence among distance distributions from MD simulations in micelles and MD simulations in bilayers discussed above. Among the eight experimentally characterized distances in $PepT_{So}$, we found that residue pair Asn174-Ser466 has the highest KL divergence value. Hence, we chose this residue pair for further studies, specifically to perform simulations with realistic nitroxide DEER labels. We attached an MTSSL DEER probe on Asn174 and Ser466 after mutating them to cystines *via* CHARMM-GUI and simulated our protein in a BDDM micelle for $\sim$19 μs. Previous studies have compared MD simulations with DEER experimental results with short-timescale MD simulations with explicit spin probes in a variety of biological systems.[52–55] To our knowledge, this is the first study of the impact of MTSSL spin labels on a membrane protein using long-timescale unbiased simulations.

Fig. 3A shows the simulated conformational ensemble projected on the intracellular and extracellular gating residues. Comparing this landscape to those for the $PepT_{So}$ simulations without probes in micelles shows that both ensembles capture all conformational states of the protein. This follows that probe molecules do not seem to interfere with the conformational dynamics of $PepT_{So}$ protein in a way that could hinder its transport function.

To understand the local effects of the MTSSL probes on the protein, we calculate Phi and Psi dihedral angles and generate Ramachandran plots for the mutated residues 174 and 466. We see a slightly larger coverage for residue 174 with the MTSSL probe (Fig. S15B, ESI†) as opposed to when it is an Asn residue

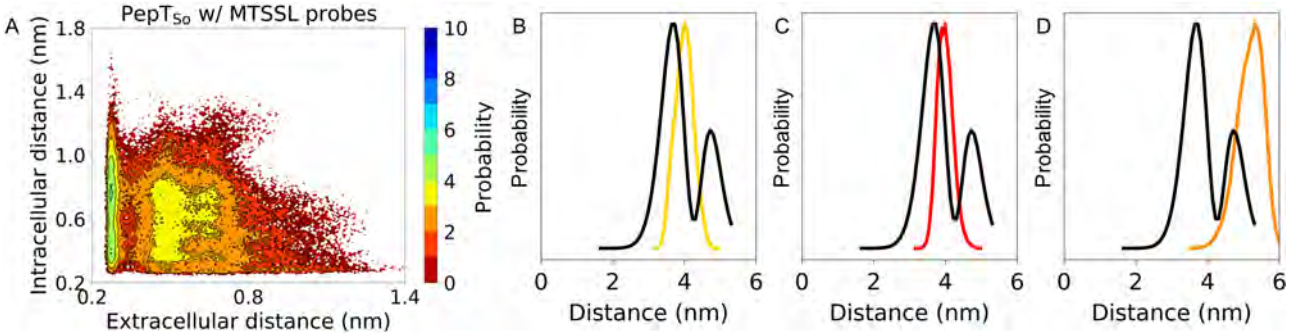

Fig. 3 (A) Conformational landscapes of $PepT_{So}$ protein are generated by projecting all simulation data on the chosen extracellular and intracellular side distances measured between Arg32-Asp316 and Ser131-Tyr431, respectively. Conformational landscape for $PepT_{So}$ MD simulations in BDDM micelles with an MTSSL labeled residue pair. Distances calculated here are between the residue pair non-hydrogen atoms that are closest in distance. These distances were calculated using "closest-heavy" distance scheme in mdtraj.[42] (B) Distance distribution for MTSSL labeled residue pair in $PepT_{So}$, 174-466, from simulations in BDDM micelle without probes (yellow), and (C) with probes (red). (D) Distance distribution for the MTSSL labeled residue pair in $PepT_{So}$, 174-466, from simulations in BDDM micelles without probes (orange) where distances are measured between ON atoms on MTSSL labels. Black lines show DEER experiment distance distributions.

(Fig. S15A, ESI†), while there is no difference for residue 466. Similarly, when we look at the regions surrounding the labeled residues, specifically two residues both before and after the labeled residues, we see a larger distribution for residue 174 (Fig. S15E and F, ESI†). Hence, we conclude that this mutant created for DEER spectroscopy experiments slightly impacts the local dynamics and secondary structure of the protein, and this effect does not alter the overall conformational dynamics of the protein.

We suggest that any alteration seen in transportation activity could be due to the kinetic rates of the transport function that would not affect the DEER observations unless functional interactions are mutated. Fowler *et al.* tested the transportation activity of their $PepT_{So}$ double cysteine mutants and 174-466 mutant although decreased activity, did not stop AlaAla transport entirely.[34] Kazmier *et al.* also tested binding of Leu to spin-labeled LeuT pairs and most double mutants retained more than 50% binding affinity as the wild type protein.

We examined the impact of a spin-probe labeled residue pair on the resulting distance distributions (Fig. S16, ESI†) by comparing micelle simulations with and without MTSSL probes. Since the probe molecules are on the extracellular side of the $PepT_{So}$ protein, two of the three extracellular side distance distributions do appear slightly perturbed. We observe that the intracellular side distances show no differences. A closer look at the distribution for the residue pair labeled with MTSSL probes shows that the quartile values are conserved. Overall, we do not see any significant changes in the distance distributions for all eight experimental distances as compared to the distance distributions obtained from simulations in BDDM micelles. As expected, if there is no overall difference in the underlying conformational landscape as we discussed above, individual residue-pair distances also would not deviate.

Distance distributions obtained from unbiased MD simulation with MTSSL probes are different from those obtained from the DEER experiment

The mean of the unpaired electron density of paramagnetic probes is concentrated close to the N–O bond.[55] Distance distributions from DEER are estimated from the dipolar coupling between these unpaired electrons. In MD simulation, they are usually estimated by measuring the distance between oxygen atoms (referred to as ON atoms) of two MTSSL labeled residues.[56] It is established in the literature that the ON–ON distance distributions are different from those of the backbone and side chain carbon atoms as the MTSSL probe introduces five flexible torsion angles and a 7 Å distance between the backbone or side chain.[57] Based on the preferred positioning of the probes, these distances can be smaller or larger compared to backbone and sidechain distances of the original residue. To observe these differences in long timescale unbiased probe simulations, we compared ON–ON atom distance distributions with the C_α and sidechain atom distances. Fig. S17 (ESI†) shows that ON–ON atom distance distributions are upward shifted with a median value of 5.22 nm in our system. We also compared closest heavy carbon distances, which are majorly used for contact prediction between protein residues, with the probe distances.[58] The median value of the closest heavy atom based distance distribution is $\sim$0.9 nm lower than the ON–ON atom distance median.

Comparing these distributions to the experimental DEER distribution, we observed single peaks from MD simulations whereas two peaks were observed in the experimental distribution, black lines in Fig. 3. Fig. 3D shows that the ON–ON atom distance distribution points to conformations captured corresponding to the second peak with a larger distance value. This can be explained using the two possibilities described by the reviewer that either the MD does not pick up conformations in the missing peaks or that the experimental distance distribution yields peaks that are artifacts of the experiments. We further compared the experimental results using the MtsslWizard software package, which creates all possible rotamer libraries of MTSSL probes to generate the distance distribution.[57] This approach makes sure that all the possible rotamers of MTSSL probes are sampled. Applying the MtsslWizard software package on simulations performed on micelles and bilayers again shows singular peaks which does not match the multiple peaks observed in

experiments Fig. S27 (ESI†). Therefore, further studies are needed to explain these differences.

Restrained-ensemble MD simulations sample spin probe dynamics, but alter protein dynamics

Our results above elucidate that MTSSL probes modulate the distance distributions obtained from DEER experiments and the experimentally characterized distance distributions are a function of both the protein's dynamics as well as the probe's dynamics. MTSSL spin labels are long and flexible molecules and their dynamics have not been examined previously over a long time. We believe that our previous simulations are not sufficient to capture the dynamics of the probes and the proteins together, making unbiased simulations intractable to explore MTSSL probe dynamics. Restrained-ensemble MD (reMD) simulations have been used previously to restrain MTSSL probes dynamics to the experimentally obtained DEER distributions and we explore this avenue to deconvolute the effect of MTSSL probe's dynamics from the experimental distributional data.

For our reMD simulations we first restrained residue pair 174-466. Since this residue pair is on the extracellular side of the $PepT_{So}$ protein, we chose another pair, 201-364, which had the highest KL divergence on the intracellular side. Hence, our next set of reMD simulations involved two restrained pairs one on each side of the protein. We dubbed these sets of simulations as reMD (1 dist) and reMD (2 dist). While in Fig. 4A and B and Fig. S18A and B (ESI†), the distance distributions between the ON–ON atoms of the MTSSL probes show a match with the experimental distribution, those between the closest-heavy atom distances do not. In Fig. 4A, the residue pair 174-466 distribution in the teal violin plot has a single dominant peak with a median value of 3.22 nm, whereas the experimental distribution has two peaks. Moreover, the same peak as seen in unbiased BDDM micelle simulation distributions shown in the yellow violin plot is 4 nm. For comparison, this value is 3.98 nm for our unbiased simulations with a labeled residue pair.

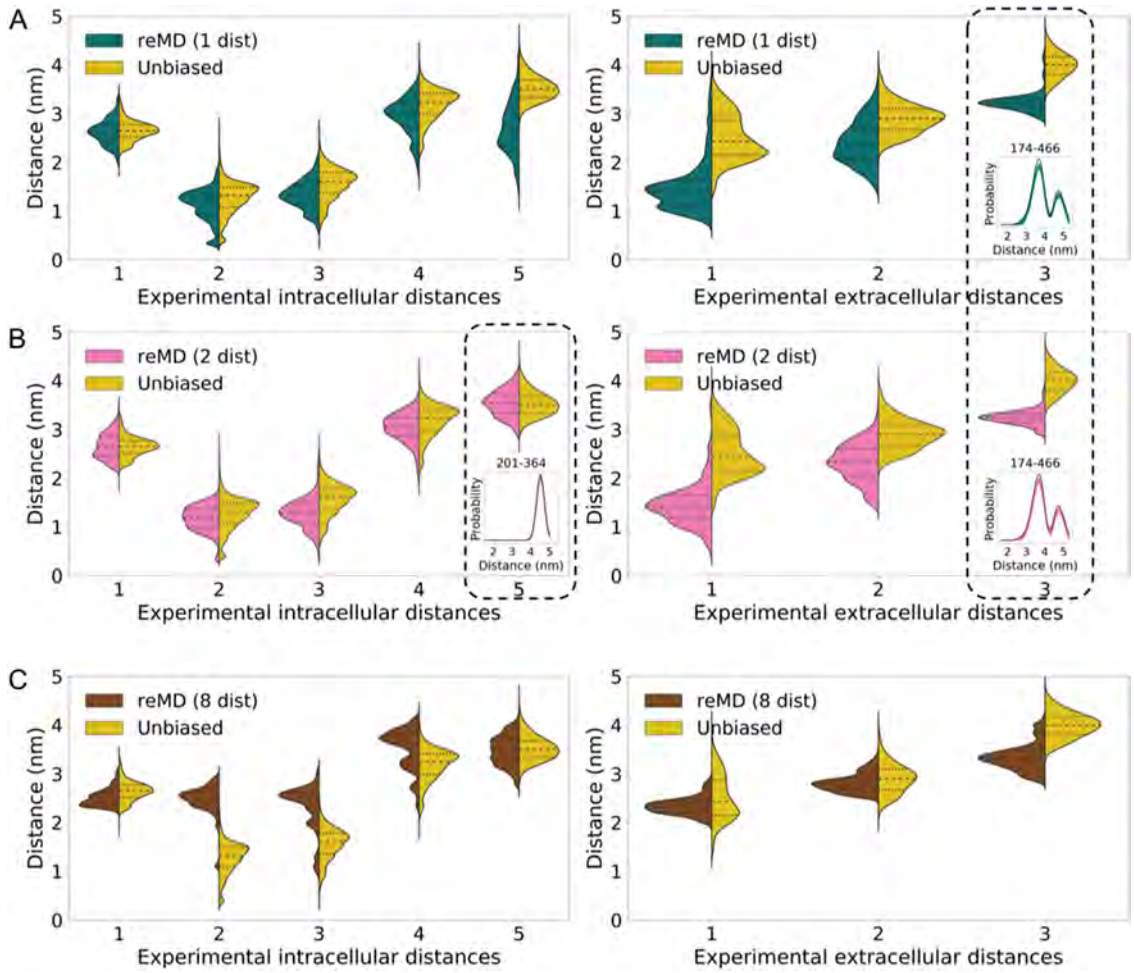

Fig. 4 (A) Violin plot shows distance distributions for five intracellular residue-pair distances and three extracellular residue-pair distances as observed from (A) reMD (1 dist) simulations where residue pair 174-466 is restrained, teal violin plots, (B) reMD (2 dist) where residue pairs 174-466 and 201-364 are restrained, pink violin plots, and (C) reMD (8 dist) where all eight residue pairs are restrained, brown violin plots. Yellow violin plots correspond to unbiased simulations of $PepT_{So}$ protein in the micelle. Black dotted outlined residues pairs in (A) and (B) are restrained pairs and probe distances are shown to match with experimental DEER distance distributions.

In general all three extracellular distances shown in Fig. 4A and B are lower shifted in reMD simulations which can be explained using a labeled residue pair on the extracellular side. While the distributions are lower shifted in reMD for the extracellular distances, neither a lower shift nor an upward shift is seen in the five intracellular distances as shown in Fig. 4A or B. Comparing the residue pair 201-364 in Fig. 4A and B, we note that when this residue pair is not restrained (teal violin plot) its mean value is 2.66 nm and when it is restrained this value is 3.54 nm, very close to the unbiased simulation value of 3.51 nm.

What happens when we restrain all eight residue pairs in system reMD (8 dist)? Three out of five distances – distances #2, 3, and 4 – on the intracellular side show an upward shift, and the median value of the brown violin plots is higher than the median value of the yellow violin plot distributions. Distances #1 and 2 on the extracellular side are also shifted upward as compared to those in systems reMD (1 dist) and reMD (2 dist), although their median values are still lower than those in unbiased simulations.

An upward shift in distance distributions is similar to what we observe in Fig. 3 where the ON–ON atom based distances shift the distribution upwards by $\sim$0.9 nm. However, the origins of these shifts are different. In particular, considering the distance distribution for residue pair 174-466 which is the third distance on the extracellular side, a lower shift in all reMD simulations compared to unbiased simulations without probes (yellow violin plots in Fig. 4) and with probes (red violin plots in Fig. S18, ESI†) indicates that reMD simulations alter the backbone dynamics in a way MTSSL probe labeled simulations did not. Vast differences in backbone dihedral angles of the relevant residues in reMD simulations support this observation (Fig. S19, ESI†). These drastic shifts in distance distributions are mirrored in the underlying conformational landscapes (Fig. S20, ESI†). Hence, the bias introduced in reMD simulations *via* additional energetic terms for force calculations affects the protein structure differently as compared to the modulation caused when MTSSL probes are attached to residues but simulated with unbiased MD simulations.

Similar to our MTSSL-labeled simulations, reMD simulations also suggest that the DEER experiment distance distributions are a convolution of both the spin probe distances as well as the inherent protein dynamics based distances. The impact of spin labels is not straightforward and unbiased simulations are ill-posed to capture their effect completely. Simulations would need to explore the conformational space of each spin label corresponding to every conformational state of the protein. This increases the computational time necessary to capture spin-label dynamics on a protein. With limited computational resources, it is not feasible to perform long-timescale residue pairwise simulations with MTSSL probes. At the same time, reMD simulations are computationally tractable, but do not solve the problem of an unbiased match with MD simulations from long-timescale MD simulations. reMD simulations with multiple restrained residue pairs also raise the unexplored concern that what number of restraints in reMD simulations would be adequate to capture an MD ensemble where all residue-pair distance distributions can correspond to their DEER experiment observables without perturbing the protein's conformational dynamics.

Conclusion

This work highlights the necessity for careful interpretation of DEER spectroscopy and MD simulations in membrane protein biophysics. The scarcity of membrane protein biophysical characterization necessitates that we salvage all information available from laboratory experiments and computational simulations. Hence DEER spectroscopy and MD simulations will continue to be important techniques in progressing our understanding of protein dynamics. It is, therefore, imperative to understand how to best compare the data obtained from both techniques, not only to show validation of MD simulations with experiments but also to avoid misleading conclusions and to draw predictive conclusions. Previous work had proposed optimization protocols to choose the ideal residue pairs for DEER experiments from already performed MD simulations.[59,60] These protocols can also be used iteratively, by performing simulations followed by experiments and then more simulations to update our understanding of a protein's conformational changes.[61] Such methods can be used to their full potential once we can decipher the structural characterization of different protein modes identified *via* multiple peaks in DEER distance distributions. Hence, in this work, we performed a comprehensive study of potential reasons for the discrepancy between the DEER experiment distributional data and residue pair distributions from atomistic MD simulations.

We show that the major reason for the difference between experiments and simulation distributions is the long length of the MTSSL label and its slow dynamics. The slow dynamics of the flexible MTSSL probes could not be captured in unbiased MD simulations and we examined this using biased simulation methods. While reMD simulations could reconcile experiments and simulations for the restrained residue pairs, reMD yielded significant changes in the protein's conformational dynamics at the residue-level and globally. It is also not feasible for researchers to perform DEER experiments on all residue pairs of a protein which can be followed by multiple residue pair biased reMD simulations. On the other hand, unbiased MD simulations do not cause any unphysical perturbations in the protein. However, it is computationally expensive to perform long-timescale MD simulations with MTSSL probes. We surmise that when using methods such as *OptimalProbes*[59,61] it would be sufficient to perform MD simulations with MTSSL probes for the top predicted choices for DEER experiments.

Our computational study of a single pair of spin labels is limited and a thorough examination of the effect of multiple spin labels and different species of spin labels on cytoplasmic and membrane proteins is necessary. Our work is also limited in examining the effect of lateral pressure in a bilayer vs micelle environment which deserves examination with different micellar sizes. We have also assumed a POPE/POPG membrane

composition and symmetry among upper and lower leaflets of the phospholipid bilayer. Our MD simulations are performed at room temperature. Spicher *et al.* proposed performing MD simulations at solvent (mixtures of glycerol and water) freezing temperatures to accurately compare with conformational ensembles explored in DEER experiments.[55] This is a potential cause for disagreement and is yet to be examined with long-timescale simulations. The simulation temperature will also impact the packing and phase behavior of membranes. Alternatively, using more rigid probe molecules such as metal cation-based probes[36] or a cross-linked side chain of the nitroxide label with pairs of cysteine residues[62] and employing biophysical experimental methods that do not require any changes in the covalent structure of the target protein that affect the protein's dynamics and sometimes function could be explored.

Author contributions

S. M. and D. S. designed the research. S. M. and S. D. performed the research and analyzed data. S.M. wrote the manuscript with inputs from D. S.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

D. S. acknowledges support through the New Innovator Award from the Foundation for Food and Agriculture Research and National Science Foundation CAREER award (MCB-1845606). S. M. acknowledges support from the University of Illinois Graduate College – Dissertation Completion Fellowship. D. S. and S. M. thank the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and *ACI*-1238993) and the State of Illinois.

References

1 O. Schiemann, C. A. Heubach, D. Abdullin, K. Ackermann, M. Azarkh, E. G. Bagryanskaya, M. Drescher, B. Endeward, J. H. Freed, L. Galazzo, D. Goldfarb, T. Hett, L. E. Hofer, L. F. Ibáñez, E. J. Hustedt, S. Kucher, I. Kuprov, J. E. Lovett, A. Meyer, S. Ruthstein, S. Saxena, S. Stoll, C. R. Timmel, M. D. Valentin, H. S. Mchaourab, T. F. Prisner, B. E. Bode, E. Bordignon, M. Bennati and G. Jeschke, *J. Am. Chem. Soc.*, 2021, **143**, 17875–17890.
2 O. Schiemann, *Studying Ribozymes with Electron Paramagnetic Resonance Spectroscopy*, 2021, DOI: 10.1002/9783527814527.ch32.
3 G. Jeschke, *Annu. Rev. Phys. Chem.*, 2012, **63**, 419–446.
4 G. Z. Sowa and P. Z. Qin, *Progress in Nucleic Acid Research and Molecular Biology*, Elsevier, 2008, pp. 147–197.
5 N. Fleck, C. A. Heubach, T. Hett, F. R. Haege, P. P. Bawol, H. Baltruschat and O. Schiemann, *Angew. Chem., Int. Ed.*, 2020, **59**, 9767–9772.
6 L. Galazzo, G. Meier, D. Januliene, K. Parey, D. D. Vecchis, B. Striednig, H. Hilbi, L. V. Schäfer, I. Kuprov, A. Moeller, E. Bordignon and M. A. Seeger, *Sci. Adv.*, 2022, **8**, 41.
7 A. Martorana, G. Bellapadrona, A. Feintuch, E. D. Gregorio, S. Aime and D. Goldfarb, *J. Am. Chem. Soc.*, 2014, **136**, 13458–13465.
8 D. Goldfarb, *Curr. Opin. Struct. Biol.*, 2022, **75**, 102398.
9 B. Joseph, A. Sikora, E. Bordignon, G. Jeschke, D. S. Cafiso and T. F. Prisner, *Angew. Chem., Int. Ed.*, 2015, **54**, 6196–6199.
10 K. Singewald, M. J. Lawless and S. Saxena, *J. Magn. Reson.*, 2019, **299**, 21–27.
11 P. Widder, J. Schuck, D. Summerer and M. Drescher, *Phys. Chem. Chem. Phys.*, 2020, **22**, 4875–4879.
12 P. P. Borbat, H. S. Mchaourab and J. H. Freed, *J. Am. Chem. Soc.*, 2002, **124**, 5304–5314.
13 G. Jeschke and Y. Polyhach, *Phys. Chem. Chem. Phys.*, 2007, **9**, 1895.
14 H. S. Mchaourab, P. R. Steed and K. Kazmier, *Structure*, 2011, **19**, 1549–1561.
15 T. Schmidt, M. A. Wälti, J. L. Baber, E. J. Hustedt and G. M. Clore, *Angew. Chem., Int. Ed.*, 2016, **55**, 15905–15909.
16 P. Koukos and A. Bonvin, *J. Mol. Biol.*, 2020, **432**, 2861–2881.
17 T. Hett, T. Zbik, S. Mukherjee, H. Matsuoka, W. Bönigk, D. Klose, C. Rouillon, N. Brenner, S. Peuker, R. Klement, H.-J. Steinhoff, H. Grubmüller, R. Seifert, O. Schiemann and U. B. Kaupp, *J. Am. Chem. Soc.*, 2021, **143**, 6981–6989.
18 M. F. Peter, C. Gebhardt, R. Mächtel, G. G. M. Muñoz, J. Glaenzer, A. Narducci, G. H. Thomas, T. Cordes and G. Hagelueken, *Nat. Commun.*, 2022, **13**, 4396.
19 S. Dutta, B. Selvam, A. Das and D. Shukla, *J. Biol. Chem.*, 2022, **298**, 101764.
20 B. Selvam, S. Mittal and D. Shukla, *ACS Cent. Sci.*, 2018, **4**, 1146–1154.
21 K. R. DeMarco, S. Bekker and I. Vorobyov, *J. Physiol.*, 2018, **597**, 679–698.
22 S. M. Islam, R. A. Stein, H. S. Mchaourab and B. Roux, *J. Phys. Chem. B*, 2013, **117**, 4740–4754.
23 R. Shen, W. Han, G. Fiorin, S. M. Islam, K. Schulten and B. Roux, *PLoS Comput. Biol.*, 2015, **11**, e1004368.
24 F. Marinelli and J. D. Faraldo-Gómez, *Biophys. J.*, 2015, **108**, 2779–2782.
25 Y. Polyhach, E. Bordignon and G. Jeschke, *Phys. Chem. Chem. Phys.*, 2011, **13**, 2356–2366.
26 L. S. Stelzl, P. W. Fowler, M. S. Sansom and O. Beckstein, *J. Mol. Biol.*, 2014, **426**, 735–751.
27 S. M. Islam and B. Roux, *J. Phys. Chem. B*, 2015, **119**, 3901–3911.
28 X. Zhuang and J. B. Klauda, *Biochim. Biophys. Acta, Biomembr.*, 2016, **1858**, 1541–1552.
29 S. J. Marrink, V. Corradi, P. C. Souza, H. I. Ingólfsson, D. P. Tieleman and M. S. Sansom, *Chem. Rev.*, 2019, **119**, 6184–6226.
30 C. Martens, R. A. Stein, M. Masureel, A. Roth, S. Mishra, R. Dawaliby, A. Konijnenberg, F. Sobott, C. Govaerts and H. S. Mchaourab, *Nat. Struct. Mol. Biol.*, 2016, **23**, 744–751.
31 I. D. Sahu, R. M. McCarrick, K. R. Troxel, R. Zhang, H. J. Smith, M. M. Dunagan, M. S. Swartz, P. V. Rajan,

B. M. Kroncke, C. R. Sanders and G. A. Lorigan, *Biochemistry*, 2013, **52**, 6627–6632.
32 R. Ward, C. Pliotas, E. Branigan, C. Hacker, A. Rasmussen, G. Hagelueken, I. R. Booth, S. Miller, J. Lucocq, J. H. Naismith and O. Schiemann, *Biophys. J.*, 2014, **106**, 834–842.
33 M. Teucher, H. Zhang, V. Bader, K. F. Winklhofer, A. J. García-Sáez, A. Rajca, S. Bleicken and E. Bordignon, *Sci. Rep.*, 2019, **9**, 13013.
34 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.
35 A. Potapov, H. Yagi, T. Huber, S. Jergic, N. E. Dixon, G. Otting and D. Goldfarb, *J. Am. Chem. Soc.*, 2010, **132**, 9040–9048.
36 X. Bogetti, S. Ghosh, A. G. Jarvi, J. Wang and S. Saxena, *J. Phys. Chem. B*, 2020, **124**, 2788–2797.
37 D. Banerjee, H. Yagi, T. Huber, G. Otting and D. Goldfarb, *J. Phys. Chem. Lett.*, 2012, **3**, 157–160.
38 E. Matalon, T. Huber, G. Hagelueken, B. Graham, V. Frydman, A. Feintuch, G. Otting and D. Goldfarb, *Angew. Chem., Int. Ed.*, 2013, **52**, 11831–11834.
39 D. P. Claxton, M. Quick, L. Shi, F. D. de Carvalho, H. Weinstein, J. A. Javitch and H. S. Mchaourab, *Nat. Struct. Mol. Biol.*, 2010, **17**, 822–829.
40 K. Kazmier, S. Sharma, M. Quick, S. M. Islam, B. Roux, H. Weinstein, J. A. Javitch and H. S. Mchaourab, *Nat. Struct. Mol. Biol.*, 2014, **21**, 472.
41 S. Newstead, D. Drew, A. D. Cameron, V. L. G. Postis, X. Xia, P. W. Fowler, J. C. Ingram, E. P. Carpenter, M. S. P. Sansom, M. J. McPherson, S. A. Baldwin and S. Iwata, *EMBO J.*, 2010, **30**, 417–426.
42 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, *Biophys. J.*, 2015, **109**, 1528–1532.
43 E. M. Quistgaard, C. Löw, F. Guettou and P. Nordlund, *Nat. Rev. Mol. Cell Biol.*, 2016, **17**, 123–132.
44 H. Krishnamurthy and E. Gouaux, *Nature*, 2012, **481**, 469–474.
45 M. C. Chan, B. Selvam, H. J. Young, E. Procko and D. Shukla, *Biophys. J.*, 2022, **121**, 715–730.
46 A. T. Weigle, M. Carr and D. Shukla, *J. Chem. Theory Comput.*, 2021, **17**, 5342–5357.
47 J. A. Coleman, D. Yang, Z. Zhao, P.-C. Wen, C. Yoshioka, E. Tajkhorshid and E. Gouaux, *Nature*, 2019, **569**, 141–145.
48 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, *Nat. Commun.*, 2020, **11**, 1005.
49 B. Selvam, Y.-C. Yu, L.-Q. Chen and D. Shukla, *ACS Cent. Sci.*, 2019, **5**, 1085–1096.
50 K. J. Cheng, B. Selvam, L.-Q. Chen and D. Shukla, *J. Phys. Chem. B*, 2019, **123**, 8411–8418.
51 D. S. Terry, R. A. Kolster, M. Quick, M. V. LeVine, G. Khelashvili, Z. Zhou, H. Weinstein, J. A. Javitch and S. C. Blanchard, *Nat. Commun.*, 2018, **9**, 230.
52 K. Sale, L. Song, Y.-S. Liu, E. Perozo and P. Fajer, *J. Am. Chem. Soc.*, 2005, **127**, 9334–9335.
53 H. Yu, Y. Mu, L. Nordenskiöld and G. Stock, *J. Chem. Theory Comput.*, 2008, **4**, 1781–1787.
54 J. L. Sarver, J. E. Townsend, G. Rajapakse, L. Jen-Jacobson and S. Saxena, *J. Phys. Chem. B*, 2012, **116**, 4024–4033.
55 S. Spicher, D. Abdullin, S. Grimme and O. Schiemann, *Phys. Chem. Chem. Phys.*, 2020, **22**, 24282–24290.
56 G. Tesei, J. M. Martins, M. B. A. Kunze, Y. Wang, R. Crehuet and K. Lindorff-Larsen, *PLoS Comput. Biol.*, 2021, **17**, e1008551.
57 G. Hagelueken, R. Ward, J. H. Naismith and O. Schiemann, *Appl. Magn. Reson.*, 2012, **42**, 377–391.
58 Z. Guo, J. Liu, J. Skolnick and J. Cheng, *Nat. Commun.*, 2022, **13**, 6963.
59 S. Mittal and D. Shukla, *J. Phys. Chem. B*, 2017, **121**, 9761–9770.
60 J. M. Hays, M. K. Kieber, J. Z. Li, J. I. Han, L. Columbus and P. M. Kasson, *Angew. Chem., Int. Ed.*, 2018, **57**, 17110–17114.
61 S. Mittal and D. Shukla, *J. Phys. Chem. B*, 2018, **122**, 10793–10805.
62 M. R. Fleissner, M. D. Bridges, E. K. Brooks, D. Cascio, T. Kalai, K. Hideg and W. L. Hubbell, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 16241–16246.

