

# *The importance of the membrane interface as the reference state for membrane protein stability*

Jakob P. Ulmschneider<sup>1,\*</sup>, Jeremy C. Smith<sup>2</sup>, Stephen H. White<sup>3</sup> & Martin B. Ulmschneider<sup>4,\*</sup>

<sup>1</sup>Department of Physics and the Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai, China

<sup>2</sup>Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge TN, USA, and Department of Biochemistry & Cellular Molecular Biology, University of Tennessee, Knoxville, TN, USA

<sup>3</sup>Department of Physiology & Biophysics, University of California at Irvine, Irvine CA, USA

<sup>4</sup>Department of Chemistry, King's College London, UK

\*Correspondence to: [jakob@sjtu.edu.cn](mailto:jakob@sjtu.edu.cn); [martin@ulmschneider.com](mailto:martin@ulmschneider.com)

## **Abstract**

The insertion of nascent polypeptide chains into lipid bilayer membranes and the stability of membrane proteins crucially depend on the equilibrium partitioning of polypeptides. For this, the transfer of full sequences of amino-acid residues into the bilayer, rather than individual amino acids, must be understood. Earlier studies have revealed that the most likely reference state for partitioning very hydrophobic sequences is the membrane interface. We have used  $\mu$ s-scale simulations to calculate the interface-to-transmembrane partitioning free energies  $\Delta G_{S \rightarrow TM}$  for two hydrophobic carrier sequences in order to estimate the insertion free energy for all 20 amino acid residues when bonded to the center of a partitioning hydrophobic peptide. Our results show that prior single-residue scales likely overestimate the partitioning free energies of polypeptides. The correlation of  $\Delta G_{S \rightarrow TM}$  with experimental full-peptide translocon insertion data is high, suggesting an important role for the membrane interface in translocon-based insertion. The choice of carrier sequence greatly modulates the contribution of each single-residue mutation to the overall partitioning free energy. Our results demonstrate the importance of quantifying the actually observed full-peptide partitioning equilibrium, which is between membrane interface and transmembrane inserted, rather than combining individual water-to-membrane amino acid transfer free energies.

## **Author summary**

The accurate determination of membrane embedded segments in genome sequences, and prediction of membrane protein structure from amino-acid sequence depend crucially on the precise quantification of polypeptide partitioning into lipid bilayers. Earlier studies have revealed that most likely reference state for partitioning very hydrophobic sequences is the membrane interface, i.e. the relevant equilibrium is that of interface-to-transmembrane partitioning. This paper describes the first full-peptide interface-to-transmembrane partitioning free energy scale for hydrophobic carrier sequences for all 20 amino acid residues. The correlation with experimental full-peptide translocon insertion data is high, suggesting an important role for the membrane interface in translocon-based insertion.

**Keywords:** Peptide partitioning, transfer free energy, translocon, membrane, molecular dynamics.

## Introduction

The transfer free energy of hydrophobic polypeptide segments into the lipid bilayer provides the thermodynamic stabilization of membrane proteins. The accurate determination of membrane embedded segments in genome sequences, and prediction of membrane protein (MP) structure from amino-acid sequence depend crucially on the precise quantification of polypeptide partitioning into lipid bilayers. Although the hydrophobic effect is the primary driver of membrane partitioning, peptide structure formation depends upon non-trivial atomic-detail interactions of peptides with membrane lipids that give rise to a large variety of complex membrane phenomena, ranging from antimicrobial peptide activity to the formation of membrane protein 3D structure.

For  $\alpha$ -helical membrane proteins, the determination of the insertion energetics of transmembrane (TM) helical segments into membranes has proven difficult. The chief experimental challenge is to overcome the tendency of non-polar helices to aggregate and precipitate out of aqueous solution [1, 2]. Nonetheless, numerous experimental and computational approaches to quantifying transfer free energies have been presented over the last decades [3]. Historically, the general approach has been to sum up water-to-membrane transfer free energies of individual amino-acid side chains, although it is clear that the peptide backbone also plays a crucial role in setting the threshold free energy for insertion [4]. A recent set of measurements has used the folding and refolding capability of the bacterial outer membrane phospholipase A (OmpLA) as a scaffold to determine the relative transfer free energies of amino acid residues into lipid bilayers from water (Figure 1b) [5]. However, unlike outer-membrane proteins, that fold directly into the membrane from an aqueous phase, multi-span  $\alpha$ -helical membrane proteins fold into the membrane bilayer aided in an incompletely characterized manner by the SecY translocase in bacteria or the structurally similar Sec61 translocase in eukaryotes [6]. To begin to understand translocon-guided insertion of TM segments, an *in vitro* assay has been used to decipher the ‘code’ that the microsomal Sec61 translocon uses to decide if nascent chain segments are inserted into the bilayer or secreted (the “biological hydrophobicity scale”, Figure 1a) [7, 8]. Although the broad energetics are similar, these two scales have proved difficult to reconcile quantitatively with other single-residue hydrophobicity scales. The key difficulty lies in interpreting the partitioning properties in the absence of detailed structural knowledge of the proteins or peptides studied and their dynamic interactions with the lipid bilayer environment. In particular, the translocon assay utilizes complex cellular systems, the mechanics and thermodynamic processes of which are at present only poorly understood [9]. For the OmpLA assay, a key issue is again the lack of the protein backbone contribution. Furthermore, and importantly, the non-inserted reference state is not known for either experiment.

In order to approach this question computationally, our approach has been to use molecular dynamics (MD) simulations to obtain directly membrane partitioning free energies for entire peptides, and report whether they are mainly TM inserted or surface bound. Importantly, these simulations reveal a very strong preference of hydrophobic peptides for the membrane interface, consistent with what is known about folding of amphipathic helices into lipid bilayers [10]. These simulations go significantly beyond earlier studies on side-chain analogs [11-13], or bulk phase transfer free energies [14, 15], in that they provide free energies

for whole sequences. This includes free energy contributions from the polar peptide backbone, peptide-induced adaptation and deformations of the lipid bilayer, and other macroscopic effects (e.g. helix dipoles, folding-partitioning coupling) not accounted for in analog studies.

In principle, full-peptide insertion values could be obtained via potential of mean force (PMF) partitioning calculations, as previously shown for both atomistic [16, 17] and coarse grained representations [18, 19]. However, we and others have shown that these simulations involve enormous convergence challenges to be reliable [20, 21], which are related to the strong bilayer perturbation caused by the transition between surface and transmembrane state. Instead, our recent simulations show that determining the partitioning of full sequences into bilayers is simpler than previously anticipated (Figure 1c) [21-28]. Hydrophobic peptides generally partition into membranes as either surface bound (S) or membrane spanning helices (TM) [3]. Water-solvated states (W) are much higher in free energy and not measurably populated at equilibrium, consistent with the experimentally known general insolubility of hydrophobic peptides in water [1, 2]. For more polar or charged peptides the equilibrium is shifted towards water soluble and interfacial states, and includes a folding component, which is generally absent for TM segments, that partition as helices, and are helical both in S-state and TM [23]. In this study, we investigate 9-residue hydrophobic carrier sequences that can monomerically insert as TM helices, but are short enough to also be observed in the S state (Figure 1c). The key difference to other computational scales is that these transfer free energies thus represent the true equilibrium behavior of a full peptide, which, as was previously shown, partitions between the S and TM state [23-27, 29, 30]. We have recently demonstrated that the direct partitioning simulations employed here are robust and yield results identical to long-scale PMF calculations [21].

## Results

### Equilibrium partitioning process

We used two carrier sequences, ac-LLLLXLLLL-nme (L8X) and ac-ALALXLALA-nme (L4A4X), with X corresponding to one of the 20 naturally-occurring amino acids. The length of these sequences ( $n = 9$ ) results in a computed range of  $|\Delta G_{S \rightarrow TM}| < 5$  kcal/mol, the optimum sensitivity for calculating free energies directly via partitioning simulations. Figures 2 and 3 show typical results from partitioning simulations for two representative peptides, L8Q and L8W. Only 3 states are populated at equilibrium: S and S' on the bilayer interface (center of mass insertion depth  $z_{CM} \approx \pm 12$  Å, tilt angle  $\theta \approx 90^\circ$ , where  $z_{CM} = 0$  is the center of the bilayer), and TM inserted ( $z_{CM} \approx 0$  Å,  $\theta \approx 10-30^\circ$ ). The TM state is mostly aligned with the bilayer normal due to the negative mismatch of short sequences that prevents a larger tilt angle. There are frequent transitions between the three states, indicating equilibrium has been obtained (Figure 2b,3b). The insertion free energy profiles ( $\Delta G$ ) were calculated from the peptide location via  $\Delta G = -RT \ln p(z)$  ( $R =$  gas constant,  $T =$  temperature,  $p(z) =$  probability distribution of peptide center of mass along the membrane normal). The above-described pattern is shared by all 20 sequences, but the equilibrium varies greatly with the choice of central residue. For example, the polar L8Q peptide populates almost equally the interfaces and the TM inserted state (Figure 2), while the much more hydrophobic L8W strongly prefers the TM state (Figure 3). Thus, a single-residue substitution leads to large change in equilibrium partitioning for these peptides.

Of particular interest are the four charged amino acids. The two cationic sequences L8K and L8R strongly prefer one S state, but they nevertheless were found to insert briefly into TM orientations. Such spontaneous insertion events are concomitant with a significant water defect and localized deformation of the bilayer [11, 31]. Snorkeling of the ammonium (lysine) and guanidinium (arginine) sidechain moieties pulls both water and lipid molecules deep into the bilayer. At the temperature range and timescales presented here L8K and L8R do not translocate across the bilayer, and cannot populate the S-state of the opposite bilayer leaflet. The negatively charged L8D and L8E behave differently: both remain in the S-state throughout the 1.5  $\mu$ s simulations, never inserting; the energetic cost to bury the negatively charged carboxylate group into the hydrophobic membrane core is clearly too high. pKa simulations on side chain analogs suggest that these two anionic residues are protonated if sufficiently inserted into the bilayer [11]. We therefore also simulated L8D and L8E in their neutral (protonated) forms, resulting in frequent TM insertion. Insertion of charged Glu and Asp residues can be achieved by increasing the hydrophobicity of the carrier sequence, e.g. by adding a flanking Leu residue on each side (L10D, L10E). These two sequences exhibited several short insertion events during the 1.5  $\mu$ s simulations but still remained overwhelmingly in the S-state, with estimates of  $\Delta G_{S \rightarrow TM} = 4.7 \pm 1.8$  kcal/mol for L10D and  $\Delta G_{S \rightarrow TM} = 6.2 \pm 2.1$  kcal/mol for L10E. It is thus apparent that the anionic residues (Glu, Asp) behave very differently from the cationic ones (Lys, Arg), and carry much higher insertion penalties.

Figure 4 shows the full-peptide surface-to-bilayer insertion scales of all 20 natural amino acids for the two carrier sequences, L8X and L4A4X. The two scales are highly correlated, with  $\Delta G_{L8X} = 1.2 \times \Delta G_{L4A4X} - 1.9$ , and  $R^2 = 0.94$ . Thus, each A $\rightarrow$ L residue substitution contributes on average  $\Delta \Delta G_{A \rightarrow L} = -0.5$  kcal/mol, similar to  $\Delta \Delta G_{app} = -0.7$  kcal/mol found in the translocon experiment [7]. Despite  $\mu$ s sampling times, many hydrophobic residues are roughly in the same range (e.g.  $\sim -1$  kcal/mol for L4A4X for A,I,V,C,M,L,F,G), indicating that the simulations are only sensitive up to  $\sim 0.3 - 0.5$  kcal/mol, which can also be seen by Alanine being predicted slightly more hydrophobic than Valine. These issues are clearly due to lack of convergence. Longer simulations are needed to determine the exact ranking. Thus, for these hydrophobic residues, some small errors in the prediction still occur and the ranking for hydrophobic residues should not be seen as absolute.

## Comparison with experimental scales

### *Side chain analog scales.*

To compare our values with existing scales, we calculated what the partitioning of the peptides would be using the available experimental and simulation data. First we consider side chain-only partitioning scales (Figure 5). In comparing full-peptide scales with single-residue scales, only the relative free energies ( $\Delta \Delta G$ ) of sidechain contributions can be analysed. The zero-point reference value is chosen to be alanine ( $\Delta \Delta G_{Ala} = 0$ ). This means we are comparing the side-chain only contribution, relative to alanine, of the various experimental partitioning values with corresponding values computed for a full-peptide S $\rightarrow$ TM process.

Water-to-cyclohexane partitioning free energies  $\Delta G_{CHX}$  of amino acid sidechain analogs have been obtained by Radzicka and Wolfenden [14, 32] (Figure 5a). Cyclohexane serves as a mimic of the hydrophobic membrane core, which is believed to have similar physicochemical

properties to bulk hydrocarbons. Indeed,  $\Delta\Delta G_{S \rightarrow TM}$  is highly correlated with  $\Delta\Delta G_{CHX}$  ( $R^2 = 0.83$  for both L8X and L4A4X). However, the computational scale is compressed by a factor of  $\sim 2.5$ . The same correlation and compression was seen relative to the potential of mean force (PMF) calculations on the transfer of side chain analogs from water to the center of a DOPC bilayer by MacCallum et al. [11, 33] (Figure 5a,  $\Delta\Delta G_{Analog}$ ,  $R^2 = 0.80$  for L8X, L4A4X). Comparison with interface-to-bilayer transfer ( $S \rightarrow bilayer$ ) of sidechain analogs from the same work shows a similar compression ( $\sim 2.3$ ) (data not shown). This comparison highlights three major limitations of sidechain analog scales: First, and evidently, the polar backbone contributions are missing. Second, peptides are much larger than small, side-chain mimicking molecules. An amino-acid sidechain attached to a peptide is therefore not ‘solvated’ in the membrane core as it would be in a hydrophobic bulk solvent. Third, consideration of bulk water as a reference state for transfer into a membrane is impractical for greasy peptides that have negligible populations solvated in aqueous solution, and populate instead the interface. Thus, good agreement was neither expected nor found, since water-to-bilayer transfer is qualitatively different than the surface-to-bilayer partitioning observed in our simulations.

**Full-peptide scales.** Important improvements over side-chain analog scales have been full-peptide measurements of the energetic cost of partitioning H-bonded peptides (Figure 5b). Wimley and White measured the partitioning of pentapeptides of the form Ace-WLXLL ( $X =$  all 20 amino acids) between water and *n*-octanol [15], and between water and the interface of POPC vesicles [34]. Water-saturated *n*-octanol, which is of a micellar structure, serves as a mimic of a heterogeneous, partly hydrophilic and hydrophobic environment such as the interior of a protein, or a lipid bilayer. The designed peptides are too short to form secondary structure or to insert into a TM orientation, and thus reside only on the interface of the vesicles. It is therefore no surprise that the correlation of the simulation data with the interfacial scale is poor ( $R^2 = 0.37$ ). The correlation with the octanol ( $\Delta\Delta G_{Octanol}$ ) hydrophobicity scale (Figure 5) is equally low ( $R^2 = 0.45$ ), as is the correlation with the difference between the two scales. Clearly, both unfolded water-to-surface and water-to-octanol transfers are very different than  $\alpha$ -helical surface-to-TM partitioning.

A significantly better fit, of  $R^2 = 0.9$ , is obtained when comparing to statistical potentials derived from an analysis of transmembrane amino-acid distributions of membrane proteins of known structure (Figure 5b,  $\Delta\Delta G_{PMF}$ ) [35]. Here, the statistical scale is compressed by a factor of  $\sim 2.2$  compared to the simulations. The origin of this scale factor – which is also seen in comparison to the translocon experiments – is discussed below.

**OmpLA sidechain scale.** Recently, a series of fluorescence spectroscopy experiments were performed by Moon and Fleming that determined the relative water-to-bilayer transfer free energies of the 20 natural amino acids by using the outer membrane phospholipase A (OmpLA) as a transmembrane scaffold [5]. OmpLA can fold reversibly into synthetic liposomes *in vitro*. Host-guest experiments using OmpLA yield relative water-to-bilayer transfer free energies ( $\Delta\Delta G_{OmpLA}$ ) of amino acid side chains on the protein surface. The equilibrium is between a functionally inserted folded beta-barrel (as verified by enzymatic activity assays), and an unfolded state in solution (Figure 1). The resulting experimental free energies are for sidechain contributions only, and are relative to alanine, which was chosen as the zero of the scale

( $\Delta\Delta G_{ALA} = 0$ ). If the computational scale is shifted in a similar way, a correlation of  $R^2 = 0.81$  is achieved, with a slope of 0.8 (Figure 5c). The pH of the experiments was 3.8, which is close to the pKa values for Asp and Glu side chains. The guest Asp and Glu side chains in the system were thus likely protonated, as they are in the simulations. The structure of the unfolded water-soluble state of OmpLA is unfortunately unknown. However, recent experiments demonstrate that it is not a specific state, or an unstructured, random-coil chain, but rather an ensemble of conformationally distinct states [36]. Likely, the mutated residue is partially or fully buried in these coiled unfolded states, and thus not fully exposed to water. The partitioning of the residue would thus then be between a partially hydrophobic state and a fully membrane buried lipid-exposed position. Since our membrane is thicker (POPC) than that used in the experiments (DLPC) additional simulations would be needed for a more valid comparison [37].

**Translocon mediated insertion scale.** Several *in vitro* experiments by Hessa et al. have yielded the code that Sec61 uses for selecting nascent chain segments for insertion into the bilayer [7, 8]. Based on a vast number of experiments, this has resulted in the construction of a ‘biological hydrophobicity scale’ that allows the prediction of whether the translocon will insert a given polypeptide segment into the membrane or translocate it across [8]. One of the great advantages of these experiments is that they are ‘full-sequence’. This has provided insights into the effect of the position of a guest residue along a TM segment and whether the effect of multiple guest residues is additive or cooperative [8], as well as the influence of the overall sequence length on insertion [38]. Figure 6 shows the values for the L8X and L4A4X scales, with the experimental data  $\Delta G_{pred}$  taken from the  $\Delta G$  prediction server (<http://dgpred.cbr.su.se>) [8]. In both cases, the correlation is very good ( $R^2 = 0.91$ ) if the only outlier, proline, which disturbs a straight helix structure, is removed. Including proline reduces  $R^2$  to 0.86.

Despite the good correlation, two key differences are apparent. First, the simulation-derived values are strongly shifted towards TM insertion, by 5-6 kcal/mol. Thus, many sequences insert in the simulations that, according to the biological scale prediction, would be secreted ( $\Delta G_{pred} > 0$ ). This shift has been previously seen for polyleucine peptides of different length, where  $\sim 2.4$  fewer leucines were required for the peptides to insert on their own than predicted by the translocon scale [23, 25]. The second difference is that the computational values are scaled up by a factor of 2.0-2.4 with respect to the translocon data.

## Discussion

The relevant equilibrium governing peptide insertion into lipid bilayers is that of full-peptide partitioning. For this, the burial of an entire sequence, rather than individual amino acids - which do not prefer to be buried in a membrane - should be investigated. The present simulations demonstrate that the partitioning of entire sequences is relatively easy to compute via direct MD simulation. Unfortunately, a comparison to an equivalent experiment is not possible yet, because in all available experimental studies different equilibria were probed.

The total insertion free energy range obtained here for the two hydrophobic carrier sequences is much smaller than prior computational and experimental scales that were based

on amino-acid analogs, because the transfer  $\Delta G_{S \rightarrow TM}$  is not from water to the bilayer core. The computed values are much closer to the experimental translocon biological hydrophobicity scale, which was found to be strongly compressed (by a factor of  $\sim 3-4$ ) compared to the analog scales - e.g. 2.6 kcal/mol for arginine vs. computed values ranging from 10-17 kcal/mol [11, 12, 33, 39-41]. There have been a number of possible explanations of the origin of this compression factor [41-48]. Clearly, different partitioning processes are involved: experimental values of  $\Delta G_{app}$  correspond to the partitioning of peptides between the translocon channel and the bilayer ( $\Delta G_{Sec61 \rightarrow TM}$ ), rather than between water-soluble and TM configurations (Figure 1a) [9, 42, 49-57].

The most relevant comparison to make using the present data is between the translocon scale and  $\Delta G_{S \rightarrow TM}$ , because both are for full sequences. The high degree of correlation, also seen in our prior studies using different carrier sequences [23, 27], could be an indicator of a significant role of the bilayer interface in translocon function, suggesting that the membrane surface is involved at some point (or at all times) during insertion. The ribosome-translocon complex is not apparently shielded from the cytoplasm and inner membrane surface [6], thus it is likely that nascent chains could make first contact with the membrane in the vicinity of the translocon, or even already form surface helices there, all without involving the lateral gate. This recently proposed sliding model suggests that the translocon acts mainly as a catalyst, with the growing nascent chain persistently exploring the states available to it [6]. Since it takes  $\sim 50$  ms to make a peptide bond (i.e the chain is elongated at this rate), there is more than enough time for the nascent sequence to fully sample the surface-to-TM energy landscape in a fashion similar to that in the present MD simulations. In the case of mostly hydrophobic sequences, this would correspond to folding and subsequent surface-to-TM sliding of the chain in the bilayer close to the translocon, rather than involving the translocon interior in a series of discrete steps via opening and closing events of the lateral gate. This view could explain the notable correlation between the simulation data for spontaneous partitioning and the experiments. In addition to the interface, the much more hydrophobic reference state in the microsomal preparations could be due to other proteins (e.g. chaperones).

Of particular interest is the role of the carrier sequence and its length in the comparison between translocon and direct partitioning. In our prior studies, for ac-GL<sub>n</sub>-R-L<sub>n</sub>G-nme ( $n=5-8$ ), an almost perfect match was found [27]. For pure polyleucine ac-L<sub>n</sub>-nme ( $n=5-10$ ), an identical slope, but constant offset of  $\sim 2$  kcal/mol between translocon experimental values and simulation-derived values was found [23]. In the present study, where the central residue **X** rather than the length of the carrier sequence  $n$  was varied for the sequences ac-LLLLXLLLL-nme (L8X) and ac-ALALXLALA-nme (L4A4X), large offsets of  $\sim 6$  kcal/mol and a slope of  $\sim 2.4$  were found.

These differences originate from the length and sequence dependence of  $\Delta G_{pred}$ . To cover the complete range of the 20 amino acids **X**, we had to choose particularly short carrier sequences ( $n = 9$ ) in order to obtain  $\Delta G_{S \rightarrow TM}$  within  $\pm 5$  kcal/mol. While this appears short, it has been shown both *via* the translocon assay as well as by oriented synchrotron radiation circular



dichroism that such short sequences can indeed be inserted TM [38]. Recently, this has also been confirmed for short sequences with more varied residue combinations, and so is not observed only for synthetic leucine-rich peptides [58]. The offset to the MD simulations indicates that in the simulations even shorter sequences insert TM than in the experiments. What could be the reasons for this? Firstly, we don't know what the non-inserted reference state is in the experiments, and we don't know how the peptides get inserted into the membrane via the translocon. In the MD simulations, short sequences induce some adaptation of the surrounding bilayer to reduce the hydrophobic mismatch by bilayer surface deformation. This significantly lowers the energetic penalty for TM insertion for very short peptides. But whether a similar bilayer deformation is possible in vicinity of the translocon is unknown. A third explanation can be found in the flanking sequences (XPXX, where X = G or K) that are part of the experiments. These additional residues are difficult to include in  $\Delta G_{S \rightarrow TM}$  calculations, since adding flanking (unfolded) glycines or prolines results in large kinetic barriers in the computations. While previously we found the effect of GPGG flanks to be minor (in fact due to bilayer deformation, as mentioned above) this has not been systematically investigated, due to the limitations of the simulation approach for unfolded sequences [23]. Since the sequences in the experiments are significantly longer than in the simulations through inclusions of unfolded flanks, there will likely be some surface shift compared to an isolated, capped hydrophobic peptide, that can insert more easily TM because there are no flanking, unfolded glycines and prolines at both ends that resist TM insertion.

The choice of the carrier sequence also leads to a not-fully-understood effect on the slope in the comparisons. For example, for ac-ALALXLALA-nme, there is a slope of 2.4 between  $\Delta G_{S \rightarrow TM}$  and  $\Delta G_{pred}$ . Yet when this short sequence is compared to a much longer carrier sequence, for example the original data from the translocon experiment for a 19 residue sequence ( $\Delta G_{app}$ ), the exact same central residue mutations of **X** result in a slope of close to one [7]. The role of the carrier sequence is complex, as there is a position dependence (burial depth) of the free energy for L $\rightarrow$ A substitutions along the sequence. In addition, sidechain occlusion effects are expected to play a role [15]. To minimize these, the central **X** residue was surrounded by L in both L8**X** and L4A4**X**, resulting in similar occlusion of **X** by its neighbors. Future studies, including a completely 'naked' backbone-only carrier sequence (restrained polyglycine) are needed to answer these questions, but are beyond the scope of the present work.

Unfortunately, due to hydrophobic aggregation, no equivalent direct experimental measurements of peptide insertion currently exist, and even the most recent experimental techniques employ very different systems [59]. However, full-peptide autonomous insertion calculations, such as shown here, are critical for our understanding of antimicrobial peptides, cell-penetrating peptides, toxins, fusion peptides, and other membrane active peptides. All of these insert into the bilayer without translocons, and therefore only the free (unassisted) partitioning free energy is important.

The main advantage of our method is that it can be used to directly and rapidly compute  $\Delta G_{S \rightarrow TM}$  for many sequences. Its only limitation is that both S and TM states must be reasonably

populated in equilibrium, so  $|\Delta G_{S \rightarrow TM}| < 5$  kcal/mol. We chose 9-residue sequences, but the method also works for much longer carrier sequences, if the sequence contains enough hydrophilic residues [27]. Polar backbone contributions (missing in the analog scales) and the effect of the peptide length are fully included. In marked contrast to interfacial scales, where residue contributions can be additive, the present calculations show this is not the case for TM insertion, as previously noted for arginine [60]. The obvious difference is that the residue position has a huge effect for a TM configuration (burial location), but only a small one for a surface aligned peptide, where the burial of the residue is unaltered. The advantage of the direct simulation protocol shown here is that all (non-trivial) sequence correlations are automatically considered. The extension of the scale to more complex sequences, where **X** is in various positions rather than just in the middle of the helix, is currently being performed, and will result in a complete computational full-peptide insertion scale.

## Methods

**Molecular dynamics (MD).** Peptide sequences of the form ac-LLLLXLLLL-nme (L8X) and ac-ALALXLALA-nme (L4A4X) were constructed with **X** corresponding to the 20 naturally occurring residues. The peptides were then embedded into the water phase of a box containing a preformed POPC (palmitoyl-oleoyl-phosphocholine) lipid bilayer. The initial conformation was an ideal  $\alpha$ -helix, placed 10 Å from the bilayer surface. All simulations were performed and analyzed using GROMACS version 4.0 ([www.gromacs.org](http://www.gromacs.org)) [61] and hippo beta ([www.biowerkzeug.com](http://www.biowerkzeug.com)), using OPLS-AA for the protein [62], TIP3P for water [63], and united atom lipid parameters for POPC [64]. Electrostatic interactions were computed using Particle Mesh Ewald (PME), and a cutoff of 10 Å was used for the van der Waals interactions. Bonds involving hydrogen atoms were restrained using LINCS [65]. Simulations were run with a 2 fs integration time-step and neighbor lists were updated every 5 steps. All simulations were performed in the NPT ensemble, with no additional applied surface tension. Water, lipids, and the protein were each coupled separately to a heat bath with time constant  $\tau_T = 0.1$  ps using weak temperature coupling [66]. Atmospheric pressure of 1 bar was maintained using weak semi-isotropic pressure coupling with compressibility  $\kappa_z = \kappa_{xy} = 4.6 \cdot 10^{-5} \text{ bar}^{-1}$  and time constant  $\tau_p = 1$  ps [67]. 1.5  $\mu$ s of MD were run for each peptide sequence.

**Partitioning free energy and robustness of the simulation protocol.** The insertion propensity,  $p_{TM}$  of each peptide was calculated as the probability of the peptide being in the TM state. To distinguish the TM state from the S and S' conformations, a criterion of  $z_{CM} < 8$  Å and  $\theta < 50^\circ$  was found to be optimal. The free energy of S  $\rightarrow$  TM partitioning was then calculated as  $D G_{S \rightarrow TM} = +kT \ln(1/p_{TM} - 1)$ . Sufficient transition events were captured in the 1.5  $\mu$ s simulations by using an elevated temperature of  $T = 150$  °C. We have previously demonstrated using MD that for pure Ln  $\alpha$ -helices, no unfolding is observed under such conditions, and this was also experimentally verified up to 90 °C [22-25]. The principal cause of this thermostability is the high hydrophobicity of the carrier sequence. However, for several L8X and L4A4X sequences where **X** was polar or charged, partially unfolded S states were observed at 150 °C,

whereas the same sequences were entirely helical at 80 °C. Hence, to avoid thermal unfolding biasing the results, secondary structure restraints were introduced for all sequences at this temperature to retain the helix integrity. This is a common procedure for studying peptide partitioning, e.g., for arginine peptide burial in the bilayer [39, 41, 43], or at the Sec61 translocon [55], for Sec61 peptide release [46], and the Sec61 function itself [68]. It is also standard in all coarse-grained peptide-membrane potential of mean force calculations, e.g. for toxin insertion [69], synthetic WALP peptides [19], partitioning of TM segments of ABC proteins and chemoreceptors [18, 70], the voltage sensor S4 helix [71], and to predict integrin dimers [72].

We have performed extensive checks that both using helical restraints and using high temperatures do not bias the results. Very long (5  $\mu$ s) control simulations were performed for the pure carrier sequence (L8), as shown in Figure S1. Figure 7 reveals that  $\Delta G_{S \rightarrow TM}$  is not affected by either the simulation temperature, or by the use of helical restraints. L8 has a relatively low barrier, allowing partitioning to be simulated at 30 °C. These results confirm that high temperatures can be successfully used to obtain values of  $\Delta G_{S \rightarrow TM}$  applicable at physiological temperatures. There is no temperature effect of partitioning, chiefly because of the lack of unfolding in the membrane, as described earlier [23].

An alternative to using helical restraints for L8X with polar and charged X is to use a force field that promotes helicity more than OPLS. We have recently demonstrated that CHARMM promotes helical structures much more than OPLS in long-scale simulation of membrane active peptides [73, 74]. We therefore performed control simulations using the CHARMM force field [75] and C36 lipid parameters [76]. No loss of helical structure occurred even at temperatures as high 180 °C, for any sequence thus tested. For L8, a 5  $\mu$ s long control simulation of L8 at 210 °C showed nearly the same  $\Delta G_{S \rightarrow TM} = -1.15$  kcal/mol as the OPLS simulation (Table S1). However, greatly reduced kinetics of partitioning as well as increased computational cost (all-atom lipids) prevented us from employing CHARMM parameters for this study. We have recently demonstrated however, that results using CHARMM are identical to OPLS, and that the results of the direct partitioning simulations, as shown here, are identical to those obtained via PMF calculations [21].

## Acknowledgements

J.P.U. was supported by a China 1000 Plan's Program for Young Talents (13Z127060001). S.H.W. was supported by NIH grant GM7463. Simulation resources were provided by the Center for High Performance Computing, Shanghai Jiao Tong University and the Maryland Advanced Research Computing Center (MARCC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Competing financial interests

The authors declare no competing financial interests.

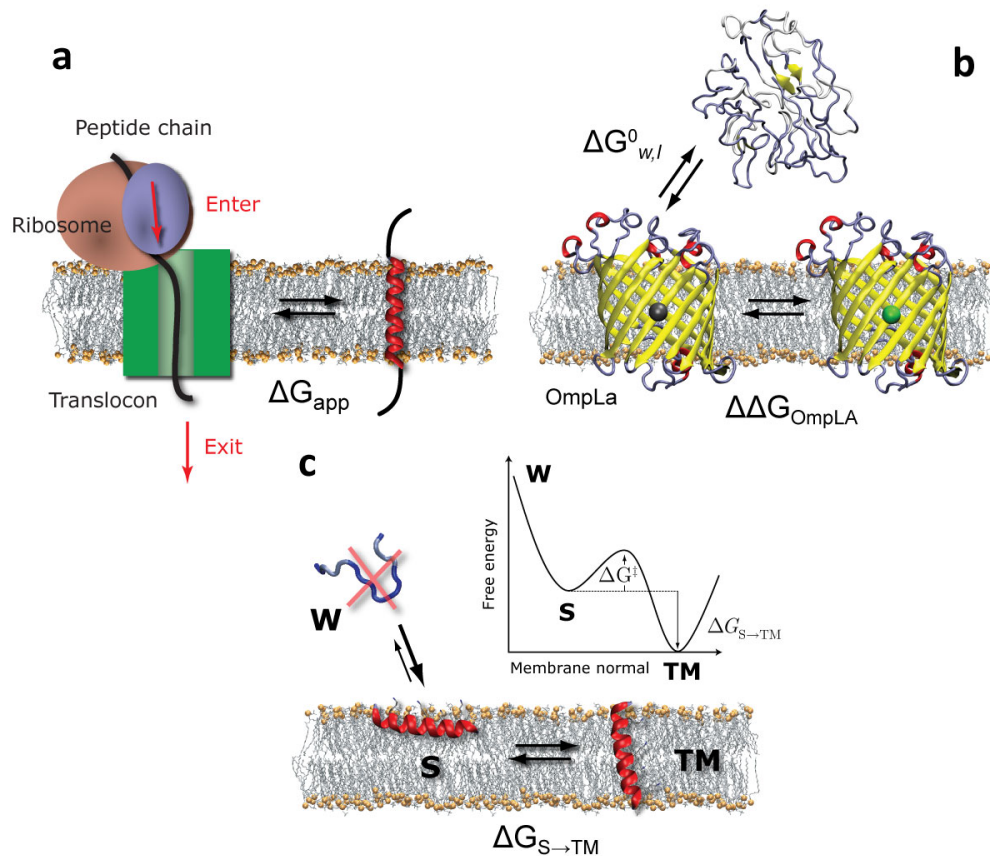
## References

- [1] A.S. Ladokhin, S.H. White, Interfacial folding and membrane insertion of a designed helical peptide, *Biochemistry*, 43 (2004) 5782-5791.
- [2] W.C. Wimley, S.H. White, Designing transmembrane  $\alpha$ -helices that insert spontaneously, *Biochemistry*, 39 (2000) 4432-4442.
- [3] S.H. White, W.C. Wimley, Membrane protein folding and stability: Physical principles, *Annu. Rev. Biophys. Biomolec. Struct.*, 28 (1999) 319-365.
- [4] S. Jayasinghe, K. Hristova, S.H. White, Energetics, stability, and prediction of transmembrane helices, *J. Mol. Biol.*, 312 (2001) 927-934.
- [5] C.P. Moon, K.G. Fleming, Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers, *Proc. Natl. Acad. Sci. USA*, 108 (2011) 10174-10177.
- [6] F. Cymer, G. von Heijne, S.H. White, Mechanisms of Integral Membrane Protein Insertion and Folding, *J. Mol. Biol.*, 427 (2015) 999-1022.
- [7] T. Hessa, H. Kim, K. Bihlmaier, C. Lundin, J. Boekel, H. Andersson, I. Nilsson, S.H. White, G. von Heijne, Recognition of transmembrane helices by the endoplasmic reticulum translocon, *Nature*, 433 (2005) 377-381.
- [8] T. Hessa, N.M. Meindl-Beinker, A. Bernsel, H. Kim, Y. Sato, M. Lerch-Bader, I. Nilsson, S.H. White, G. von Heijne, Molecular code for transmembrane-helix recognition by the Sec61 translocon, *Nature*, 450 (2007) 1026-1030.
- [9] E. Schow, J. Freitas, P. Cheng, A. Bernsel, G. von Heijne, S. White, D. Tobias, Arginine in Membranes: The Connection Between Molecular Dynamics Simulations and Translocon-Mediated Insertion Experiments, *J. Memb. Biol.*, 239 (2011) 35-48.
- [10] P.F. Almeida, A.S. Ladokhin, S.H. White, Hydrogen-bond energetics drive helix formation in membrane interfaces, *Biochim Biophys Acta*, 1818 (2012) 178-182.
- [11] J.L. MacCallum, W.F. Bennett, D.P. Tieleman, Distribution of amino acids in a lipid bilayer from computer simulations, *Biophys. J.*, 94 (2008) 3393-3404.
- [12] A.C. Johansson, E. Lindahl, Position-resolved free energy of solvation for amino acids in lipid membranes from molecular dynamics simulations, *Proteins*, 70 (2008) 1332-1344.
- [13] D. Sengupta, J.C. Smith, G.M. Ullmann, Partitioning of amino-acid analogues in a five-slab membrane model, *Biochim Biophys Acta*, 1778 (2008) 2234-2243.
- [14] A. Radzicka, R. Wolfenden, Comparing the Polarities of the Amino-Acids - Side-Chain Distribution Coefficients between the Vapor-Phase, Cyclohexane, 1-Octanol, and Neutral Aqueous-Solution, *Biochemistry*, 27 (1988) 1664-1670.
- [15] W.C. Wimley, T.P. Creamer, S.H. White, Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides, *Biochemistry*, 35 (1996) 5109-5124.
- [16] A. Babakhani, A.A. Gorfe, J.E. Kim, J.A. McCammon, Thermodynamics of peptide insertion and aggregation in a lipid bilayer, *J. Phys. Chem. B*, 112 (2008) 10528-10534.
- [17] A. Sandoval-Perez, K. Pluhackova, R.A. Bockmann, Critical Comparison of Biomembrane Force Fields: Protein-Lipid Interactions at the Membrane Interface, *J Chem Theory Comput*, 13 (2017) 2310-2321.
- [18] A. Chetwynd, C.L. Wee, B.A. Hall, M.S.P. Sansom, The Energetics of Transmembrane Helix Insertion into a Lipid Bilayer, *Biophys. J.*, 99 (2010) 2534-2540.
- [19] Benjamin A. Hall, Alan P. Chetwynd, Mark S.P. Sansom, Exploring Peptide-Membrane Interactions with Coarse-Grained MD Simulations, *Biophys. J.*, 100 (2011) 1940-1948.
- [20] C. Neale, W.F.D. Bennett, D.P. Tieleman, R. Pomès, Statistical Convergence of Equilibrium Properties in Simulations of Molecular Solutes Embedded in Lipid Bilayers, *J. Chem. Theory Comput.*, 7 (2011) 4175-4188.
- [21] J.C. Gumbart, M.B. Ulmschneider, A. Hazel, S.H. White, J.P. Ulmschneider, Computed Free Energies of Peptide Insertion into Bilayers are Independent of Computational Method, *J Membrane Biol*, (2018).
- [22] J.P. Ulmschneider, J.P.F. Doux, J.A. Killian, J.C. Smith, M.B. Ulmschneider, Peptide partitioning and folding into lipid bilayers, *J. Chem. Theory Comput.*, 5 (2009) 2202-2205.
- [23] J.P. Ulmschneider, J.C. Smith, S.H. White, M.B. Ulmschneider, In silico partitioning and transmembrane insertion of hydrophobic peptides under equilibrium conditions, *J. Am. Chem. Soc.*, 133 (2011) 15487-15495.
- [24] M.B. Ulmschneider, J.P.F. Doux, J.A. Killian, J. Smith, J.P. Ulmschneider, Mechanism and kinetics of peptide partitioning into membranes, *J. Am. Chem. Soc.*, 132 (2010) 3452-3460.

- [25] M.B. Ulmschneider, J.C. Smith, J.P. Ulmschneider, Peptide partitioning properties from direct insertion studies, *Biophys. J.*, 98 (2010) L60–L62.
- [26] J.P. Ulmschneider, M. Andersson, M.B. Ulmschneider, Determining Peptide Partitioning Properties via Computer Simulation, *J. Memb. Biol.*, 239 (2011) 15–26.
- [27] M.B. Ulmschneider, J.P. Ulmschneider, N. Schiller, B.A. Wallace, G. von Heijne, S.H. White, Spontaneous transmembrane helix insertion thermodynamically mimics translocon-guided insertion, *Nat Commun*, 5 (2014) 4863.
- [28] J.P. Ulmschneider, M.B. Ulmschneider, Molecular Dynamics Simulations Are Redefining Our View of Peptides Interacting with Biological Membranes, *Accounts of Chemical Research*, (2018).
- [29] Y. Wang, C.H. Chen, D. Hu, M.B. Ulmschneider, J.P. Ulmschneider, Spontaneous formation of structurally diverse membrane channel architectures from a single antimicrobial peptide, *Nat. Comm.*, 7 (2016) 13535.
- [30] J.P. Ulmschneider, Charged Antimicrobial Peptides Can Translocate across Membranes without Forming Channel-like Pores, *Biophys. J.*, 113 (2017) 73–81.
- [31] J.A. Freites, D.J. Tobias, G. von Heijne, S.H. White, Interface connections of a transmembrane voltage sensor, *Proc. Natl. Acad. Sci. USA*, 102 (2005) 15059–15064.
- [32] R. Wolfenden, Experimental measures of amino acid hydrophobicity and the topology of transmembrane and globular proteins, *J. Gen. Physiol.*, 129 (2007) 357–362.
- [33] J.L. MacCallum, W.F.D. Bennett, D.P. Tieleman, Partitioning of amino acid side chains into lipid bilayers: Results from computer simulations and comparison to experiment, *J. Gen. Physiol.*, 129 (2007) 371–377.
- [34] W.C. Wimley, S.H. White, Experimentally determined hydrophobicity scale for proteins at membrane interfaces, *Nat. Struct. Biol.*, 3 (1996) 842–848.
- [35] M.B. Ulmschneider, M.S. Sansom, A. Di Nola, Properties of integral membrane protein structures: derivation of an implicit membrane potential, *Proteins*, 59 (2005) 252–265.
- [36] G. Krainer, P. Gracia, E. Frotscher, A. Hartmann, P. Groger, S. Keller, M. Schlierf, Slow Interconversion in a Heterogeneous Unfolded-State Ensemble of Outer-Membrane Phospholipase A, *Biophys J*, 113 (2017) 1280–1289.
- [37] J. Gumbart, B. Roux, Determination of Membrane-Insertion Free Energies by Molecular Dynamics Simulations, *Biophys. J.*, 102 795–801.
- [38] S. Jaud, M. Fernandez-Vidal, I. Nilsson, N.M. Meindl-Beinker, N.C. Hubner, D.J. Tobias, G. von Heijne, S.H. White, Insertion of short transmembrane helices by the Sec61 translocon, *Proc. Natl. Acad. Sci. USA*, 106 (2009) 11588–11593.
- [39] S. Dorairaj, T.W. Allen, On the thermodynamic stability of a charged arginine side chain in a transmembrane helix, *Proc. Natl. Acad. Sci. USA*, 104 (2007) 4943–4948.
- [40] I. Vorobyov, L. Li, T.W. Allen, Assessing Atomistic and Coarse-Grained Force Fields for Protein & Lipid Interactions: the Formidable Challenge of an Ionizable Side Chain in a Membrane, *The Journal of Physical Chemistry B*, 112 (2008) 9588–9602.
- [41] L. Li, I. Vorobyov, T.W. Allen, Potential of mean force and pKa profile calculation for a lipid membrane-exposed arginine side chain, *J. Phys. Chem. B*, 112 (2008) 9574–9587.
- [42] A.C.V. Johansson, E. Lindahl, Titratable Amino Acid Solvation in Lipid Membranes as a Function of Protonation State, *The Journal of Physical Chemistry B*, 113 (2008) 245–253.
- [43] L. Li, I. Vorobyov, A.D. MacKerell, Jr., T.J. Allen, Is arginine charged in a membrane?, *Biophys. J.*, 94 (2008) L11–L13.
- [44] A.C. Johansson, E. Lindahl, The role of lipid composition for insertion and stabilization of amino acids in membranes, *The Journal of Chemical Physics*, 130 (2009) 185101–185108.
- [45] A.C.V. Johansson, E. Lindahl, Protein contents in biological membranes can explain abnormal; solvation of charged and polar residues, *Proc. Natl. Acad. Sci. USA*, 106 (2009) 15684–15689.
- [46] A. Rychkova, S. Vicatos, A. Warshel, On the energetics of translocon-assisted insertion of charged transmembrane helices into membranes, *Proc. Natl. Acad. Sci. USA*, 107 (2010) 17598–17603.
- [47] N.M. Meindl-Beinker, C. Lundin, I. Nilsson, S.H. White, G. von Heijne, Asn- and Asp-mediated interactions between transmembrane helices during translocon-mediated membrane protein assembly, *EMBO Rep*, 7 (2006) 1111–1116.
- [48] K. Xie, T. Hessa, S. Seppälä, M. Rapp, G. von Heijne, R.E. Dalbey, Features of Transmembrane Segments That Promote the Lateral Release from the Translocase into the Lipid Phase†, *Biochemistry*, 46 (2007) 15153–15161.

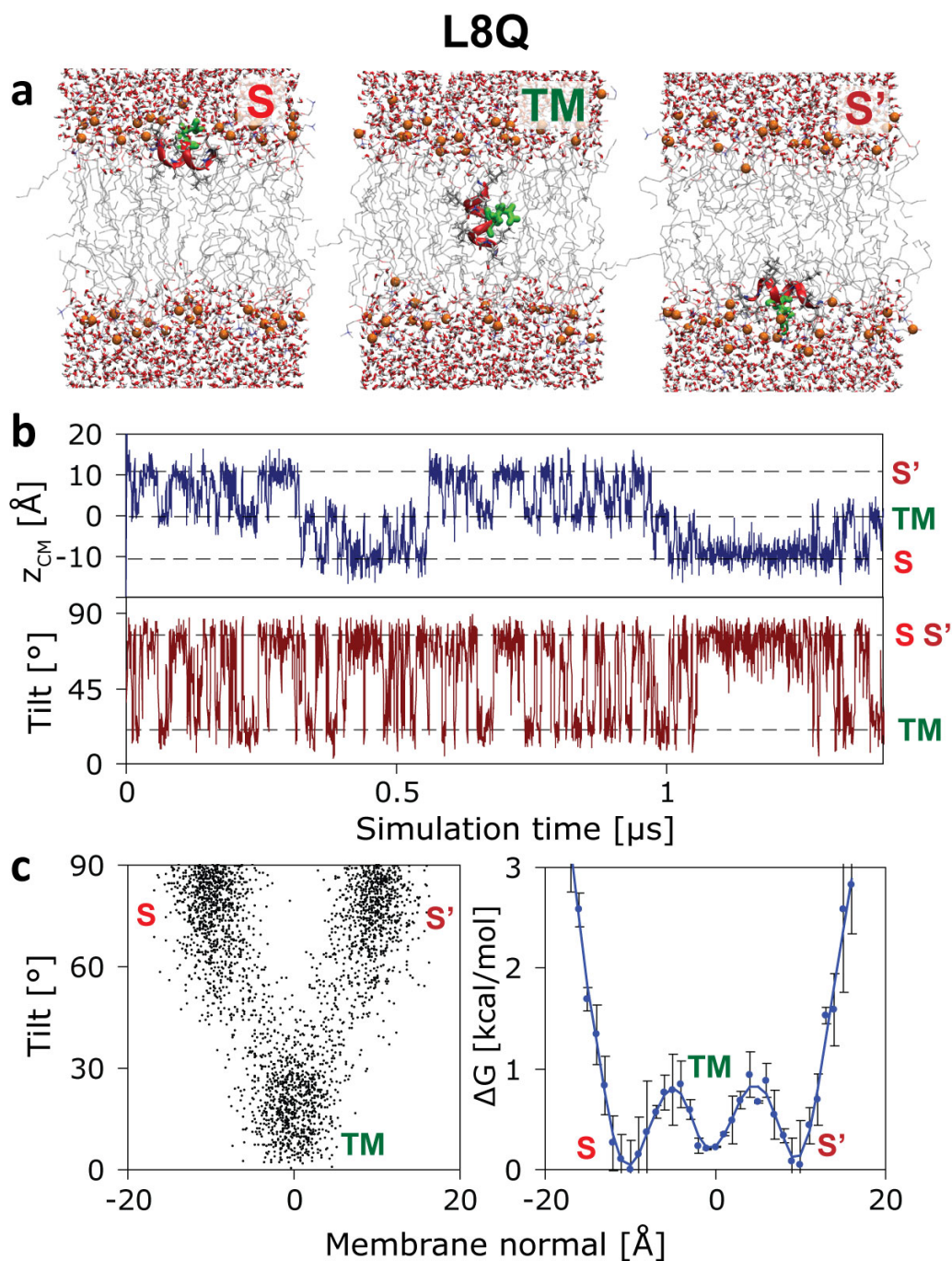
- [49] S.H. White, G. von Heijne, Do protein-lipid interactions determine the recognition of transmembrane helices at the ER translocon?, *Biochem. Soc. Trans.*, 33 (2005) 1012-1015.
- [50] S.H. White, G. von Heijne, Transmembrane helices before, during, and after insertion, *Curr. Opin. Struct. Biol.*, 15 (2005) 378-386.
- [51] S.H. White, Membrane protein insertion: the biology-physics nexus, *J. Gen. Physiol.*, 129 (2007) 363-369.
- [52] G. Von Heijne, Formation of transmembrane helices *in vivo*--Is hydrophobicity all that matters?, *J. Gen. Physiol.*, 129 (2007) 353-356.
- [53] S.U. Heinrich, W. Mothes, J. Brunner, T.A. Rapoport, The Sec61p Complex Mediates the Integration of a Membrane Protein by Allowing Lipid Partitioning of the Transmembrane Domain, *Cell*, 102 (2000) 233-244.
- [54] S.H. White, G. von Heijne, How translocons select transmembrane helices, *Annual Review of Biophysics*, 37 (2008) 23-42.
- [55] J. Gumbart, C. Chipot, K. Schulten, Free-energy cost for translocon-assisted insertion of membrane proteins, *Proc. Natl. Acad. Sci. USA*, 108 (2011) 3596-3601.
- [56] T. Junne, L. Kocik, M. Spiess, The Hydrophobic Core of the Sec61 Translocon Defines the Hydrophobicity Threshold for Membrane Integration, *Mol. Biol. Cell*, 21 (2010) 1662-1670.
- [57] E. Demirci, T. Junne, S. Baday, S. Bernèche, M. Spiess, Functional asymmetry within the Sec61p translocon, *Proc. Natl. Acad. Sci. USA*, (2013).
- [58] C. Baeza-Delgado, G. von Heijne, M.A. Marti-Renom, I. Mingarro, Biological insertion of computationally designed short transmembrane segments, *Scientific Reports*, 6 (2016) 23397.
- [59] A. Elazar, J.J. Weinstein, J. Prilusky, S.J. Fleishman, Interplay between hydrophobicity and the positive-inside rule in determining membrane-protein topology, *Proc. Natl. Acad. Sci. USA*, 113 (2016) 10340-10345.
- [60] J.L. MacCallum, W.F.D. Bennett, D.P. Tieleman, Transfer of Arginine into Lipid Bilayers Is Nonadditive, *Biophys. J.*, 101 (2011) 110-117.
- [61] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: A message-passing parallel molecular dynamics implementation, *Comp. Phys. Comm.*, 95 (1995) 43-56.
- [62] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, *J. Am. Chem. Soc.*, 118 (1996) 11225-11236.
- [63] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.*, 79 (1983) 926-935.
- [64] J.P. Ulmschneider, M.B. Ulmschneider, United atom lipid parameters for combination with the Optimized Potentials for Liquid Simulations all-atom force field, *J. Chem. Theory Comput.*, 5 (2009) 1803-1813.
- [65] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: A linear constraint solver for molecular simulations, *J. Comp. Chem.*, 18 (1997) 1463-1472.
- [66] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.*, 126 (2007) 014101.
- [67] H.J.C. Berendsen, J.P.M. Postma, W.F. Vangunsteren, A. Dinola, J.R. Haak, Molecular-dynamics with coupling to an external bath, *J. Chem. Phys.*, 81 (1984) 3684-3690.
- [68] B. Zhang, T.F. Miller, Hydrophobically stabilized open state for the lateral gate of the Sec translocon, *Proc. Natl. Acad. Sci. USA*, 107 (2010) 5399-5404.
- [69] C.L. Wee, M.B. Ulmschneider, M.S.P. Sansom, Membrane/Toxin Interaction Energetics via Serial Multiscale Molecular Dynamics Simulations, *J. Chem. Theory Comput.*, 6 (2010) 966-976.
- [70] B.A. Hall, J.P. Armitage, M.S.P. Sansom, Transmembrane Helix Dynamics of Bacterial Chemoreceptors Supports a Piston Model of Signalling, *PLoS Comput Biol*, 7 (2011) e1002204.
- [71] C.L. Wee, A. Chetwynd, M.S.P. Sansom, Membrane Insertion of a Voltage Sensor Helix, *Biophys. J.*, 100 (2011) 410-419.
- [72] Antreas C. Kalli, Benjamin A. Hall, Iain D. Campbell, Mark S.P. Sansom, A Helix Heterodimer in a Lipid Bilayer: Prediction of the Structure of an Integrin Transmembrane Domain via Multiscale Simulations, *Structure (London, England : 1993)*, 19 (2011) 1477-1484.
- [73] M. Andersson, Jakob P. Ulmschneider, Martin B. Ulmschneider, Stephen H. White, Conformational States of Melittin at a Bilayer Interface, *Biophys. J.*, 104 (2013) L12-L14.
- [74] Y. Wang, T. Zhao, D. Wei, E. Strandberg, A.S. Ulrich, J.P. Ulmschneider, How reliable are molecular dynamics simulations of membrane active antimicrobial peptides?, *Biochim. Biophys. Acta Biomem.*, (2014).

- [75] A.D. MacKerell, D. Bashford, 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, M. Karplus, All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†, *J. Phys. Chem. B*, 102 (1998) 3586-3616.
- [76] J.B. Klauda, R.M. Venable, J.A. Freites, J.W. O'Connor, D.J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A.D. MacKerell, R.W. Pastor, Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types, *J. Phys. Chem. B*, 114 (2010) 7830-7843.

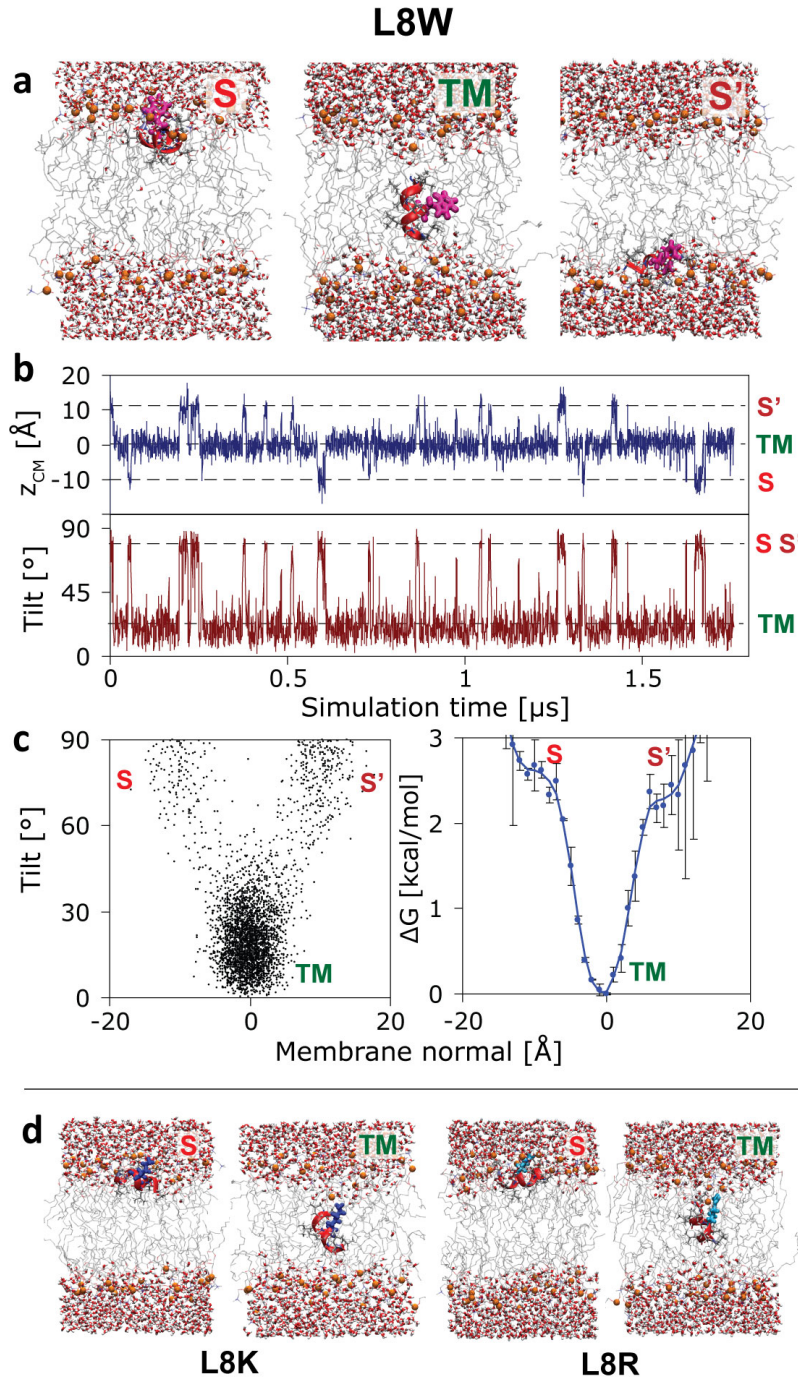


**Figure 1.** Schematic depiction of the partitioning processes studied to quantify the interaction of peptide sequences with lipid bilayers. (a) The partitioning equilibrium probed by the translocon-mediated insertion experiments is presumably between the translocon pore and the bilayer ( $\Delta G_{app}$ ), while the entry of the peptide into the translocon (“Enter”), and subsequent secretion (“Exit”) are thought to be non-equilibrium processes. (b) The outer membrane phospholipase A (OmpLA) can fold reversibly into synthetic liposomes in vitro. Host-guest experiments using OmpLA as a scaffold yield relative water-to-bilayer transfer free energies ( $\Delta\Delta G_{OmpLA}$ ) of amino acid side chains on the protein surface. (c) Direct partitioning simulations of freely inserting, sufficiently hydrophobic peptides reveal the equilibrium between surface adsorbed (S) and transmembrane inserted (TM) states. No soluble state (W) exists for these peptides, which unfold in water and precipitate out of solution.

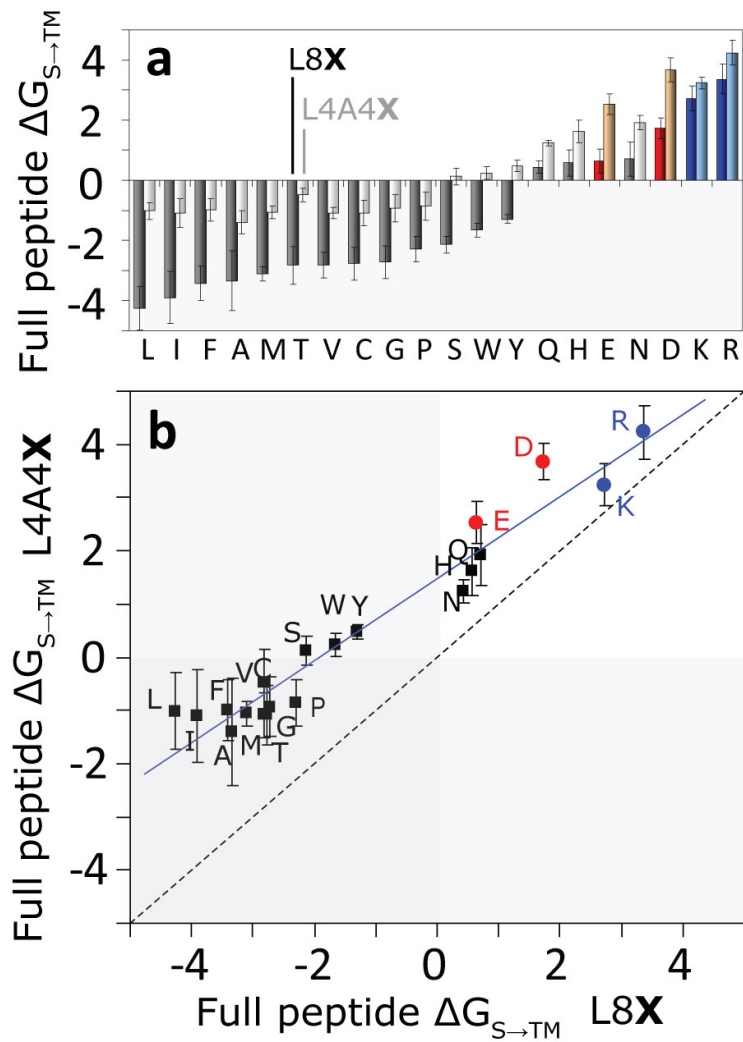




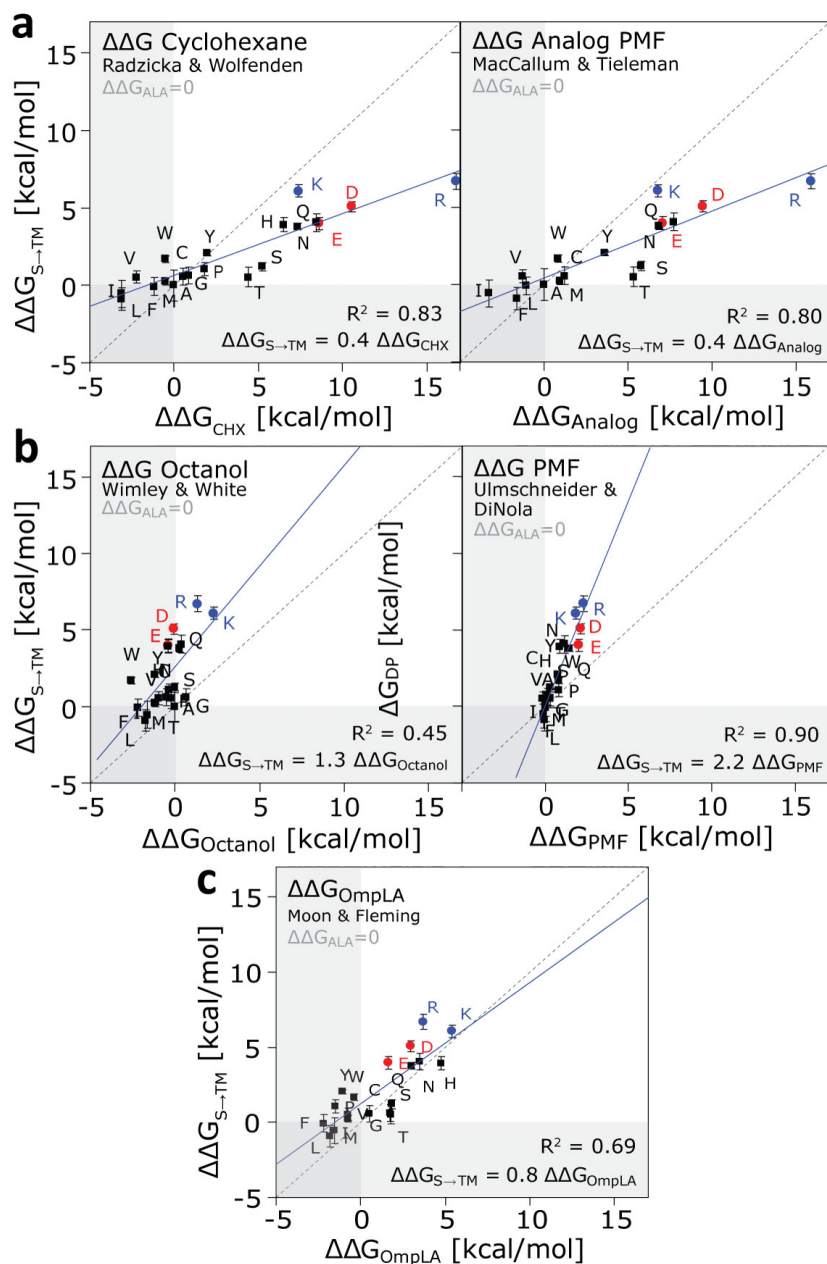
**Figure 2.** Surface-to-TM partitioning simulation of the L8Q peptide. (a) Two surface aligned (S, S') and one TM inserted states are seen (Q side chain in green). (b) Multiple transitions between these 3 states occur during 1.5  $\mu$ s, illustrated here by plotting the tilt angle and the center-of-mass position along the membrane normal as a function of simulation time. (c) A scatter plot reveals the clustering into the 3 distinct orientations. The insertion free energy profile ( $\Delta G$ ) of the peptide along the membrane normal is directly obtained from the peptide position via  $\Delta G = -RT \ln(p_z)$ , shifted so that the lowest value of  $\Delta G$  is zero. Error bars are derived from block averaging.



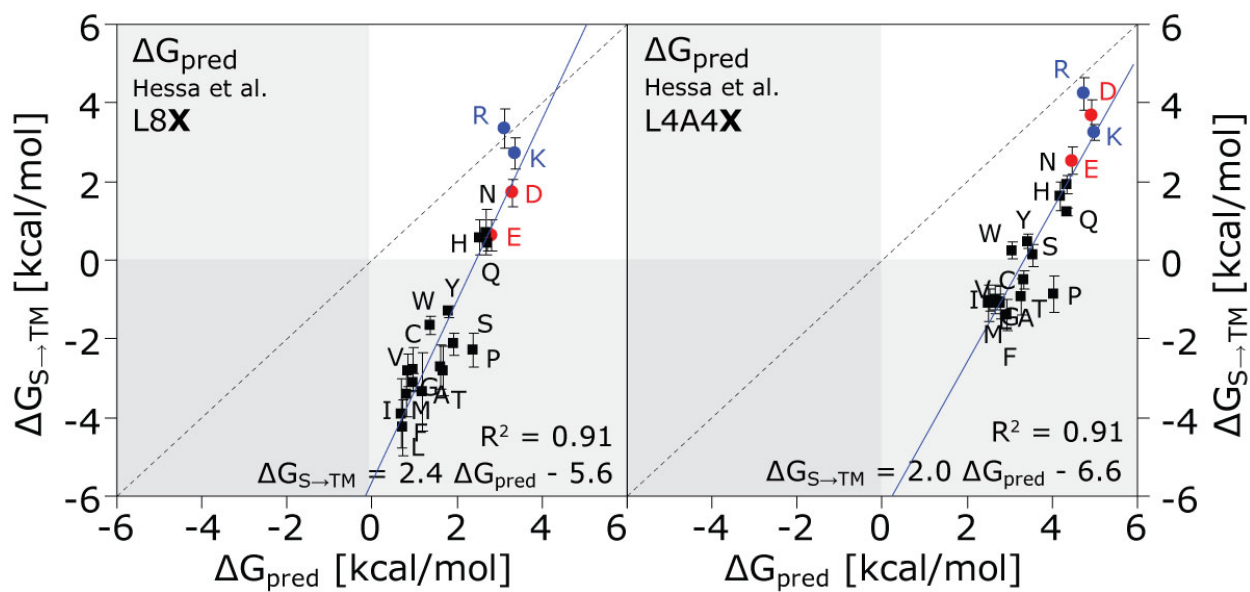
**Figure 3.** Surface-to-TM partitioning simulation of the L8W peptide. (a) The same principal states (S, S', TM) as for L8Q are found (W side chain in pink). (b) Multiple transitions are observed, but the equilibrium is strongly shifted towards the TM orientation compared to L8Q. (c) The shift is apparent in the scatter plot as well as the insertion free energy profile ( $\Delta G$ ). (d) The cationic charged sequences containing Lys (L8K) and Arg (L8R) follow the same pattern and can insert TM via snorkeling of the sidechain, pulling water molecules and lipid headgroups into the hydrophobic core. The S' state does not occur because the peptide cannot translocate. The only sequences found not to insert are the anionic peptides L8E and L8D.



**Figure 4.** (a) Full-peptide insertion scales for the surface-to-bilayer transfer of the L8X and L4A4X sequences. The values represent  $\Delta G_{S \rightarrow TM}$  for the full sequence, with X = central residue. Color coding: grey = hydrophobic, red = negatively charged residues, blue = positively charged residues. (b) Comparing the two scales reveals the underlying effect of the carrier sequence on the partitioning results, with  $\Delta G_{L8X} = 1.2 \times \Delta G_{L4A4X} - 1.9$ . The more hydrophobic carrier sequence inserts more strongly, but the slope is also affected.

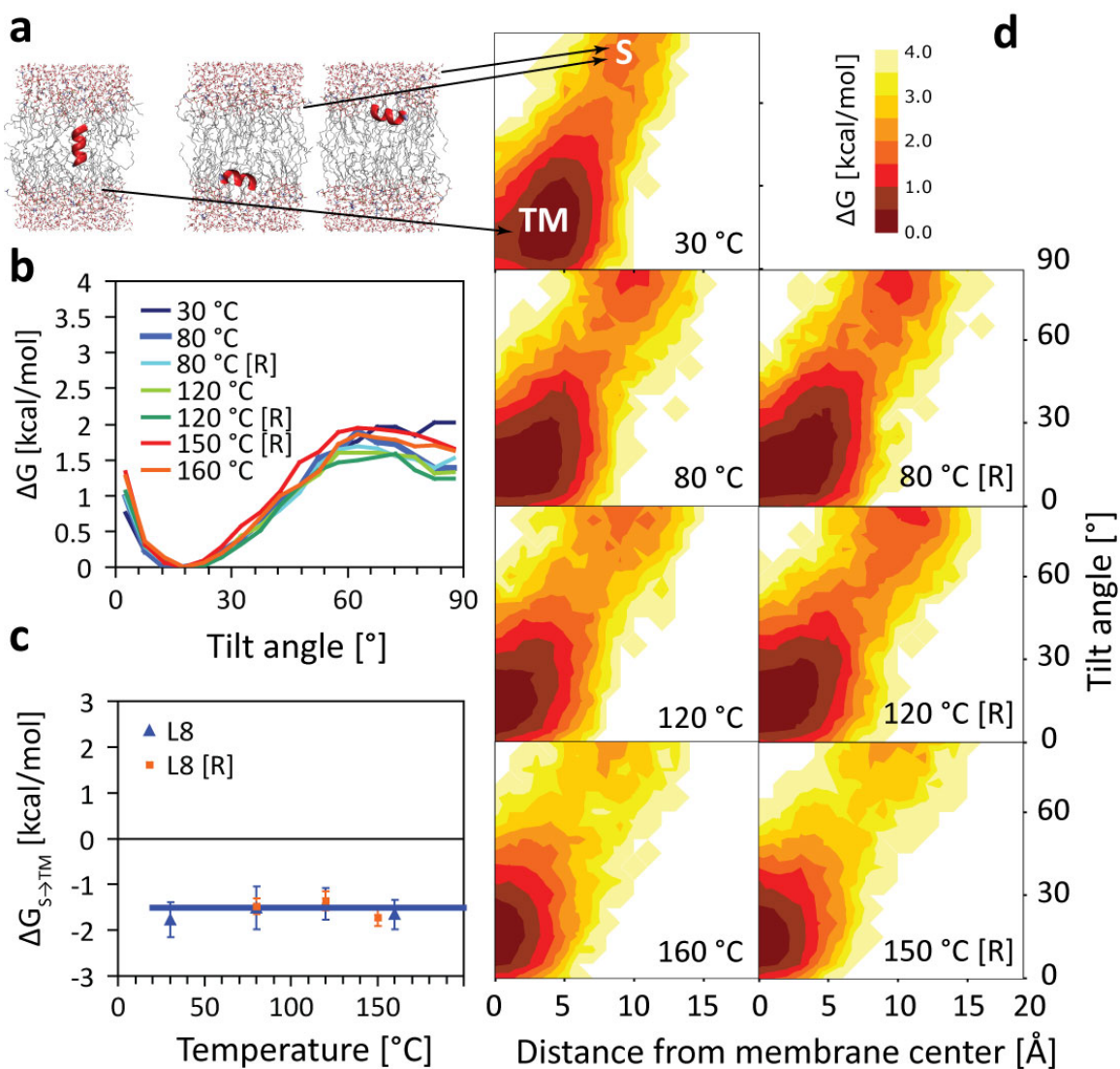


**Figure 5.** Side-chain only correlations with several experimental and computational scales. Relative free energies (to alanine,  $\Delta\Delta G_{ALA} = 0$ ) are compared as the underlying system in each case is different and absolute comparison of  $\Delta G_{S \rightarrow TM}$  not appropriate. (a). Both water-cyclohexane partitioning free energies ( $\Delta\Delta G_{CHX}$ ) of amino acid side chain analogs as well as PMF calculations of sidechain analogs into lipid bilayers ( $\Delta\Delta G_{Analog}$ ) correlate well with  $\Delta\Delta G_{S \rightarrow TM}$ , but are scaled up by a factor of  $\sim 2.5$ . These scales lack the contribution of the polar peptide backbone. (b). Whole-residue water/octanol hydrophobicity scales ( $\Delta\Delta G_{Octanol}$ ) derived from small pentapeptides correlate poorly with the simulation data, indicating that peptide water-to-octanol transfer is very different from surface-to-bilayer partitioning. The best correlation is achieved with statistical potentials derived from transmembrane amino acid distributions of membrane proteins of known structure ( $\Delta\Delta G_{PMF}$ ), which are compressed w.r.t.  $\Delta\Delta G_{S \rightarrow TM}$  by a factor of 2.2. (c). A similarly good agreement is seen with the recent partitioning experiments using the outer membrane phospholipase A (OmpLA) as a transmembrane scaffold ( $\Delta\Delta G_{OmpLA}$ ). The blue line indicates the best linear fit.



**Figure 6.** Comparison of full peptide insertion energies  $\Delta G_{S \rightarrow TM}$  to full-sequence translocon-mediated insertion data ('biological scale',  $\Delta G_{pred}$ ). For both L8X and L4A4X, the correlation (w/o the outlier proline) is high ( $r^2 = 0.91$ ). The computational values are stretched by a factor of  $\sim 2$ , and strongly shifted towards the TM state.





**Figure 7.** Temperature independence of hydrophobic S-to-TM partitioning, as shown using MD for a polyleucine with 8 residues (L8). No temperature effect is visible over the large range 30-160 °C. The simulations are 5  $\mu$ s each, so one-dimensional free energy profiles (b), and the overall transfer free energy  $\Delta G_{S \rightarrow TM}$  (c), are well converged. Simulations using helical restraints are denoted with [R]. No unfolding is observed even at very high temperatures, and both unrestrained and restrained simulations yield the same  $\Delta G_{S \rightarrow TM} \approx -1.56 \pm 0.16$  kcal/mol (c). Two-dimensional free energy profiles (as a function of membrane insertion and tilt angle) reveal that the position of the TM minimum is slightly shifted towards the bilayer center upon raising the temperature, which occurs due to slight thinning of the membrane (d). However, the S-to-TM free energy is unaffected.

TOC graphic

