J Comput Aided Mol Des (2011) 25:895–911
DOI 10.1007/s10822-011-9470-9

Efficient molecular mechanics simulations of the folding, orientation, and assembly of peptides in lipid bilayers using an implicit atomic solvation model

Andrew J. Bordner · Barry Zorman · Ruben Abagyan

Received: 7 April 2011 / Accepted: 25 August 2011 / Published online: 9 September 2011

Abstract Membrane proteins comprise a significant fraction of the proteomes of sequenced organisms and are the targets of approximately half of marketed drugs. However, in spite of their prevalence and biomedical importance, relatively few experimental structures are available due to technical challenges. Computational simulations can potentially address this deficit by providing structural models of membrane proteins. Solvation within the spatially heterogeneous membrane/solvent environment provides a major component of the energetics driving protein folding and association within the membrane. We have developed an implicit solvation model for membranes that is both computationally efficient and accurate enough to enable molecular mechanics predictions for the folding and association of peptides within the membrane. We derived the new atomic solvation model parameters using an unbiased fitting procedure to experimental data and have applied it to diverse problems in order to test its accuracy and to gain insight into membrane protein folding. First, we predicted the positions and orientations of peptides and complexes within the lipid bilayer and compared the simulation results with solid-state NMR structures. Additionally, we performed folding simulations for a series of host–guest peptides with varying propensities to form alpha helices in a hydrophobic environment and compared the structures with experimental measurements. We were also able to successfully predict the structures of amphipathic peptides as well as the structures for dimeric complexes of short hexapeptides that have experimentally characterized propensities to form beta sheets within the membrane. Finally, we compared calculated relative transfer energies with data from experiments measuring the effects of mutations on the free energies of translocon-mediated insertion of proteins into lipid bilayers and of combined folding and membrane insertion of a beta barrel protein.

Keywords Membrane proteins · Implicit solvation model · Molecular mechanics · Peptide folding and association · Amphipathic peptide

ICM implementation: ruben@molsoft.com

A. J. Bordner (✉) · B. Zorman
Mayo Clinic, 13400 E Shea Blvd, Scottsdale, AZ 85259, USA
e-mail: bordner.andrew@mayo.edu

R. Abagyan (✉)
Molsoft LLC, 11199 Sorrento Valley Road #209, San Diego, CA 92121, USA
e-mail: ruben@molsoft.com

Introduction

Integral membrane proteins are encoded by a significant portion (20–30%) of the genomes of sequenced organisms [1] and fulfill diverse functions as receptors, transporters, channels, structural anchors, and enzymes. Furthermore, they are the targets of approximately 60% of approved drugs [2] and are therefore of considerable biomedical interest. In spite of their ubiquity and medical importance, comparatively few experimental structures have been solved in comparison with non-membrane proteins due to technical challenges [3, 4].

Computational methods can help by providing atomistic models of membrane proteins that can be used to generate experimentally verifiable hypotheses about a protein's structure, function, interactions, and folding energetics. However, computer simulations of membrane proteins have their own challenges, one of which is accounting for the energetics of solvent-protein interactions in the spatially

heterogeneous environment comprised of the lipid bilayer and water. Molecular dynamics simulations of membrane proteins often include explicit all-atom lipids, but at a large computational cost [5]. Coarse-grained models use large particles to represent clusters of neighboring atoms, which reduces the computational burden; but they still remain computationally expensive [6, 7]. Implicit solvation models offer a further significant increase in speed so that even larger systems can be modeled. Furthermore, the solvent degrees of freedom are averaged out so that implicit solvation models can be used with molecular mechanics global energy optimization, thus providing an efficient method for predicting membrane protein structures.

Implicit solvation models for membrane proteins

Several different implicit solvation models for membrane proteins have been previously developed. One approach is to derive residue-level knowledge-based potentials that smoothly vary as a function of each residue's depth within the membrane [8, 9]. While such potentials are useful for determining the orientation of a protein relative to the membrane, they use only a single point to describe each residue and so are inappropriate for the atomic level modeling considered in this study. A number of studies [10–13] have modified the popular Generalized Born (GB) electrostatics model to represent the membrane as a low dielectric slab region within the high dielectric ($\varepsilon \approx 80$) aqueous environment and added a non-polar surface tension contribution proportional to the solvent accessible surface area (SASA) to create so-called GBSA models that are applicable to molecular dynamics simulations of membrane proteins. Another membrane solvation model, called IMM1 [14], is a modified version of the EFF1 aqueous implicit solvation model [15] in which the parameters vary along the direction of the membrane normal. Finally, another type of model calculates the total solvation free energy as a sum over the solvent accessible surface area (SASA) multiplied by corresponding atomic solvation parameters (ASPs). Such ASP models were originally developed for aqueous solvation [16–20] and have been adapted to membrane solvation by Efremov et al. [21]. Another study used an ASP membrane solvation model in order to calculate the position of membrane proteins relative to the lipid bilayer [22]. However, because the ASPs in that work represent water–membrane transfer free energies and no terms for aqueous solvation were included, that model is only applicable to rigid optimization of a protein's position in which the atoms' SASA values do not change. In this study, we derive and employ an ASP membrane solvation model for use in all-atom modeling of membrane proteins.

The procedure for fitting ASPs starts with a relatively large number of initial atom types that are first clustered into a fixed number of groups before using linear regression to fit the ASPs to gas→cyclohexane transfer free energy data. We also examined the effects of the number of atom types and membrane geometry on the accuracy of predicting the orientations of transmembrane peptides solved by solid state NMR [23]. A set of values that yielded good agreement with the experimental structures was then chosen for use in all subsequent simulations.

Additionally, we have also investigated several applications of the solvation model in order to assess its accuracy and to gain insights into the positioning, folding, and non-covalent association of peptides within a lipid bilayer. First, we predicted the positions and orientations of different transmembrane alpha-helical peptides and complexes and compared the results with experimental solid state NMR structures. As mentioned above, these results were used to select the number of atom types and membrane geometry for the solvation model employed in the other simulations in this study. We next calculated the optimal positions and orientations of various amphipathic peptides relative to the membrane. In addition, we attempted the more difficult task of ab initio prediction of the structures of these amphipathic peptides and their positions relative to the membrane starting from an unfolded fully extended structure. As another test, we performed molecular mechanics folding simulations for a series of host–guest peptides studied by Liu and Deber [24] in order to see whether they reproduce the observed alpha-helical content. The structural preferences of five different 20-mer homopolymers were also examined and compared with the corresponding results for the Liu-Deber peptides. We also performed dimer simulations for a series of host–guest peptides, some of which were experimentally found to form beta sheets within the membrane [25, 26]. Finally, we compared calculated relative transfer free energies for variant peptides with data from two experimental studies. One study by Hessa et al. [27] measured the biological apparent insertion free energies deduced from experiments on translocon-mediated insertion of peptides into lipid bilayers while the other study by Moon and Fleming [28] measured the relative free energies of folding and membrane insertion for the wild type OmpLA beta barrel membrane protein and single point mutants.

Methods

Molecular mechanics simulations

The ICM program (Molsoft LLC) was used for all molecular mechanics simulations. ICM performs global energy optimization in torsion angle space using biased probability Monte Carlo sampling [29]. A large reduction of about a factor 5–10 in the number of conformational

degrees of freedom for torsion angle coordinates compared with Cartesian atomic coordinates leads to more efficient sampling and a larger radius of convergence [30]. In addition to the solvation energy term, discussed below, the energy function included van der Waals, hydrogen bond, electrostatics, and torsion energy components calculated using the ECEPP/3 force field [31–33]. The electrostatic energy was calculated using a distance dependent dielectric constant, $\varepsilon = 4r$, and the van der Waals potential was modified to approach a finite value of 7 kcal/mol at small separation distances [34] in order to improve sampling and avoid numerical instabilities. Simulations were run for a total of 10^6 function calls for rigid optimization of the peptide positions and orientations relative to the membrane and 2×10^8 function calls for the simulations of one and two flexible peptides, unless otherwise noted. Because the NMR structures of the bacterial coat protein and MerF transporter (PDB entries 1MZT and 2H3O) only provided backbone atom coordinates, their rigid optimization simulations included sampling side chain torsion angles and so were run longer (10^8 function calls) to insure adequate sampling. Three independent simulation runs, using different random number seeds, were performed in the cases with $\geq 10^8$ function calls in order to verify convergence, and the lowest energy conformation from all simulations was selected for subsequent analysis. Except for rigid optimization, in which the experimental structure was used, all ICM simulations began with the peptide backbone in a fully extended (unfolded) conformation.

Membrane solvation model

The membrane solvation model uses atomic solvation parameters (ASPs) that depend on the local solvent environment. The membrane is oriented with its normal direction parallel to the z-axis and the center of the membrane is defined by $z = 0$. The solvent environment is then defined by three regions in terms of the z coordinate of each atom: the membrane core region for $|z| \leq a$, the interface region for $a < |z| < b$, and the aqueous region for $|z| \geq b$. One set of ASPs, σ_i^{mem}, describe solvation in the membrane core and another set, σ_i^{aq}, describe solvation in the aqueous environment with linearly interpolating values in the transition region. In other words, the ASP for an atom with a particular z coordinate is

$$\sigma_i(z) = \begin{cases} \sigma_i^{\mathrm{mem}} & |z| \leq a \\ \dfrac{1}{b-a}\left[(b-|z|)\,\sigma_i^{\mathrm{mem}} + (|z|-a)\sigma_i^{\mathrm{aq}}\right] & a < |z| < b. \\ \sigma_i^{\mathrm{aq}} & |z| \geq b \end{cases} \quad (1)$$

Thus the solvation energy varies continuously with each atom's position, which improves energy optimization convergence compared with a discontinuous (two region) definition. The solvation energy is then

$$E_{\mathrm{solv}} = \sum_{i=1}^{N_{\mathrm{atoms}}} \sigma_i(z_i) A_i, \quad (2)$$

in which atom i has z coordinate z_i and solvent accessible surface area (SASA) A_i. For simplicity and also to maintain continuity of the solvation energy across the membrane-solvent boundary, the SASA is calculated using a constant probe radius of 1.4 Å in all regions. The larger size of lipid molecules compared with water molecules suggests that a larger probe radius may be appropriate in the membrane region. However, other than an overall approximate scaling of the ASPs by a constant factor depending on the probe radius, the differences in fit ASPs are expected to be small since they depend only on small differences in the shape of the solvent accessible surface near boundaries between contributions from different surface atoms. The study by Efremov et al. [21] found a strong correlation between ASPs derived with probe radii of 1.4 Å for water and 3.3 Å for cyclohexane solvation, supporting their approximate linear dependence.

Fitting atomic solvation parameters

Previously reported aqueous solvation parameters [35], which are implemented in the ICM program, were used for σ_i^{aq}. The membrane ASPs, σ_i^{mem}, were fit to best reproduce experimental gas→cyclohexane transfer free energies of amino acid analogs [36], including a size correction [37]. It should be noted that the aqueous ASPs were fit to the gas→water transfer free energies of the same amino acid analogs reported in that paper [37]. All available free energy data for 19 residue analogs, excluding proline, were used. The compounds consisted of the corresponding side chain truncated at the β-carbon and were assumed to be in a fully extended conformation for the calculation of SASAs. The fitting procedure involved first determining an optimal grouping of an initial set of 14 protein atom types defined in the ICM program, given the number of groups, N, using k-means clustering [38] with $k = N$. The first step in clustering was to calculate a vector $\mathbf{v}_j$ for atom type j, in which each component $(\mathbf{v}_j)_i$ is the total SASA of atom type j in compound i. These vectors are then clustered using the k-means algorithm. This procedure groups together atom types that have similar radii and that also co-occur in the same compounds, for instance amide nitrogen (type 3) and carbonyl oxygen (type 7). The definitions of the initial atom types and their radii are shown in Table 1. Next, for each number of atom type groups N, the ASPs were fit to the experimental data using linear regression assuming that all atom types within the same group have the same ASP.

Table 1 Initial atom types and radii used to calculate the SASA

Atom type number	Description	Atomic radius (Å)
1	Primary aliphatic and carbonyl C	1.95
2	Aromatic C	1.80
3	Amide N	1.70
4	Lysine$^+$ N_ζ	1.70
5	Arginine N (N_ε, $N_{\eta 1}$, and $N_{\eta 2}$)	1.70
6	Hydroxyl O	1.60
7	Carbonyl O	1.40
8	Aspartic acid$^+$ and glutamic acid$^+$ O	1.40
9	Cysteine (sulfhydryl) S	2.00
10	Methionine or disulfide S	1.85
11	Secondary aliphatic C	1.95
12	Tertiary aliphatic C	1.95
13	Histidine N_ε	1.70
14	Secondary amine N (backbone N, histidine N_δ, tryptophan N_ε)	1.70

Then, both the membrane geometry, which is defined by the parameters a and b in Eq. 1, and the number of groups, N, were varied in order to find which combinations yield good agreement between the calculated orientations of a set of TM alpha-helical peptides and the values experimentally measured by solid state NMR. N values from 4 to 10 were examined. Simulations were run using all combinations of total membrane thickness values excluding the transition region (or $2a$ from Eq. 1) from 18 Å to 30 Å in 2 Å steps and transition region thickness values (or $b-a$ from Eq. 1) of 0.01 Å (e.g. effectively no transition region), 2, 5, 7, and 10 Å.

Rigid optimization of position and orientation of transmembrane proteins relative to the membrane

The position and orientation relative to the membrane for a set of five transmembrane α helical peptides or peptide complexes were calculated using Monte Carlo sampling of the six translational and rotational degrees of freedom in order to find the global minimum of the solvation energy. The experimental structures were used for the simulations. The results were then compared with the tilt angle of the helix axis relative to the membrane normal axis, α, and the rotation angle about the helix axis, θ, obtained from solid state NMR structures. For NMR experiments with multiple structures, simulations were performed for each structure and the average angles calculated from the results.

Only the position and orientation of each peptide were optimized through sampling the six relevant degrees of freedom using the ICM molecular mechanics program. In other words, we performed rigid body optimization of the placement of the experimental structure relative to the membrane. The orientation of a TM alpha-helical peptide relative to the membrane was defined by two angles: α is the angle between the helix axis and the membrane normal direction, or z-axis, and θ is the rotation angle about the helix axis relative to an arbitrary reference conformation. Solid state NMR [23] only determines these two degrees of freedom. While α is uniquely defined for a single TM peptide, θ is not so that only differences between θ angles (e.g. between calculated and experimental values) are meaningful. The helix axis was calculated by applying singular value decomposition (SVD) to the centered backbone atom (C_α, C, and N) coordinates, $\mathbf{x}-\overline{\mathbf{x}}$, in which $\overline{\mathbf{x}}$ is the backbone centroid. The helix axis is the principal component corresponding to the largest eigenvalue.

Finally, the differences between the calculated and experimental angles were used to select the optimal number of ASPs and membrane geometry parameters, as discussed in the Results section. These selected parameters were then used in all subsequent simulations.

Folding of amphipathic peptides

In the ab initio folding simulations of amphipathic peptides, a modest boundary constraint was applied in order to reduce the initially infinite space of translational degrees of freedom in the Monte Carlo optimization. Because the membrane is laterally homogenous, i.e. the solvation energy does not change in the x- and y-axis directions, a constraint was applied during each simulation in order to reduce unproductive movement of the peptide in these directions. The constraint energy was a function of the distance of the C_α atom of the central peptide residue (or residue $m/2$ for peptides with an even number of residues m) from the origin, which is at the membrane center, and was zero for distances less than 30 Å and increased quadratically with weight 3.5 kcal/(mol Å^2) for larger distances. We also ran an additional 3 simulations for the longest peptides, delta-hemolysin and magainin, and selected the lowest energy conformation from among all simulations (as in the other cases) in order to improve convergence. The next two sections describe the procedures for calculating transfer free energies for comparison with the experimental studies of Hessa et al. [27] and Moon and Fleming [28], respectively.

Transfer energies for individual residues in an α helix

We performed molecular mechanics simulations of 13 residue poly-alanine peptides with different uncharged

residues substituted at the central position in order to calculate the contribution of each residue type to the total free energy of transferring helical peptides from water into the membrane. These were then compared with the apparent transfer free energies determined by Hessa et al. [27] (see Results section). The peptide backbone was constrained to be an α helix with $\phi = -60°$, $\psi = -45°$. Simulations were performed using both the water and membrane solvation models by sampling only side chain torsion angles of the substituted residues and its two adjacent residues for a total of 8×10^6 energy evaluations. The transfer free energy was then calculated as the difference between the total energies of the final structures in each environment. Finally, the relative transfer free energies were calculated by subtracting the energy of the poly-alanine peptide from all values (so that by definition $\Delta G = 0$ for alanine).

Transfer energies for residues in a β barrel membrane protein

Simulations were also run to calculate transfer free energies for all uncharged residue substitutions at a central lipid-facing site in OmpLA (residue 210) for comparison with the experimental results of Moon and Fleming [28] (see "Results and discussion" section). In a membrane environment, side chains for residues 208–212 were relaxed via Monte Carlo sampling starting from the X-ray structure of OmpLA (PDB entry 1QD5 [39]). Next, in a water environment, the side chains for the same segment, residues 208–212, were relaxed starting with the backbone in an extended conformation with charged N- and C-termini. Simulations were run for 8×10^6 function evaluations, and the transfer energy calculated as the difference between the total energies of the final structures in the two solvent environments. Finally, the transfer free energies relative to alanine were calculated by subtracting the transfer free energy for alanine from all values.

Results and discussion

We first discuss fitting the solvation model parameters. Next, we examine a series of different simulations to predict peptide structures, many of which compare favorably with experimental results. Finally, we compare the calculated effects of residue substitutions on transfer free energies with experimentally determined values.

Effects of membrane geometry and atom types on rigid optimization results

We first examined the effect of increasing the total membrane thickness for the case of an infinitesimal transition region thickness (0.01 Å). In this case, the predicted tilt angle, α, for the proteins with a single transmembrane helix (1A11, 1MZT, and 2KB7) steadily decreased with increasing membrane thickness. This is expected since the optimal orientation of the helix has its hydrophobic central portion within the membrane so that its axis moves closer to the membrane normal as the membrane region becomes thicker. No such trend is apparent for the multi-span protein 2H3O and the multi-subunit protein complex 2KQT. Also, there is no monotonic trend in the helix rotation angle, θ, as the membrane thickness is increased.

The small number of solid state NMR structures (5) in Table 3 and the stochastic global energy optimization procedure used for structure prediction precludes rigorous optimization of the number of atom types, N, and membrane geometry, specified by a and b; however, acceptable ranges for these parameters can be determined by examining the prediction accuracy for different values. The overall prediction accuracy was assessed by the average absolute differences in α and θ, in which the average was calculated over all NMR structures. We first examined all parameter combinations that yielded the smallest average absolute difference in the tilt angle, $<|\Delta\alpha|>$, since this angle is the most accurately predicted by the simulations and also is the most sensitive to membrane geometry. In all cases with the lowest error in α, with $<|\Delta\alpha|> < 6.0°$, the membrane thickness, including the transition layers, was within the fairly narrow range 28 Å $\leq 2a + 2b \leq$ 32 Å. Similarly all cases with the lowest errors in θ, with $<|\Delta\theta|> < 21.0°$, had an overall membrane thickness within a wider but overlapping range, 18 Å $\leq 2a + 2b \leq$ 32 Å. In contrast, the transition region thickness, b, and number of atom types, N, had relatively little effect on prediction accuracy. Only the largest transition region thickness, $b = 10$, and the smallest number of atom types, $N = 4$, were excluded from the high accuracy cases with $<|\Delta\alpha|> < 6.0°$. Based on these results, we chose intermediate parameter values within the ranges observed in high accuracy predictions: the number of atomic solvation parameters $N = 8$, the central membrane thickness, $2a = 20$ Å, and an overall membrane thickness, $2a + 2b = 30$ Å (or transition layer thickness $b = 5$ Å).

Comparison with membrane thickness derived from other studies

The 30 Å overall membrane thickness is in reasonable agreement with the thickness of the hydrophobic core observed in experimental X-ray and neutron diffraction studies of DOPC bilayers [40]. Importantly, that study also emphasized the dynamic nature of the solvated lipid bilayer as well as the incursion of water molecules into the bilayer, whose extent depends on the degree of hydration. The

overall membrane thickness also is in general agreement with computational estimates derived from the extent of the hydrophobic regions in membrane proteins [22]. That study also found variation in the estimated thicknesses of different biological membranes; however, only membranes from 2 out of 9 sources, namely the outer membrane of gram-negative bacteria and the cell wall membrane of mycobacteria, were significantly different from 30 Å. The hydrophobic thickness in that study was defined as twice the depth at which the aqueous hydration component is 50%, which would correspond to 25 Å in our model; however, that model assumed a sharper sigmoidal cutoff between the hydrophobic and aqueous solvation regions, preventing a direct comparison. Overall, the estimated thicknesses for different membranes suggest that although membrane properties, including the average thickness of the aliphatic lipid tail region, vary with their composition, the membrane geometry chosen here is typical and should apply to biological membranes from many different sources. In any case, the membrane thickness can be easily adjusted in our solvation model to accommodate significant deviations from the typical values derived here. Finally, we note three complicating factors in unambiguously defining the membrane geometry: (1) membranes are dynamic structures so that no sharp boundaries exist between different regions, (2) experimental structures of membrane proteins are usually solved in detergent rather than in a lipid bilayer so that they do not provide direct information on the position of the protein relative to the membrane, and (3) the membrane protein may perturb the thickness and/or shape of the immediately surrounding lipid bilayer [41].

Membrane atomic solvation parameters

The optimal ASP values are shown in Table 2. The RMSD between the experimental gas→cyclohexane transfer energies and the values calculated with the ASPs was only 0.15 kcal/mol, indicating a close fit. A plot of the calculated versus the experimental free energy values is shown in Fig. 1. Next we compared our ASPs with those obtained by Efremov et al. [21], which is the only other study to derive similar parameters. We first note that Efremov et al. fit ASPs to the same gas→cyclohexane transfer free energies used in this study (from [37]); however, we employed a different set of atom types and a different fitting procedure, so that some distinct ASP values are expected. Our ASPs for aliphatic carbon (atom type 1) and aromatic carbon (atom type 2), namely −8.24 and −24.46 cal/(mol Å^2) respectively, are quite similar to those obtained in that study for the same atom types, −11 and −26 cal/(mol Å^2) respectively. Likewise the ASPs for the two charged oxygen atom types, −26.70 cal/(mol Å^2) for type 7 and −21.75 cal/(mol Å^2) for type 8, are fairly close to the ASP for the single charged oxygen atom type in that study of −20 cal/(mol Å^2). In contrast, most of the remaining ASPs differ significantly. The largest deviation is for the sulfur ASP, which was found to be −33.90 cal/(mol Å^2) for both atom types 9 and 10 in this study but was only −2 cal/(mol Å^2) in Efremov et al. In addition, the ASP for uncharged or hydroxyl oxygen (atom type 6) is −19.97 cal/(mol Å^2) here but was determined to be 3 cal/(mol Å^2) in that study. Finally, uncharged nitrogen atom types (3, 13, and 14) had ASPs of −26.70 cal/(mol Å^2) and −33.90 cal/(mol Å^2), which differ from the value of −59 cal/(mol Å^2) in that study. Importantly, these differences in ASP values are greater than the estimated parameter uncertainty in Efremov et al. and are too large to be explained by differences in atomic SASAs. Thus it

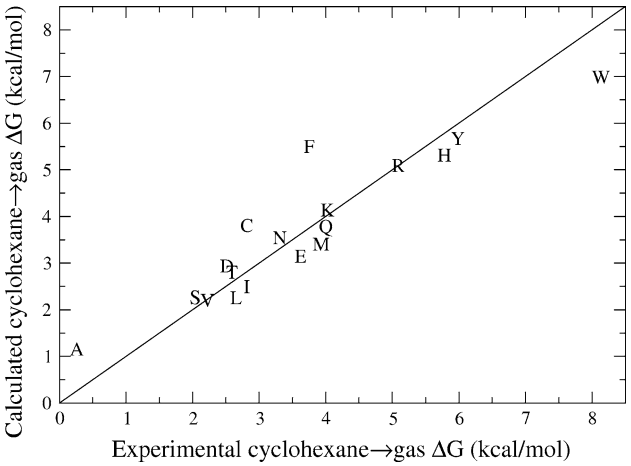

Fig. 1 Plot of the calculated versus experimental cyclohexane → gas transfer energy for residue side chain analog compounds. Solvation free energy values were calculated using the ASPs in Table 2, which minimize the RMSD between the computed and experimental free energy values, while experimental values were taken from Sharp et al. [37]

Table 2 Membrane atomic solvation parameters

Atom type numbers	Atomic solvation parameter (cal/(mol Å^2))
1	−8.24
2	−26.46
3, 7	−26.70
4, 9, 10, 13, 14	−33.90
6, 12	−19.97
5	−24.30
8	−21.75
11	−11.54

Atom type numbers are defined in Table 1

appears that the use of different atom types has yielded a different partitioning of the free energy among individual atoms. Unlike Efremov et al., in which atom type definitions, given a fixed number of types *N*, were chosen a priori, we instead began with a large number of atom types and then used clustering in order to ascertain the best way to partition them into N groups. In the following sections, we will discuss the results from a series of membrane peptide simulations that demonstrate what accuracy can be expected when the new solvation parameters are combined with the ECEPP/3 force field, and that also may help in interpreting the experimental data for these systems.

Transmembrane peptide positions and orientations

Table 3 compares the predicted positions and orientations of five transmembrane peptides with values from solid state NMR experiments, in which average experimental tilt angles are in the range of approximately 11°–23°. One structure, a viral proton channel (PDB entry 2KQT), is a homotetramer, while the remaining structures are monomers. Both the experimental and calculated orientations of the structures relative to the membrane are illustrated in Fig. 2.

Using the chosen solvation parameters and membrane geometry, the helix tilt angle is accurately determined for all proteins with absolute differences between the predicted and experimental values averaged over all NMR structures varying from 3.61° to 17.7°. The largest tilt angle error (17.7°) is for the MerF mercury transport protein (PDB entry 2H3O). This is likely due to the fact that neither of the TM helices fully traverse the membrane and also because the side chain atoms are missing from the NMR structure and so needed to be predicted along with the overall protein orientation. The errors in the rotation angles about the helix axes are generally higher in magnitude than for the helix tilt angles. This can be explained by the smaller "lever arm" over which the orientating effects of differences in local surface hydrophobicity and consequently lower torque due to solvation. For the helix rotation angle, the relevant lever arm distance is approximately the radius of the α-helix, or about 3 Å, whereas for the helix tilt angle, the lever arm distance is approximately the half the width of the membrane, or 15 Å.

We note that a null model without the membrane solvation energy would predict all helix orientations with equal probability. Thus an estimate of the statistical significance for the difference between the predicted and experimental helix rotation angles, ignoring multiple NMR structures, is simply $|\Delta\theta|/180$. Because the tilt angle is a spherical polar angle, the statistical significance is calculated as

Table 3 Comparison between the calculated and experimental orientations of transmembrane helices relative to the membrane

PDB entry	Number of NMR structures	Description	Helix residues	Average experimental tilt angle[b] $\langle\alpha_{exp}\rangle$	Average calculated tilt angle[b] $\langle\alpha_{calc}\rangle$	Average tilt angle difference $\langle\lvert\alpha_{exp} - \alpha_{calc}\rvert\rangle$	Average rotation angle difference $\langle\lvert\theta_{exp} - \theta_{calc}\rvert\rangle$
1A11 (1CEK)[c] [65]	10	Acetylcholine receptor M2	1–25 (all)	11.3° ± 0°	11.7° ± 5.65°	4.21°	23.4°
1MZT[a] [66]	1	Fd bacteriophage pVIII major coat protein	22–45	18.4°	36.1°	17.7°	12.6°
2H3O[a] [67]	1	MerF mercury transport protein	50–66	20.3°	19.1°	1.17°	23.1°
2KB7 [68]	20	Phospholamban	25–53	22.9° ± 2.69°	26.2° ± 3.44°	3.66°	46.8°
2KQT [69]	17	Influenza A virus M2 proton channel	Chain A; 22–46	18.7° ± 0.108°	19.0° ± 4.38°	3.61°	16.7°

Angles are averaged over all NMR structures. For consistency, the average tilt angles for the experimental structures were recalculated using the same procedure as for the simulation results, which is described in the Methods section

[a] Simulations included side chain conformational sampling since the experimental structures only include backbone atom coordinates

[b] Average value ± standard deviation is shown

[c] Solution NMR structure (PDB entry 1A11) was used for simulations and the results compared with the solid state NMR structure (PDB entry 1CEK) because the latter has only backbone atom coordinates

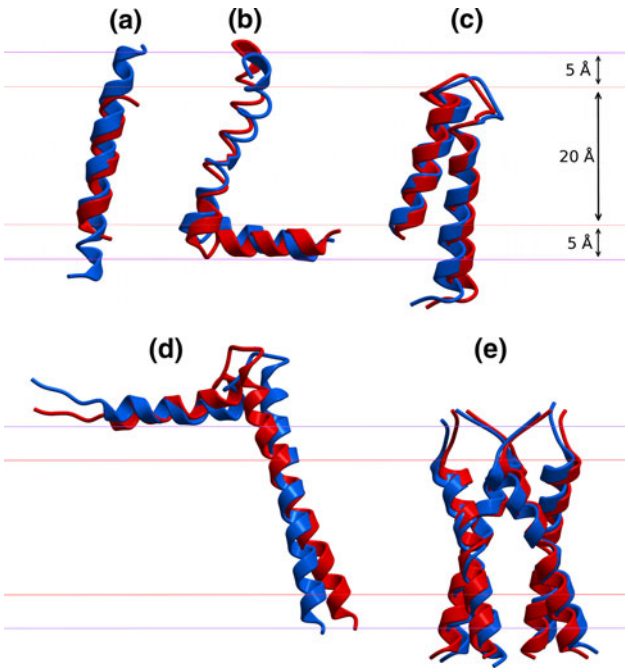

Fig. 2 Predicted orientation of proteins relative to the membrane, shown in *blue*, compared with the orientations measured by solid-state NMR experiments, shown in *red*. *Violet lines* denote the boundaries of the 30 Å thick membrane while the *red lines* denote the boundaries of the 20 Å thick core region. The thicknesses of the corresponding solvation regions are indicated above. Results for the following proteins are shown: **a** acetylcholine receptor M2, **b** fd bacteriophage coat protein, **c** MerF mercury transport protein, **d** phospholamban, and **e** influenza A virus M2 proton channel. Details on the proteins and simulation results are given in Table 3

$$\int_{\alpha_{exp}-\Delta\alpha}^{\alpha_{exp}+\Delta\alpha} d(\cos\chi) = \frac{1}{2}\left[\cos\left(\alpha_{exp}-\Delta\alpha\right) - \cos\left(\alpha_{exp}+\Delta\alpha\right)\right] \tag{3}$$

for an experimental tilt angle α_{exp} and tilt angle difference $\Delta\alpha$. The statistical significance varies from 0.71 to 9.6% for the predicted tilt angles and from 7.0 to 26% for the rotation angles. The upper significance value for the rotation angles is reduced to only 13% if the one outlier, phospholamban, is excluded.

Orientation and folding of amphipathic peptides

The position and orientations of five different amphipathic antimicrobial peptides relative to the membrane were also calculated using the same methods as for the transmembrane α helical peptides described in the last section. The experimental backbone structures of the peptides, as determined by NMR, were used and kept unchanged throughout the simulations. The results are summarized in Table 4. Because all structures were solved by solution NMR, the calculated and experimental orientations could not be compared. All helical peptides were found to bind at the membrane-solvent interface with tilt angles near 90°, or parallel to the membrane boundary. Furthermore, all structures were solved in detergent micelles and therefore are likely to be similar but not identical to the peptides' structures in lipid bilayers. The cecropin A-magainin 2 hybrid (PDB entry 1F0D) and the ovispirin-1 G10 mutant (PDB entry 1HU6) had the largest centroid distances from the membrane center due to non-helical hydrophilic termini that extend into the aqueous region in the lowest energy orientations.

We also attempted the more challenging task of predicting the structure, position, and orientation of the same amphipathic peptides but starting from a fully extended unfolded conformation. As shown in Fig. 3, all predicted structures, except for the cecrophin A-magainin 2 hybrid (1F0D), were predominantly helical and all were located at the membrane-solvent interface. The 1F0D peptide structure also had a helical segment that was in the interface and parallel to the membrane but the N-terminus had no regular secondary structure, *i.e.* is a coil. The 1F0D simulations had the highest deviation of lowest energies ($\sigma = 3.4$ kcal/mol) so that additional sampling may be needed for adequate convergence. We therefore ran additional simulations, for a total of 15, and did find a lower energy conformation with a qualitatively different structure, in which the N-terminal portion contained a short helical segment so that the peptide had a helix-coil-helix structure, also located at the interface. Interestingly, all 20 NMR structures have a

Table 4 Calculated orientation of amphipathic helices relative to the membrane

PDB entry	Number of NMR structures	Description	Helix residues	Average calculated tilt angle (α)	Average distance of the centroid from the membrane center (Å)
1F0D [70]	20	Cecropin A-Magainin 2 hybrid peptide	9–20	84.7° ± 36.2°	22.5
1HU5 [71]	20	Ovispirin-1	4–18	87.6° ± 10.6°	13.8
1HU6 [71]	20	Ovispirin-1 G10 mutant	4–11	81.2° ± 31.2°	21.0
1HU7 [71]	20	Novispirin T7 mutant	7–18	80.5° ± 27.6°	18.0
2KAM	20	Delta-hemolysin	1–26 (all)	98.7° ± 15.8°	17.1
2MAG [72]	10	Magainin	1–23 (all)	104.0° ± 22.0°	19.6

Angles and centroid distances are averaged over all NMR structures

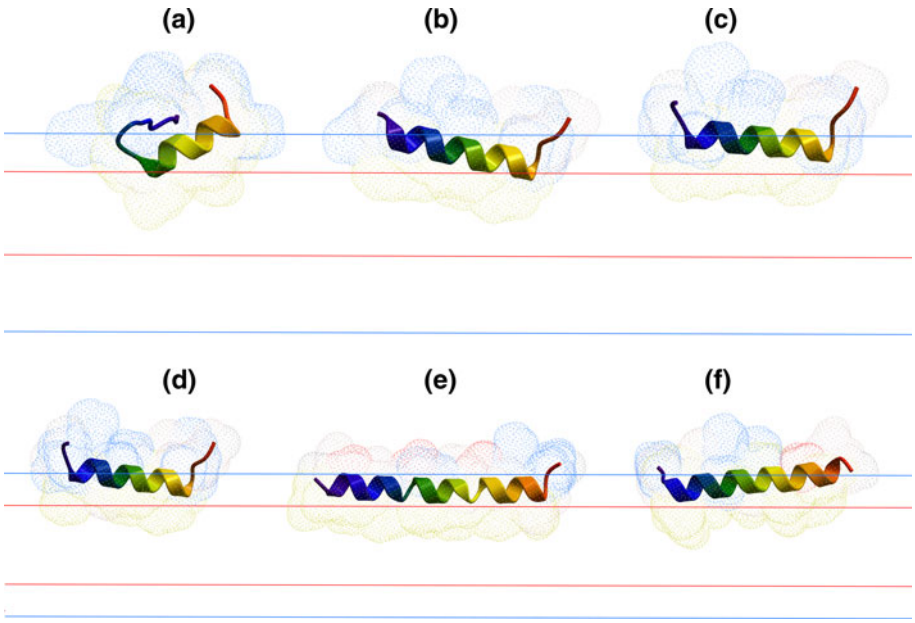

Fig. 3 Predicted structures of the amphipathic antibiotic peptides listed in Table 4. Note that the above structures are from ab initio folding simulations of fully flexible peptides, while the results in Table 4 are from simulations using rigid NMR backbone structures. The peptides are shown in ribbon representation with the color varying from blue at the N-terminus to red at the C-terminus. The solvent accessible surface of each peptide is also shown and colored according to the chemical properties of the corresponding residues: hydrophobic (*yellow*), uncharged polar (*pink*), negatively charged (*blue*), or positively charged (*red*). The peptides are **a** cecropin A-magainin 2 hybrid (1F0D), **b** ovispirin 1 (1HU5), **c** ovispirin-1 G10 mutant (1HU6), **d** novispirin T7 mutant (1HU7), **e** delta-hemolysin (2KAM), and **f** magainin (2MAG)

helical C-terminal segment with a coil segment in the middle and either a coil or a short helix in the N-terminal segment, in agreement with the simulation results. The presence of the structurally variable central coil in both the NMR and predicted structures is presumably due to the flexible linker segment, Gly-Ile-Gly, in the center of the peptide that disrupts the α helix. Furthermore, the slower convergence of the simulations for this peptide may be due to its flexibility. Figure 3 also shows that the charged groups, colored in red and blue, are facing the solvent while the hydrophobic groups, colored in yellow, are facing the membrane as expected. These results demonstrate that the energy function is accurate enough to discriminate between the native-like conformation and the many non-native conformations, which are incorrectly folded or incorrectly positioned relative to the membrane.

Helix-forming propensities of different amino acids in a lipid bilayer

An interesting series of experiments by Liu and Deber [24] examined the helical propensities of different residues both in an aqueous buffer and in a non-polar solvent (n-butanol) chosen to mimic the membrane environment. A "host–guest" approach was used in which guest residues (denoted by X) at three positions in the host peptide (KKAAAX AAAAAXAAWAAXAAAKKKK-amide) were substituted with all 20 natural amino acids. The lysine residues at each end of the peptide reduced overall hydrophobicity, and thus aggregation that would interfere with purification. The tryptophan residue in the center provided a fluorescent probe to monitor the local solvent environment.

We performed simulations of these peptides in both pure aqueous and membrane core environments and characterized their helicities by the number of central residues (19 residues excluding terminal Lys residues) with an α-helical backbone conformation. The DSSP algorithm [42] was used to determine the secondary structure for these results and also for other structures analyzed in this study. The energy optimized peptide structures were then compared with the corresponding helicities measured by Liu and Deber using circular dichroism spectroscopy. These results are presented in Table 5. They show that the numbers of helix residues was significantly correlated with the measured helicities for the membrane core but not for water (Kendall τ test at 5% significance). It is not clear why the aqueous ASPs did not reproduce the observed trends in helicities for peptides in water.

We also ran similar simulations of 20 residue homopolypeptides containing only alanine, glycine, isoleucine, leucine, or valine residues in both hydrophobic and aqueous environments. The results are summarized in Table 6. Poly-alanine was found to form a partial α helix of 6 residues in water, giving an average helical content of 42%.

Table 5 Alpha-helical content of Liu and Deber [24] host–guest peptides (KKAAAXAAAAAXAAWAAXAAAKKKK-amide) with X indicating each of the 20 naturally occurring amino acids

Residue	Membrane core			Water		
	Number of helix residues[a]	Helicity in n-butanol[b]	Lowest energy (kcal/mol)	Number of helix residues[a]	Helicity in aqueous buffer[b]	Lowest energy (kcal/mol)
A	19.0	3.72	−165 ± 0.0208	14.0	2.06	−234 ± 2.18
C	19.0	3.28	−169 ± 0.0569	17.7	0.91	−237 ± 1.82
D	7.7	2.66	−197 ± 4.70	10.0	1.47	−282 ± 0.378
E	6.3	2.54	−195 ± 4.03	17.0	2.25	−288 ± 0.234
F	15.7	3.78	−183 ± 0.234	10.0	1.12	−249 ± 1.84
G	9.7	3.44	−164 ± 2.18	6.67	0.34	−239 ± 1.85
H	16.3	2.90	−187 ± 0.0436	14.0	0.79	−273 ± 0.974
I	19.0	3.88	−159 ± 0.0200	18.3	1.54	−225 ± 0.958
K	13.0	2.65	−171 ± 0.985	9.0	1.12	−278 ± 0.121
L	19.0	3.84	−169 ± 0.0173	12.3	1.80	−236 ± 0.214
M	19.0	3.67	−170 ± 0.113	13.3	1.33	−244 ± 0.0751
N	8.0	2.82	−183 ± 1.05	9.7	0.59	−273 ± 0.720
P	4.7	1.70	−164 ± 1.09	4.0	0.0	−230 ± 2.17
Q	13.3	2.87	−183 ± 0.997	19.0	1.19	−272 ± 0.215
R	19.0	2.84	−191 ± 0.808	18.3	1.35	326 ± 0.307
S	16.7	2.99	−176 ± 0.264	19.0	0.79	−252 ± 0.259
T	19.0	3.27	−175 ± 0.0700	19.0	0.90	−248 ± 0.110
V	17.7	3.82	−160 ± 0.176	7.7	1.03	−231 ± 2.26
W	19.0	3.20	−197 ± 0.0404	9.0	1.13	−272 ± 0.230
Y	16.3	3.33	−181 ± 1.44	13.7	1.16	−260 ± 0.240

The mean and sample standard deviation for the lowest energies achieved in each simulation are shown

[a] Average number of residues in the longest α-helical segment within the central 19-residue portion of the peptide, excluding the terminal lysine residues. Residues with an α-helix backbone conformation were determined by the DSSP algorithm [42]

[b] Helicity measurements ($\theta_{222\ nm}$) from Ref. [24]

Table 6 Lowest energy structures for 20-mer poly-X peptides with X = (alanine, glycine, isoleucine, leucine, or valine) in both water and the membrane core

Peptide residues	Water			Membrane core		
	Lowest energy (kcal/mol)	Lowest energy structure	Average helicity (%)[a]	Lowest energy (kcal/mol)	Lowest energy structure	Average helicity (%)[a]
A	−102 ± 0.475	Partial α helix	42	−116 ± 0.0289	α helix	90
G	−118 ± 1.23	Random coil	0	−109 ± 2.03	π helix	0
I	−48.2 ± 1.94	Small 3-strand β sheet	0	−76.0 ± 0.0231	α helix	90
L	−114 ± 0.932	α helices connected by loop	77	−148 ± 0.0493	α helix	90
V	−91.1 ± 0.948	2-strand antiparallel β sheet	0	107 ± 2.59	2-strand antiparallel β sheet	0

[a] Average over the lowest energy structures from three independent simulations of the percentage of peptide residues occurring in an alpha helix as determined by DSSP [42]

Different studies have found conflicting values of the helix-forming propensity of alanine in water. One study by Ingwall et al. [43] found that short 10 residue poly-alanine peptides were non-helical, although much longer peptides with 160–1,000 residues were mostly helical. The parameters that they determined for the Zimm-Bragg model [44] of helix formation at ambient temperature, s = 1.05 and $\sigma = 8 \times 10^{-4}$, yields a low average helicity of only 8.6% for the poly-alanine 20-mer examined here. In contrast, an experimental study by Chakrabartty et al. [45] found that alanine is a strong helix former in water. A calculation using the parameters that they determined for a Lifson-Roig model [46], modified to include N-capping parameters, yields a considerably higher average helical content of

69% for the poly-alanine 20-mer. There are two possible reasons for the different conclusions from these two experiments: (1) each examined different peptides; the former examined lysine containing block copolymers with 10–1,000 alanine residues while the latter examined shorter host–guest peptides with residue substitutions at defined sites and (2) they fit the experimental data using different theoretical models of helix-coil transitions. Since these comparisons with experiment were made through intermediate theoretical models we also ran simulations for three of the peptides tested in the Chakrabartty et al. study, Ac-YGKA$_4$KA$_4$KA$_4$K-CONH$_2$, Ac-YGGKA$_4$KA$_4$KA$_4$K-CONH$_2$, and Ac-YGGGKA$_4$KA$_4$KA$_4$K-CONH$_2$. We found that these peptides had average helicities of 42.6, 54.4, and 41.7% compared with the experimental values of 73.7, 67.7, and 61.6%. Taken together, these results show that the simulations yielded helical content for the poly-alanine peptide in water that is somewhat lower than experimental results from the later study of Chakrabartty et al., but in qualitative agreement. As Chakrabartty et al. found that only alanine is a strong helix former, the low predicted helical content of the poly-Gly, Ile, and Val peptides are in agreement, but the high helical content of poly-Leu is not.

In the hydrophobic membrane environment we found that three of the peptides, poly-Ala, Ile, and Leu, were almost entirely helical whereas two peptides, poly-Gly and poly-Val contained no helical segments. This is consistent with the rank order of helical propensities found by Li and Deber [47] for host–guest peptides in a membrane, namely Ala $\approx$ Leu $<$ Ile $<$ Val $<$ Gly. The predicted structures are also consistent with the results of Monera et al. [48], who found similar helical propensities for residues substituted on the hydrophobic face of amphipathic peptides.

Interestingly, the simulations showed that the polyvaline 20-mer formed a β hairpin both in water and in the pure hydrophobic environment. Additional simulations of the KV$_{20}$K peptide likewise showed that it formed a β hairpin with the charged N- and C-termini in solution. This is consistent with the fact that the beta-branched amino residues valine and isoleucine have among the highest β strand propensities in water as assessed by both occurrence statistics in proteins [49] and experimental measurements in host–guest systems [50–52]. This can be attributed to steric and entropic contribution from the bulky side chains of these residues that disfavors an α helical conformation in which nearby side chains are closely packed and favors a β strand conformation in which the side chains extend outwards in alternating directions. Interestingly, our results differ from those of Efremov et al. [21], who found that a poly-valine peptide formed an α helix in the hydrophobic membrane environment and a random coil in water with their solvation model. We did however find that valine leads to an α helix in the context of the Liu-Deber host–guest peptide, in which it is flanked by alanine residues and so is less conformationally constrained. Such neighboring residue interactions are not directly accounted for in simplified theoretical models of the helix-coil transitions, such as the Lifson-Roig model [46], although their relative importance remains to be determined. The large fraction of alanine residues, which favor helix formation, in the Liu-Deber peptides also likely contributes to the significant helicity of the valine-substituted peptide. We also found that poly-isoleucine peptides, containing another β branched amino acid, also formed predominantly β strand structures in solution but, unlike poly-leucine, formed an α helix in the membrane core. Poly-glycine formed an unusual π helix in the membrane. Although it is unlikely that an isolated poly-glycine peptide would actually forms a stable extended π helix in membranes, it is plausible that this may be a low energy conformation because of the energetically favorable backbone hydrogen bonds, compact structure that somewhat reduces the exposure of the hydrophilic backbone, and the lack of steric clash from a side chain. Furthermore, π helices are quite rare in globular proteins [53], but are considerably more prevalent in transmembrane segments of membrane proteins [54].

Assembly of short peptides into beta strand dimers within the membrane

A paper by Wimley et al. [25] describes the experimental characterization of a short hexameric peptide, acetyl-WLLLLL, that was found to aggregate into antiparallel β sheets in phosphatidylcholine (POPC) membranes. A subsequent paper [26] examined POPC membrane binding and folding for a series of related host–guest peptides, acetyl-WLLXLL-OH, in which X is one of the 20 natural amino acids. Studies of these model host–guest peptides can help in elucidating the energetic factors driving β sheet formation in membranes, which is relevant for the folding of β barrel membrane proteins and also potentially for the formation of amyloids linked to human diseases [55, 56].

In order to test the membrane solvation model and gain insight into the folding of these peptides, we have run Monte Carlo simulations of all 20 peptides with neutral C-termini. These can be directly compared with the experimental results obtained at pH 2.5, in which all peptides had significant membrane binding. The simulation results are summarized in Table 7. Overall, the peptides found to form β sheets in the experiments, namely with X = C, F, I, L, M, and V, were all found to form β sheets. Indeed, all except the cysteine peptide had the maximum number (6) of interstrand hydrogen bonds and backbone torsion angles that were the closest to typical values for β sheets ($\varphi = -120°$ and $\psi = 115°$ for parallel or $\varphi = -140°$ and $\psi = 135°$ for antiparallel sheets [57]). Likewise, peptides with X = A and

Table 7 Lowest energy structures for dimers of the hexameric host–guest peptides, acetyl-WLLXLL with X = each of the 20 natural amino acids, studied by Bishop and Wimley [26]

X-residue in AcWLLXLL	Lowest energy structure	Lowest energy (kcal/mol)	Number of H-bonds in β sheet	Backbone angles RMSD (degrees)
β sheet forming				
C	Parallel β sheet	-121 ± 2.10	3	58.6
F	Antiparallel β sheet	-126 ± 1.40	6	41.5
I	Antiparallel β sheet	-113 ± 1.16	6	42.8
L	Antiparallel β sheet	-118 ± 1.79	6	37.5
M	Parallel β sheet	-122 ± 0.640	6	30.2
V	Antiparallel β sheet	-116 ± 2.00	6	35.4
Marginal β sheet forming				
A	Bent antiparallel β sheet	-115 ± 2.06	6	40.7
W	Partial parallel β sheet	-137 ± 2.48	3	70.2
Random coil				
D	Bent antiparallel β sheet	-144 ± 2.62	4	44.5
E	Random coil	-147 ± 0.866	NA	NA
G	Random coil	-114 ± 0.755	NA	NA
H	Random coil	-136 ± 1.27	NA	NA
K	Partial parallel β sheet	-145 ± 1.50	2	58.6
N	Random coil	-136 ± 0.767	NA	NA
P	Partial parallel β sheet	-113 ± 0.728	3	NA
Q	Random coil	-135 ± 1.52	NA	NA
R	Random coil	-176 ± 1.11	NA	NA
S	Partial parallel β sheet	-124 ± 2.07	3	60.0
T	Parallel β sheet	-124 ± 1.62	5	57.8
Y	Random coil	-131 ± 1.71	NA	NA

The peptides are divided into three groups depending on whether they were found to mostly form β sheets, only marginally form β sheets, or form random coils in membranes in that study. The last column shows the RMSD in backbone angles from the typical β sheet values ($\varphi = -120°$ and $\psi = 115°$ for parallel or $\varphi = -140°$ and $\psi = 135°$ for antiparallel sheets). The mean and sample standard deviation for the lowest energies achieved in each simulation are shown

W, which showed a more modest propensity to form β sheets in the experiments, also formed stable β sheets with 6 and 3 interstrand hydrogen bonds respectively. Finally, almost all of the peptides that formed random coils in the experiments, with X = D, E, G, H, K, N, P, Q, R, S, T, and Y, either formed random coils or, in some cases, shorter and consequently less stable β sheets with only 2–4 interstrand hydrogen bonds. The exception was threonine, which formed a parallel β sheet stabilized by 5 hydrogen bonds. Except for a few cases, these general results demonstrate that molecular mechanics optimization using the implicit solvation model accurately reproduces the non-covalent association and overall conformational preferences of the acetyl-WLLXLL-OH peptides studied by Bishop et al. [26]. Furthermore the peptides with high β sheet propensities likely form larger structures than the dimers simulated here, with >2 peptides, as was found for acetyl-WLLLLL [25]. Simulations with more subunits would be considerably more computationally expensive, and are expected to largely agree with the dimer results. This is because dimers with near-ideal β sheet geometries are consistent with larger β sheet structures whereas those with distorted structures, such as the aspartic acid peptide dimer with only a short bent β strand segment, would not provide a good nucleus for extending the β sheet as more subunits bind.

The original peptide studied, acetyl-WLLLLL, was found to form an antiparallel β sheet at the membrane interface and to be oriented perpendicular to the membrane, in agreement with the experimental results of [25]. Importantly, the assembly of the two β strands into an antiparallel β sheet is not completely longitudinally symmetric so that the dimer has a preferred orientation direction relative to the membrane. This can be seen by the unique global minimum in the solvation energy at the position and orientation of the calculated structure, as shown in Fig. 4. It is also interesting that although the acetyl-WLLLLL peptide formed an antiparallel β sheet, as observed in the experiment, several of the other peptides instead formed parallel β sheets in the simulations, thus suggesting more heterogeneity in the β sheet structures for this series of host–guest peptides. The relative orientations of the strands for these other peptides were not measured and so await experimental validation.

The solvation energy for the tryptophan residue in acetyl-WLLLLL makes a large contribution to the perpendicular orientational preference of this peptide (data not shown). The tryptophan residue nearest to the membrane-solvent interface in the lowest energy structure is oriented with the relatively hydrophilic nitrogen atom towards the aqueous solvent and the relatively hydrophobic aromatic

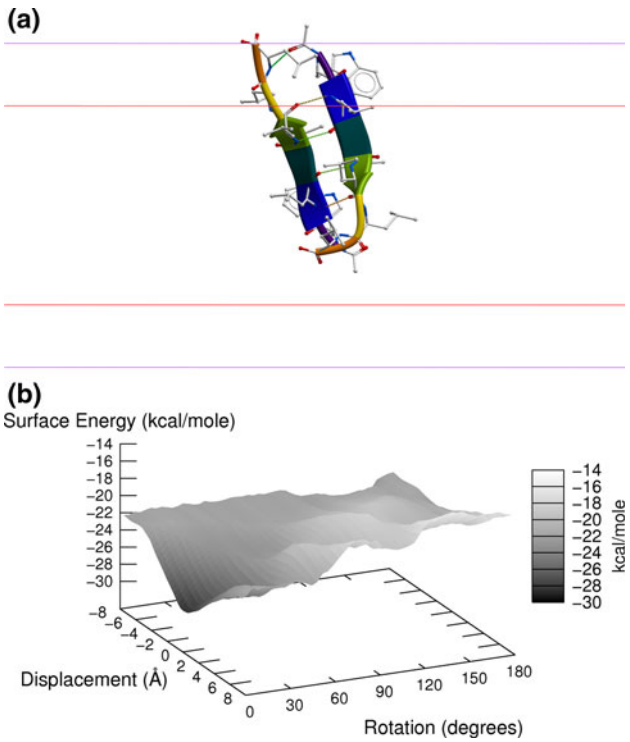

Fig. 4 **a** The lowest energy structure for the acetyl-WLLLLL peptide dimer and **b** a three-dimensional plot of the solvation energy for the dimer as a function of displacement along the membrane normal axis and rotation angle relative to the normal axis. This shows the global energy minimum is at the depth and orientation found in the simulation. The predicted conformation, an antiparallel β strand located near the membrane-solvent interface and oriented approximately perpendicular to the membrane, is consistent with the experimental results obtained by Wimley et al. [25]

carbon atoms towards the membrane. A simulation of an analog of the tryptophan side chain, indole, showed that it is preferentially in the interface with the same orientation. Likewise, an analog of the tyrosine side chain, p-cresol, also localizes to the interface with the hydrophilic hydroxyl group pointing towards the aqueous solvent region. Both the apparent insertion free energies of Hessa et al. [27] and statistical residue potentials derived from the distribution of residue types in membrane protein structures [9] indicate that tryptophan and tyrosine have the most favorable solvation free energy in the interface near the lipid head groups. The amphipathic nature of the side chains for these residues combined with their rigidity contributes to their interface localization and preferred orientation. Favorable specific interactions between these aromatic side chains and the lipid head groups that are mediated by cation-π, hydrogen bonding, and dipole interactions also likely contribute but are not explicitly accounted for in the present solvation model. Experimental data on the orientation of indole in lipid bilayers are inconclusive with, for example, an NMR study [58] confirming the perpendicular orientation observed in our simulations while linear dichroism spectroscopy measurements indicated an orientation that is closer to parallel to the membrane boundary [59]. Finally, the former NMR study found that the lipid composition of the bilayer somewhat affected the positional and orientational preferences of indole relative to the membrane, which is an effect that would be missed by general membrane solvation models.

In order to further understand the conformational preferences of the β sheet pairing, the energy components of parallel β sheet structures for the leucine and valine cases were compared to the lowest energy antiparallel structures, which were approximately 2.8 and 2.2 kcal/mol lower in energy, respectively. In both cases, lower total intrastrand energy favors the antiparallel structure. In fact, the parallel structure considered for the valine substituted peptide has seven hydrogen bonds, one more than the anti-parallel structure; and the interstrand binding energy alone would have favored the parallel structure. Chou, Nemethy, and Scheraga [60] have reported calculations on unsolvated six residue poly-leucine β sheet structures that also showed intrastrand energy as a dominant factor in favoring antiparallel structures over parallel ones. In contrast, calculations in that report for short poly-alanine β sheets showed a preference for antiparallel binding favored by lower interstrand energy.

Comparison with experimental membrane insertion free energies

Since the ASPs were determined using experimental data for a simple model system, namely the transfer of amino acid side chain analogs from water into cyclohexane, we next examined how well the new solvation model agrees with data from two different experiments. One experiment, by Hessa et al. [27], derived the apparent free energy of inserting different peptide fragments into a lipid bilayer by measuring the fraction of membrane-integrated to non-integrated peptides resulting from Sec61 translocon-mediated membrane insertion of peptides forming α helices within the membrane of intact cells. Their experimental setup closely follows the actual biological process in which newly synthesized integral membrane proteins are integrated into the lipid bilayer by translocons, which are large molecular machines. The other experiment, by Moon and Fleming [28], measured the free energies of the reversible combined process of folding and membrane insertion for a native β barrel protein (OmpLA) and also for mutants with residue substitutions at a site that faces the lipid and is located near the central plane of the membrane. Figure 5 shows plots of calculated versus experimental transfer free energies for these two experiments.

Hessa et al. combined the transfer energies from many different peptides in order to derive an apparent insertion free energy for each residue that depends on its location within the membrane, which they denoted as $\Delta G_{app}^{aa(i)}$, where i is the amino acid's sequence position in the transmembrane helix. We calculated water to membrane transfer energies for different uncharged residues substituted into the center of a poly-alanine helical peptide and compared the results with the biological apparent free energy for the center of the membrane $\Delta G_{app}^{aa(0)}$ from Hessa et al. The results, plotted in Fig. 5(a), show the high correlation between the calculated and experimental values ($r = 0.93$). We note that despite the high correlation, the magnitudes of the Hessa et al. apparent free energies were lower than the values calculated with the solvation model.

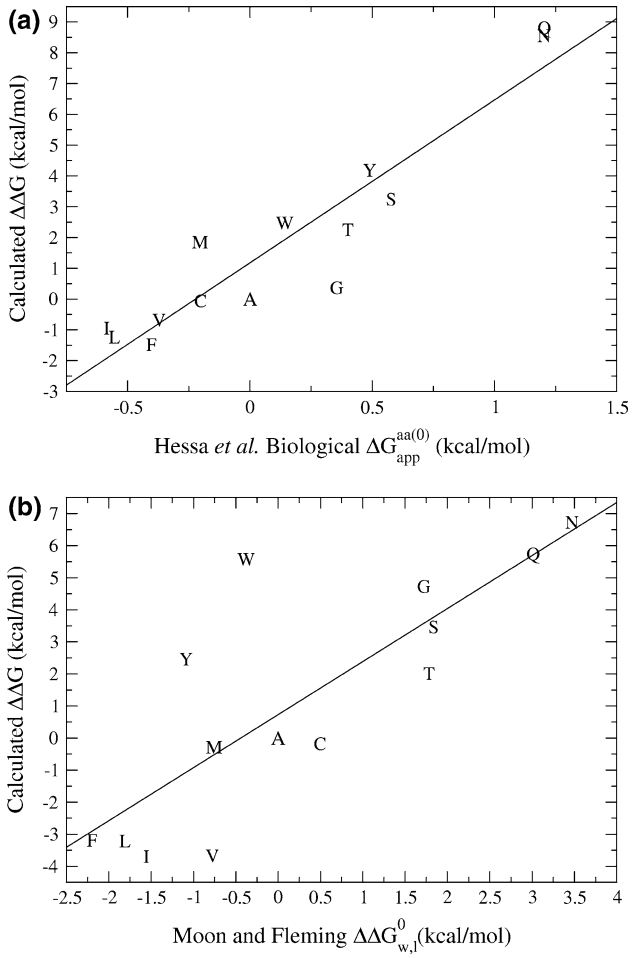

Fig. 5 Calculated free energies for transferring different uncharged residues from water to the membrane central hydrophobic core versus experimental values from Hessa et al. [27] in (**a**) and from Moon and Fleming [28] in (**b**). All values are relative to alanine. The $\Delta\Delta G_{app}^{aa(0)}$ values in (**a**) were obtained from Supplementary Table S2 of Ref. [27] while the $\Delta\Delta G_{w,l}^{0}$ values in (**b**) are from Supplementary Table S1 of Ref. [28]. The linear regression fits to the data are also shown in each plot

This is likely due to the fact that the Hessa et al. free energies represent the transfer of peptides from the translocon channel into the membrane while the calculated free energies are for transfer from water into the membrane. If the interior of the translocon channel is considered to have hydrophobicity that is intermediate between that of water and the membrane core then this difference would explain the lower magnitude biological apparent free energy values.

We did not perform calculations for charged residues because their apparent translocon-mediated insertion free energies are significantly lower than what one would expect for simple transfer from aqueous solvent into a hydrophobic environment, which is dominated by an unfavorable electrostatic energy penalty. Two computational studies [61, 62] have attributed this difference to both interactions with surrounding proteins in the membrane as well as the presence of water in the central aliphatic region of the lipid bilayer, which solvates the charged residue and thus reduces the insertion free energy penalty. Also, we did not examine the dependence of the transfer free energies with the residue's position within the membrane. Hessa et al. did fit the position dependence but used a simple functional form (single Gaussian for all residues except Trp and Tyr), presumably due to the uncertainty in the position dependence from, for example, differences in the positions of the TM peptides relative to the membrane. The simple spatial dependence of the solvation free energy in our model is qualitatively similar to the dependence in Hessa et al. for non-aromatic residues (i.e. excluding Trp and Tyr). Namely, it has maximum magnitude in the center of the membrane and then smoothly decreases to zero at the membrane-aqueous solvent interface around $|z| = 15$ Å. The width of the central region in which the residue of interest is fully within the hydrophobic core will be smaller than 10 Å for polar or charged residues due to snorkeling, in which the hydrophilic side chain will bend towards the aqueous solvent. Reproducing the spatial dependence of the solvation free energy for aromatic residues would require accounting for interactions with the lipid head groups, which is not included in the present model.

The correlation ($r = 0.80$) between the calculated and Moon and Fleming experimental values was lower than for Hessa et al. but still significant. Interestingly, the two largest outliers were Trp and Tyr, which are also the residues most preferentially occurring in the membrane-solvent interface region [63], presumably due to energetically favorable interactions. This could be explained by W210 and Y210 in the corresponding OmpLA mutants interacting with the lipid head groups. However, this seems unlikely given that residue 210 is thought to reside near the central plane of the membrane, so that a large perturbation of the

membrane or the position of OmpLA relative to it, possibly in combination with favorable side chain conformations (i.e. snorkeling), would be required for such interactions. In addition, the range of the calculated free energies was higher than for the experimental values but the discrepancy was not as large as for the Hessa et al. data. The presence of water molecules in the central lipid tail region of the membrane, discussed above, could explain this difference.

We also examined the contribution of the protein backbone to the water-membrane transfer free energy. Because glycine has no non-hydrogen atoms in its side chain, this contribution was estimated by calculating the solvation energy difference between the hydrophobic membrane core and water for a series of poly-glycine peptides one to five residues in length. Linear regression analysis was then used to determine the contribution of each residue, or backbone segment, to the transfer free energy. We obtained a value of 1.16 kcal/mol for backbone transfer. This is consistent with the value of 1.2 ± 0.1 kcal/mol inferred by Wimley and White from the experimental transfer free energies for short peptides from water into the membrane interface [64]. Because the peptides remained unstructured in both water and the membrane, the experimental value reflected the transfer free for the backbone without the contribution from hydrogen bond formation.

Conclusions and outlook

The atomic solvation model for membranes derived in this work was shown to reproduce key experimental results on the orientation, folding, energetics, and association of peptides in a lipid bilayer. Taken together these results suggest that the simplifying approximations made in the solvation model still allow it to adequately describe the energetics of peptides in the membrane. Importantly, the model was also computationally efficient so that it allowed us to run ab initio folding simulations of peptides using relatively modest computer resources.

Several future extensions of the solvation model are possible. First, another region representing the interface portion of the membrane can be added. This portion of the lipid bilayer is composed of solvated lipid head groups that appear to form energetically favorable interactions with the aromatic residues tryptophan and tyrosine, as deduced from both physical measurements [27] and the occurrence statistics obtained from membrane proteins with available structures [9]. Another possible extension would be to account for aqueous solvation regions within transmembrane pores in channel and transporter membrane proteins. Regions for either aqueous or membrane solvation with spherical, ellipsoidal, cylindrical, or rectilinear shapes have been implemented in the ICM program and can be used for this purpose. In addition, it may be possible to utilize additional statistical information from the relative occurrence frequencies of different atom types and their surface accessibilities observed in available membrane protein structures, which is independent of the gas→cyclohexane transfer free energies, to increase the quantity of data used to fit the ASPs. Finally, further molecular mechanics simulations on the folding and association of peptides and proteins within the membrane using this solvation model will help further validate its applicability and provide useful structure predictions for future experimental study.

Acknowledgments We thank Dr. Totrov for helpful discussions. This work was funded by the Mayo Clinic.

References

1. Wallin E, von Heijne G (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein Sci 7((4):1029–1038
2. Yildirim MA et al (2007) Drug-target network. Nat Biotechnol 25(10):1119–1126
3. Lacapere JJ et al (2007) Determining membrane protein structures: still a challenge!. Trends Biochem Sci 32((6):259–270
4. Love J et al (2010) The New York consortium on membrane protein structure (NYCOMPS): a high-throughput platform for structural genomics of integral membrane proteins. J Struct Funct Genomics 11(3):191–199
5. Lindahl E, Sansom MS (2008) Membrane proteins: molecular dynamics simulations. Curr Opin Struct Biol 18(4):425–431
6. Marrink SJ, de Vries AH, Mark AE (2004) Coarse grained model for semiquantitative lipid simulations. J Phys Chem B 108(2): 750–760
7. Bond PJ et al (2007) Coarse-grained molecular dynamics simulations of membrane proteins and peptides. J Struct Biol 157(3): 593–605
8. Ulmschneider MB, Sansom MS, Di Nola A (2005) Properties of integral membrane protein structures: derivation of an implicit membrane potential. Proteins 59(2):252–265
9. Senes A et al (2007) E(z), a depth-dependent potential for assessing the energies of insertion of amino acid side-chains into membranes: derivation and applications to determining the orientation of transmembrane and interfacial helices. J Mol Biol 366(2):436–448
10. Spassov VZ, Yan L, Szalma S (2002) Introducing an implicit membrane in generalized Born/solvent accessible continuum solvent models. J Phys Chem B 106(34):8726–8738
11. Tanizaki S, Feig M (2005) A generalized Born formalism for heterogeneous dielectric environments: application to the implicit modeling of biological membranes. J Chem Phys 122(12):124706
12. Im W, Feig M, Brooks CL III (2003) An implicit membrane generalized born theory for the study of structure, stability, and interactions of membrane proteins. Biophys J 85(5):2900–2918
13. Ulmschneider MB et al (2007) A generalized born implicit-membrane representation compared to experimental insertion free energies. Biophys J 92(7):2338–2349
14. Lazaridis T (2003) Effective energy function for proteins in lipid membranes. Proteins 52(2):176–192
15. Lazaridis T, Karplus M (1999) Effective energy function for proteins in solution. Proteins 35(2):133–152

16. Wesson L, Eisenberg D (1992) Atomic solvation parameters applied to molecular dynamics of proteins in solution. Protein Sci 1(2):227–235
17. Eisenberg D, McLachlan AD (1986) Solvation energy in protein folding and binding. Nature 319(6050):199–203
18. Ooi T et al (1987) Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proc Natl Acad Sci USA 84(10):3086–3090
19. Cummings MD, Hart TN, Read RJ (1995) Atomic solvation parameters in the analysis of protein–protein docking results. Protein Sci 4(10):2087–2099
20. Schiffer CA et al (1993) Protein structure prediction with a combined solvation free energy-molecular mechanics force field. Molecular Simulation 10(2):121–149
21. Efremov RG et al (1999) A solvent model for simulations of peptides in bilayers. I. Membrane-promoting alpha-helix formation. Biophys J 76(5):2448–2459
22. Lomize AL et al (2006) Positioning of proteins in membranes: a computational approach. Protein Sci 15(6):1318–1333
23. Hong M (2007) Structure, topology, and dynamics of membrane peptides and proteins from solid-state NMR spectroscopy. J Phys Chem B 111(35):10340–10351
24. Liu LP, Deber CM (1998) Uncoupling hydrophobicity and helicity in transmembrane segments. Alpha-helical propensities of the amino acids in non-polar environments. J Biol Chem 273(37): 23645–23648
25. Wimley WC et al (1998) Folding of beta-sheet membrane proteins: a hydrophobic hexapeptide model. J Mol Biol 277(5): 1091–1110
26. Bishop CM, Walkenhorst WF, Wimley WC (2001) Folding of beta-sheets in membranes: specificity and promiscuity in peptide model systems. J Mol Biol 309(4):975–988
27. Hessa T et al (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature 450(7172): 1026–1030
28. Moon CP, Fleming KG (2011) From the cover: side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. Proc Natl Acad Sci USA 108(25): 10174–10177
29. Abagyan R, Totrov M (1994) Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. J Mol Biol 235(3):983–1002
30. Abagyan R, Totrov M, Kuznetsov D (1994) ICM—A new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. J Comput Chem 15(5):488–506
31. Momany FA et al (1975) Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, ydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. J Phys Chem 79:2361–2381
32. Nemethy GN, Pottle MS, Scheraga HA (1983) Energy parameters in polypeptides. 9. Updating of gemetrical parameters, nonbonded interactions and hydrogen bond interactions for the naturally occurring amino acids. J Phys Chem 87:1883–1887
33. Nemethy GN et al (1992) Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. J Phys Chem 96:6472–6484
34. Fernandez-Recio J, Totrov M, Abagyan R (2002) Soft protein–protein docking in internal coordinates. Protein Sci 11(2): 280–291
35. Abagyan R (1997) Protein structure prediction by global energy optimization. In: van Gunsteren WF, Weiner PK, Wilkinson AJ (eds) Computer simulation of biomolecular systems: theoretical and experimental applications. Kluwer Academic, Norwell, pp 363–394
36. Wolfenden R et al (1981) Affinities of amino acid side chains for solvent water. Biochemistry 20(4):849–855
37. Sharp KA et al (1991) Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. Biochemistry 30(40):9686–9697
38. Hastie T, Tibshirani R, Friedman J (2001) In: The elements of statistical learning. Springer, New York, p 461–465
39. Snijder HJ et al (1999) Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. Nature 401(6754):717–721
40. Wiener MC, White SH (1992) Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data. III. Complete structure. Biophys J 61(2):434–447
41. Phillips R et al (2009) Emerging roles for lipids in shaping membrane-protein function. Nature 459(7245):379–385
42. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22(12):2577–2637
43. Ingwall RT et al (1968) Conformational studies of poly-L-alanine in water. Biopolymers 6(3):331–368
44. Zimm BH, Bragg JK (1959) Theory of the phase transition between helix and random coil in polypeptide chains. J Chem Phys 31(2):526–531
45. Chakrabartty A, Kortemme T, Baldwin RL (1994) Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. Protein Sci 3(5):843–852
46. Lifson S, Roig A (1961) On the theory of helix-coil transition in polypeptides. J Chem Phys 34(6):1963–1974
47. Li SC, Deber CM (1994) A measure of helical propensity for amino acids in membrane environments. Nat Struct Biol 1(6): 368–373
48. Monera OD et al (1995) Relationship of sidechain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix. J Pept Sci 1(5):319–329
49. Chou PY, F. G.D. (1974) Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. Biochemistry 13(2):211–222
50. Kim CA, Berg JM (1993) Thermodynamic beta-sheet propensities measured using a zinc-finger host peptide. Nature 362(6417): 267–270
51. Minor DL Jr, Kim PS (1994) Measurement of the beta-sheet-forming propensities of amino acids. Nature 367(6464):660–663
52. Minor DL Jr, Kim PS (1994) Context is a major determinant of beta-sheet propensity. Nature 371(6494):264–267
53. Barlow DJ, Thornton JM (1988) Helix geometry in proteins. J Mol Biol 201(3):601–619
54. Riek RP et al (2001) Non-alpha-helical elements modulate polytopic membrane protein architecture. J Mol Biol 306(2): 349–362
55. Cerf E et al (2009) Antiparallel beta-sheet: a signature structure of the oligomeric amyloid beta-peptide. Biochem J 421(3):415–423
56. Butterfield SM, Lashuel HA (2010) Amyloidogenic protein-membrane interactions: mechanistic insight from model systems. Angew Chem 49(33):5628–5654
57. IUPAC-IUB Commision (1970) IUPAC-IUB commission on biochemical nomenclature. Abbreviations and symbols for the description of the conformation of polypeptide chains. Tentative rules (1969). Biochemistry 9(18):3471–3479
58. Gaede HC, Yau WM, Gawrisch K (2005) Electrostatic contributions to indole-lipid interactions. J Phys Chem B 109(26): 13014–13023

59. Esbjorner EK et al (2007) Tryptophan orientation in model lipid membranes. Biochem Biophys Res Commun 361(3):645–650
60. Chou KC, Nemethy G, Scheraga HA (1983) Effect of amino acid composition on the twist and the relative stability of parallel and antiparallel β-sheets. Biochemistry 22(26):6213–6221
61. Johansson AC, Lindahl E (2009) Protein contents in biological membranes can explain abnormal solvation of charged and polar residues. Proc Natl Acad Sci USA 106(37):15684–15689
62. Rychkova A, Vicatos S, Warshel A (2010) On the energetics of translocon-assisted insertion of charged transmembrane helices into membranes. Proc Natl Acad Sci USA 107(41):17598–17603
63. Yau WM et al (1998) The preference of tryptophan for membrane interfaces. Biochemistry 37(42):14713–14718
64. Wimley WC, White SH (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat Struct Biol 3(10):842–848
65. Opella SJ et al (1999) Structures of the M2 channel-lining segments from nicotinic acetylcholine and NMDA receptors by NMR spectroscopy. Nat Struct Biol 6(4):374–379
66. Marassi FM, Opella SJ (2003) Simultaneous assignment and structure determination of a membrane protein from NMR orientational restraints. Protein Sci 12(3):403–411
67. De Angelis AA et al (2006) Structure determination of a membrane protein with two trans-membrane helices in aligned phospholipid bicelles by solid-state NMR spectroscopy. J Am Chem Soc 128(37):12256–12267
68. Traaseth NJ et al (2009) Structure and topology of monomeric phospholamban in lipid membranes determined by a hybrid solution and solid-state NMR approach. Proc Natl Acad Sci USA 106(25):10165–10170
69. Cady SD et al (2010) Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers. Nature 463(7281):689–692
70. Oh D et al (2000) Role of the hinge region and the tryptophan residue in the synthetic antimicrobial peptides, cecropin A(1–8)-magainin 2(1–12) and its analogues, on their antibiotic activities and structures. Biochemistry 39(39):11855–11864
71. Sawai MV et al (2002) Impact of single-residue mutations on the structure and function of ovispirin/novispirin antimicrobial peptides. Protein Eng 15(3):225–232
72. Gesell J, Zasloff M, Opella SJ (1997) Two-dimensional 1H NMR experiments show that the 23-residue magainin antibiotic peptide is an alpha-helix in dodecylphosphocholine micelles, sodium dodecylsulfate micelles, and trifluoroethanol/water solution. J Biomol NMR 9(2):127–135