PROTEINS: Structure, Function, and Bioinformatics 59:252–265 (2005)

Properties of Integral Membrane Protein Structures: Derivation of an Implicit Membrane Potential

Martin B. Ulmschneider,[1]* Mark S.P. Sansom,[2] and Alfredo Di Nola[1]
[1]*Department of Chemistry, University of Rome 'La Sapienza,' Roma, Italy*
[2]*Department of Biochemistry, University of Oxford, Oxford, United Kingdom*

ABSTRACT **Distributions of each amino acid in the trans-membrane domain were calculated as a function of the membrane normal using all currently available α-helical membrane protein structures with resolutions better than 4 Å. The results were compared with previous sequence- and structure-based analyses. Calculation of the average hydrophobicity along the membrane normal demonstrated that the protein surface in the membrane domain is in fact much more hydrophobic than the protein core. While hydrophobic residues dominate the membrane domain, the interfacial regions of membrane proteins were found to be abundant in the small residues glycine, alanine, and serine, consistent with previous studies on membrane protein packing. Charged residues displayed nonsymmetric distributions with a preference for the intracellular interface. This effect was more prominent for Arg and Lys resulting in a direct confirmation of the positive inside rule. Potentials of mean force along the membrane normal were derived for each amino acid by fitting Gaussian functions to the residue distributions. The individual potentials agree well with experimental and theoretical considerations. The resulting implicit membrane potential was tested on various membrane proteins as well as single trans-membrane α-helices. All membrane proteins were found to be at an energy minimum when correctly inserted into the membrane. For α-helices both interfacial (i.e. surface bound) and inserted configurations were found to correspond to energy minima. The results demonstrate that the use of trans-membrane amino acid distributions to derive an implicit membrane representation yields meaningful residue potentials. Proteins 2005;59:252–265.**

Key words: amino acid distribution; membrane protein; implicit membrane; potential of mean force; α-helices

INTRODUCTION

Integral membrane proteins play a crucial role in cell function and communication. Current estimates indicate that 20–30% of the human genome encodes membrane proteins.[1–3] Even though the majority of drug targets are membrane proteins such as receptors and ion-channels[4] only 46 high-resolution structures of different membrane proteins are known at present. The scarcity of structural data is mainly a result of substantial difficulties with over-expression and crystallization of membrane proteins.[5] Recently, promising developments in the methodology of membrane protein structure determination have been reported.[6–9] Nevertheless it seems unlikely that the rate of structure determination will increase significantly in the near future.

The relative paucity of structural data has impeded the development of knowledge-based potentials that have been successfully applied in globular protein structure prediction.[10] Instead, a set of methods with increasing levels of sophistication has been developed to predict the topology of trans-membrane (TM) α-helices in membrane protein sequences, reaching accuracies close to 100%.[11–13] The prediction methods can be divided into two broad classes: i) hydrophobicity analyses of membrane protein sequences based on theoretical or experimental physiochemical considerations[13–18] and ii) statistical analyses based on known membrane protein structures or databases of experimentally confirmed membrane protein topologies.[19–26]

These methods have been employed to analyse the residue distributions and general properties of TM helices[27,28] and membrane protein structures[29–31] in order to extract common features. Others have concentrated on the role of individual residues such as proline-induced kinking of TM helices[32] or the importance of glycine in TM helix association.[33] Analyses of residue distributions have also been used to study the packing of membrane proteins[34–36] and to derive knowledge-based scales for membrane protein prediction and folding.[37,38]

The present work can be divided into two parts. The first is a detailed analysis of the distributions and preferred locations of each amino acid in the membrane domain using all currently available α-helical membrane protein structures. This analysis closely follows a previous publication,[29] which suffered from the scarcity of structures available at the time. Recent years, however, have seen a considerable increase in the number of membrane protein

Grant sponsor: Wellcome Trust

*Correspondence to: Martin B. Ulmschneider, Department of Chemistry, University of Rome 'La Sapienza,' Piazzale Aldo Moro 5, I-00185 Roma, Italy

Received 22 July 2004; Accepted 6 August 2004

Published online 18 February 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20334

structures available at atomic resolution thus making a reanalysis timely (an up to date summary of current structures is provided by White - http://blanco.biomol.uci.edu/).

The second part represents an assessment of the usefulness of these amino acid distributions to derive potentials of mean force for membrane protein folding and simulation. The calculation of the potentials of mean force was adapted from a method used for globular proteins.[39,40]

Due to the extremely high computational cost of molecular mechanics simulations using explicit lipid-bilayer membranes[41–43] there has been an increasing interest in implicit membrane representations.[44,45] The immense success of the generalized Born implicit solvation model[46,47] in globular protein and peptide folding simulations (see e.g., Ulmschneider and Jorgensen[48]) has spurred attempts to introduce the generalized Born formalism to represent the membrane environment implicitly.[49,50] These methods describe the membrane environment as a uniform hydrophobic slab and have been used successfully to fold and assemble small helical membrane peptides.[50]

In the present approach an implicit membrane representation was derived from the distributions of amino acids along the membrane normal. These distributions were calculated from all currently available α-helical membrane protein structures at resolutions better than 4 Å. Since the lipid bilayer environment provides the dominant driving forces for membrane protein folding and integration[51,52] it was assumed that the preference of different amino acids for clearly defined regions along the membrane normal is a direct result of the specific interactions of these amino acid with the membrane environment. Therefore, the basic hypothesis was that these distributions can be used to calculate a potential of mean force along the membrane normal for each amino acid, which correspond to an effective implicit membrane potential. The resulting implicit membrane representation can be integrated into a Monte Carlo or Molecular Dynamics algorithm.

METHODS

Membrane Proteins

The present study involved all 46 α-helical membrane protein structures currently (March 2004) available in the protein database with resolutions greater than 4 Å. A list of all proteins used is given in the Appendix. Where several structures of the same protein were available the highest resolution structure was used. Any identical chains were removed before the analysis. The present dataset represents a threefold increase in the number of proteins since our previous study.[29]

Aligning the Proteins

The crystal structures of all proteins were positioned in the membrane by minimizing the sum of the angles of their TM α-helices with respect to the bilayer normal (the z-axis) and centering their membrane domains on the membrane center. The membrane domain of the proteins was defined by the space between the intracellular and extracellular termini of the TM helices, determined using DSSP.[53] This method has been shown to produce good alignments with respect to the membrane normal particularly for membrane proteins with several identical domains like the KcsA potassium channel and multimeric membrane proteins such as bacteriorhodopsin.

The membrane center was placed at the origin $z = 0$ and proteins were aligned so that residues in the TM region facing the "outside" are along the negative z-axis and residues facing the "inside" along the positive z-axis. Thus the modulus of the distance represents the normal distance of the residue from the plane in the centre of the bilayer. The "inside" was defined as the cytoplasmic side of the (plasma) membrane of gram-positive (single membrane) bacteria, the cytosolic side of the inner (plasma) membrane of gram-negative (double membrane) bacteria, the stroma side of the thylakoid membrane in chloroplasts and the matrix side of the inner mitochondrial membrane. The "outside" is thus defined as the extracellular, periplasmic, lumen (thylakoid space) and inter-membrane side respectively.

Residue Distributions

The normal distance z of the backbone carbon α-atom from the membrane center was measured for each residue. Subsequently, the distribution $n_i(z)\,\Delta z$ along the bilayer normal was derived by counting the number of amino acids of type i = Ala, Arg, Asp, etc. in the interval $z \to z + \Delta z$. Unless stated otherwise values were averaged over the width of the interval which was chosen to be $\Delta z = 2.0$ Å.

Potentials of Mean Force

For each amino acid type i a potential of mean force $E_i(z)$ was calculated as a function of the membrane normal (z-axis) only. The potentials were derived by adapting a method used for globular proteins.[39,40] The measured frequency of residues $n_i(z)\,\Delta z$ was normalized giving

$$f_i(z)\Delta z = \frac{n_i(z)}{N_i}\,\Delta z, \tag{1}$$

where $N_i = \sum_z n_i(z)\Delta z$.

This normalized frequency distribution corresponds to a potential of mean force

$$E_i(z) = -kT \ln f_i(z). \tag{2}$$

Here k is the Boltzmann constant and T is the temperature of the native state of the protein. However, this potential is biased by the overall residue distribution $\sum_i f_i(z)$. To eliminate this bias, the potential of mean force of the overall residue distribution was chosen as the reference state

$$E_{ref}(z) = -kT \ln \sum_i^n f_i(z), \tag{3}$$

where the sum is over all amino acid types i. The resulting potentials of mean force are thus given by

$$\Delta E_i(z) = E_i(z) - E_{ref}(z). \tag{4}$$

TABLE I. Fitting Data[†]

Residue	a_0	a_1	a_2	a_3	a_4	a_5	a_6	χ^2	C [%]	RMS [%]
ALA	0.039	0.042	0.0081	−2.7				0.00081	95	0.11
ARG	0.055	−0.057	0.0039	0.8	0.050	0.009	17.0	0.00228	96	1.07
ASN	0.051	−0.040	0.0075	−1.6				0.00426	81	0.26
ASP	0.064	−0.063	0.0028	1.1	0.026	0.020	25.4	0.00262	96	0.33
GLN	0.054	−0.047	0.0071	0.0				0.00173	93	0.33
GLU	0.066	−0.064	0.0028	−0.7	0.012	0.015	29.4	0.00435	94	0.34
GLY	0.040	0.028	0.0060	−3.5				0.00233	80	0.19
HIS	0.025	0.056	0.0090	−20.0	0.040	0.015	15.0	0.01020	68	0.45
ILE	0.034	0.071	0.0058	−0.4				0.00178	97	0.17
LEU	0.032	0.062	0.0049	−0.1				0.00141	97	0.16
LYS	0.061	−0.060	0.0029	−2.8	0.035	0.015	23.6	0.00209	97	0.24
MET	0.034	0.050	0.0042	0.5				0.00281	91	0.22
PHE	0.024	0.075	0.0029	−1.5				0.00246	96	0.25
PRO	0.061	−0.041	0.0051	−1.1				0.00534	79	0.25
TRP	0.028	0.064	0.0143	−14.1	0.064	0.015	12.7	0.00216	96	0.20
TYR	0.030	0.045	0.0120	−16.0	0.030	0.020	13.7	0.00880	55	0.26
VAL	0.035	0.057	0.0084	−0.4				0.00394	89	0.20

[†]The parameters a_i are defined in Equation 5, a_3 and a_6 are the centers of the respective Gaussians. The χ^2 error is defined by Equation 6. C is the correlation coefficient and RMS is the root mean square percent error of the fitted function with respect to the measured data. Cys, Ser and Thr were not fitted (see text).

Fitting Gaussian Functions

The basic hypothesis of the present study was that each residue on its own would prefer a certain well defined region along the membrane normal. Therefore single or double peak Gaussians were fitted to the residue distributions giving smooth potential functions. Membranes are extremely fluid, and there is strong evidence that any residue not positioned in its proper environment causes a significant rearrangement of the peptide and/or surrounding lipids.[54,55] X-ray and neutron diffraction experiments on fluid liquid-crystalline bilayers have shown clear spatial separation (in the form of Gaussians) for the different principal structural groups of the lipids (carbonyl, phosphates, etc.).[56–58] In addition, experiments on a variety of tryptophan analogs[59,60] were found to have clearly defined positions at the membrane interfaces. Recent simulations of membrane proteins in explicit fluid lipid bilayers have confirmed a strong correlation of the positioning of the aromatic residue belts and lipid headgroups.[61] On the other hand the hydrophobic molecule hexane was found to have a Gaussian probability distribution centered on the core of the membrane.[62] At present the exact energetic reasons behind these preferences are not fully understood, but it seems very likely that they are the result of a complex interplay of electrostatic and hydrophobic (i.e., entropic) forces of the lipid bilayer with the unique electrostatic and surface/volume properties of each amino acid.[54,63–65]

The distributions were fitted using the trial function

$$f(z) = a_0 + a_1\exp(-a_2(z - a_3)^2) + a_4\exp(-a_5(z - a_6)^2), \quad (5)$$

by iteratively minimizing the χ^2 error

$$\chi^2 = \sum_{j=1}^{n} [y_j - f(z_j)]^2. \quad (6)$$

Here y_j are the measured values and $f(z_j)$ are the corresponding fitted values. n is the number of measured data points. The fitting parameters, minimum χ_2, correlation coefficients and root mean square errors of each amino acid type are shown in Table I.

Solvent and Lipid Accessible Residues

Not all residue side chains in the TM region face the lipid bilayer, some are buried within the TM domain of the membrane protein itself while others line the pores of ion channels. In order to differentiate these residues the accessible surface area of each residue was calculated using the Connolly method[66,67] with a probe radius of 1.4 Å. Missing nonhydrogen atoms in the protein structures were added using standard side chain stereochemistry.

Hydrogen atoms were not included. The accessibility fraction x is calculated by dividing the accessible surface area reached by the probe sphere by the total surface area of the respective residue in isolation.

RESULTS AND DISCUSSION

Statistics

The current study involved 46 α-helical membrane proteins containing 440 nonredundant TM helices. The results for helix length, height (i.e., the projection of the helix length onto the membrane normal) and tilt angles are in very close agreement with previous studies.[29,31,32] It is noteworthy that the results of Bowie, using just 45 TM α-helices, gave almost identical values to the current study. The mean number of residues per TM helix was found to be 26.3 (± 5.6) compared to Bowie's figure of 26.4. The mean height, defined as the projection of the TM helix onto the membrane normal, was found to be 33.7 (± 8.2) Å.

The average tilt angle of the helices has increased slightly to 24° (±14°) up from 21° in Bowie's analysis and 22° in our own previous analysis. This increase highlights that the current dataset has until recently been somewhat

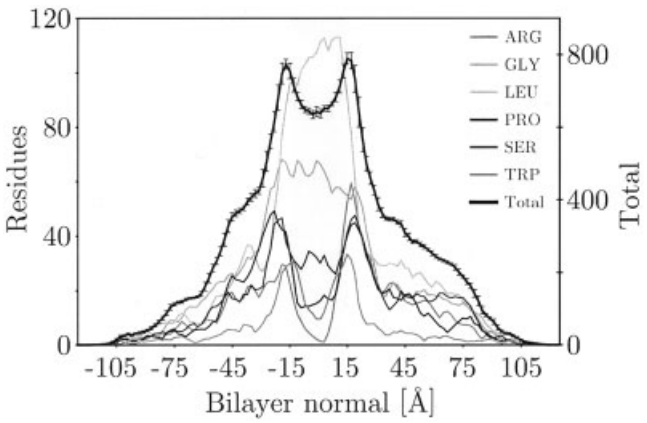

Fig. 1. Total residue distribution along the membrane normal (right side) and contributions from representative amino acid types, i.e., Arg = *charged*, Leu = *hydrophobic*, Ser = *polar*, Trp = *aromatic*. Gly and Pro were included because of their importance in membrane protein packing. The bin width was 2 Å and values were averaged over 3 Å (slightly overlapping bins).

biased towards topologically "simple" membrane proteins (see Byrne and Iwata[68] for a review of recent structures). Most of the early structures like the photosynthetic reaction center[69] have straight and only slightly tilted helices, while recently solved structures, like rhodopsin,[70] have revealed strongly kinked and titled helices, and yet others are curved rather than tilted and run almost parallel to the plane of the membrane.[71]

Membrane Domain Properties

The overall distribution of residues (i.e., the sum for all amino acid types) along the membrane normal is shown in Figure 1 together with the distribution of a few individual amino acids. It can be seen that α-helical proteins follow a saddle like distribution with two peaks at the interfacial regions, caused by aromatic, charged, and polar groups. The fanning out of residue types in the membrane region can be clearly observed and it should be noted that apart from the charged residues all other distributions are symmetric.

The six residues Leu, Ala, Val, Ile, Gly, and Phe together account for two thirds (63%) of residues in the TM domain and half of all residues (49%) in the protein structures investigated. As previously reported[72] the hydrophobic residues Ala, Ile, Leu, and Val make up the bulk of residues in the TM domains α-helical proteins accounting for almost half (45%) of residues in the membrane, with Leu being by far the most frequent residue (15%). Also significant is the high frequency of glycine in TM segments (9%). It has been reported that glycine residues occur frequently at helix–helix interfaces and crossing points[33] and it has been suggested that this may facilitate closer packing of TM helices,[36,73] especially in motifs combining Gly and β-branched side chains.[34] This packing has been explored via a series of NMR experiments on glycophorin A dimers,[74–77] which firmly established the essential role of Gly in the dimerization motif.

Gly was found to be the most frequent residue in the interfacial domains of membrane proteins (defined by the regions: −25 to −15 Å and 15 to 25 Å, c.f., White and Wimley[63]), accounting for 9% of all residues in this domain. Generally the interfacial regions show a much more homogeneous distribution compared to the membrane domain, with most residues having frequencies between 3–5%. Interestingly the four residues Gly, Ala, Ser, and Pro have the highest propensities in the interfacial regions (9%, 8%, 7%, and 7% respectively), excluding only Leucine (8%). It seems that small residues (Gly, Ala, and Ser are the three smallest residues) are advantageous in the interfaces because good packing in the loop regions of TM helices is more difficult to achieve with larger side chains. Proline on the other hand allows for unique backbone kinks and hence its presence at the interfaces might be advantageous because it allows backbone conformations not accessible to other amino acids. It should be noted that these values are in excellent agreement with a recent packing study which found Gly, Ala, Ser, and Pro to have among the highest packing values at the interfaces.[36]

Charged Residues

Energetic considerations suggest that charged amino acids should generally be excluded from TM segments.[15] In fact in the current analysis they account for less than 6% of the residues in the TM domain. However, charged residues make up one fifth (19%) of residues in the interfacial regions. Membrane proteins generally have an asymmetric charge distribution along the membrane normal. This provides for the correct orientation of the protein in the membrane as well as preventing the loss of the protein to the extracellular space. Indeed Figure 2 shows that there is a bulk of charge on the expression side of the membrane protein (intracellular, matrix or stroma, see Methods).

Figure 3 demonstrates the asymmetry of the charge distribution. The net charge along the membrane normal was calculated by assuming all ionizable residues (Arg, Asp, Glu, Lys) with a surface accessibility greater than 10% to be charged, while all others were taken to be neutral. The averaged net total charge per protein on the "inside" (i.e., cytoplasm, matrix or stroma, $0 < z < +\infty$) was found to be $+3.8 \pm 0.2\,e$, compared to $-4.5 \pm 0.2\,e$ on the "outside" ($-\infty < z < 0$), giving strong support to the "positive-inside rule".[17] The ratio of net surface charge (outside/inside) was found to be -1.34 ± 0.2, averaging over a surface accessibility range of $10\% < x < 70\%$. Thus for every three positive residues on the intracellular side there are four negative residues on the extracellular side.

Hydrophobic Residues

As expected all four hydrophobic residues Ala, Ile, Leu, and Val show a clear preference for the trans-bilayer region [Fig. 2(B)], in good agreement with previous results.[72]

An analysis of the surface accessible residue propensities is summarized in Table II. It shows the percentage propensity of hydrophobic and charged amino acid types on the protein surface for the interfacial and transmembrane domains. The results were averaged over a

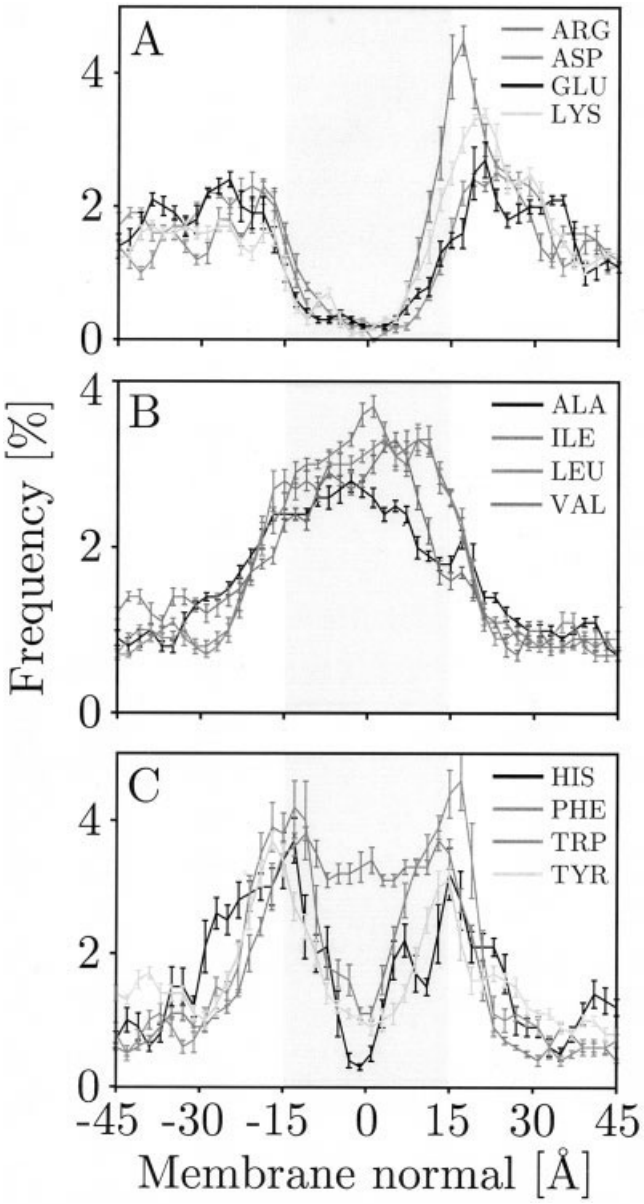

Fig. 2. Normalized distributions for charged (A), hydrophobic (B) and aromatic residues (C). The membrane domain is shaded with the cytoplasmic side is to the right (positive).

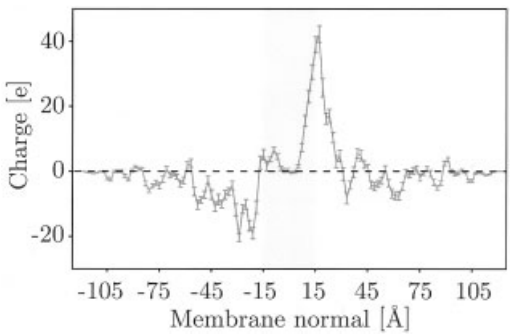

Fig. 3. Net charge along the membrane normal summed over all 46 proteins. The cytoplasm, matrix or stroma is to the right of the membrane (shaded).

surface accessibility range of 10–50%. Our previous analysis calculated the surface fraction (i.e., the percentage of residues of a certain type located on the surface), but it was found that this ratio is strongly dependent on the surface accessibility. In contrast the standard deviations in Table II show that the surface propensity does not change considerably with surface accessibility.

Table II shows that the surface propensities of Phe, Leu, Ile, Val, Ala, and Gly are only slightly higher than in the remainder of the membrane domain (63%, see Membrane domain properties). However, the larger hydrophobic side chains Phe, Leu, Ile, and Val account for half of all surface residue in the TM region, compared to 43% for the remainder of the membrane domain (i.e., they show a somewhat clearer preference for the TM domain surface).

The behavior of the small side chains Ala and Gly is also interesting; their surface propensity does not change from the interface to the membrane (c.f., total is 19% in the TM domain and 17% at the interfaces). As stated above, these residues play an important role in helix–helix packing due to their short side chains[34,36,73] thus explaining their preference for the loop and core regions of α-helical membrane proteins.

Hydropathy Analysis

In order to further investigate the nature of the transmembrane domain a hydropathy analysis was performed using various hydrophobicity scales. The average hydrophobicity with respect to the membrane normal was calculated for buried and surface-exposed residues as a function of the membrane normal (see Fig. 4 caption). The results are displayed in Figure 4 using a recent knowledge-based hydrophobicity scale.[20] It can be seen that the protein surface changes from very hydrophilic in the aqueous domains to strongly hydrophobic in the trans-membrane region with steep gradients at the membrane interfaces. Buried residues, on the other hand, are only slightly more hydrophobic in the membrane compared to the soluble domains. However, they are influenced by the exterior protein environment of the interfacial regions, being more hydrophilic here than elsewhere in the protein. Thus the remarkable feature of this analysis is that the protein surface in the membrane domain is much more hydrophobic than the protein interior, suggesting that membrane proteins are indeed somewhat "inside-out," compared to globular proteins, at least in their trans-membrane domain.

This result contradicts a previous study of seven α-helical membrane proteins[30] that found no correlation between hydrophobicity and surface accessibility in the TM domain. It also suggests that the highly debated issue whether the TM surface of an α-helical membrane protein is more hydrophobic than its core can be justified on the basis of the current dataset of structures (see references 30, 78, and 79 for an extended discussion). It should also be noted that this result is contrary to our own previous conclusions based on a reduced dataset of 15 α-helical membrane proteins,[29] which found no preference of Phe, Leu, Ile, and Val for the trans-membrane domain surfaces.

TABLE II. Surface Propensities of Hydrophobic and Charged Residues[†]

	Residue Propensity [%]			
	Surface		Total	
Residues	Membrane	Interface	Membrane	Interface
Phe, Leu, Ile, Val, Ala, Gly	65.2 ± 0.5	37.3 ± 1.3	62.6	40.4
Phe, Leu, Ile, Val	50.4 ± 1.0	22.4 ± 1.1	43.2	23.8
Ala, Gly	14.8 ± 0.6	14.9 ± 0.2	19.4	16.6
Arg, Asp, Glu, Lys	7.3 ± 0.6	24.1 ± 1.4	5.8	19.0

[†]A residue was defined to be on the surface if it has an accessibility greater than χ. The data was averaged for the range $10\% < \chi < 50\%$, standard deviations are given.

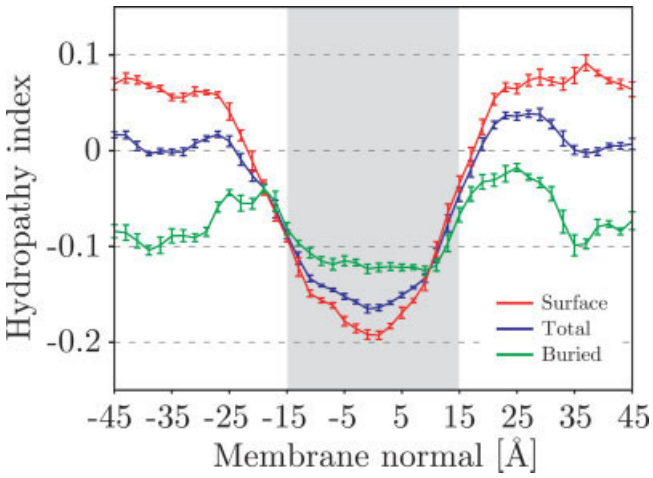

Fig. 4. The hydrophobicity along the membrane normal was calculated by assigning each residue of type i a hydropathy value and multiplying these values to the measured residue distributions $n_i(z)\,\Delta z$ along the bilayer normal. Dividing by the total number of residues in the interval Δz gives the average hydropathy along the membrane normal. In this figure all residues with a surface accessibility of $x > 20\%$ were considered to be on the surface.

The current analysis was repeated for a number of widely used hydrophobicity scales.[14,16,80–83] The results are in excellent agreement, with each scale reproducing the effect described above faithfully. A residue was defined to be on the protein surface if it has an accessibility greater than x. Variation of the data was tested for the range $5\% < x < 75\%$. The result was found to be qualitatively invariant against change in x. With increasing x the *buried* curve approaches the *total* (which does not change), while the *surface* curve becomes more hydrophilic in the aqueous domains and more hydrophobic in the membrane domain (moving away from the *total* curve). Thus it can be concluded with certainty that the effect shown in Figure 4 is genuine and not a computational artifact.

Aromatic Residues

Aromatic residues have been suggested to play a special role in membrane proteins (see e.g., references 59 and 84). They are believed to anchor the proteins into the membrane through an interaction of their aromatic rings with the lipid head groups. A preferred localization of aromatic residues in the interfacial regions has previously been noted for both the photosynthetic reaction center[69] and bacterial porins.[85] Such anchoring has been explored via NMR,[59,86] molecular dynamics simulations[42,61] and by experimental studies of synthetic trans-membrane peptides,[87,88] which found that even though tryptophan has the most hydrophobic side chain of all residues it resists partitioning with its indole NH group below the carbonyl region of a bilayer.[88]

Tryptophan, tyrosine, and histidine were found to have highly symmetrical distributions with a pronounced peak at each membrane interface [Fig. 2(C)]. In contrast, phenylalanine is distributed throughout the trans-bilayer region, behaving similar to hydrophobic residues [c.f., Fig. 2(B)]. This different behavior indicates that there is a strong penalty associated with the burial of hydrogen bonding groups in the other aromatics. These results are in general agreement with the kPROT analysis of all predicted α-helical membrane proteins in the SWISS-PROT database[37] and with the earlier analysis of Landolt-Marticorena et al.[27]

Polar Residues

The uncharged polar residues display two different types of behavior. Asparagine and glutamine follow the distribution pattern of charged residues avoiding the TM region (data not shown). This presumably reflects their need to form multiple hydrogen bonds. In contrast serine and threonine closely follow the overall residue distributions thus showing no preference for either the transmembrane or extra-membrane region (see Fig. 1). It has been noted[89] that serine and threonine side chains in a helix can form hydrogen bonds to the carbonyl oxygen of the preceding turn of the helix, thus enabling such side chains to occur in the TM region. Furthermore, as noted by Eilers et al.[73] serine and threonine may be associated with tight packing of TM α-helices.

Proline and Glycine Residues

Proline plays a special role in TM α-helices[90–93] due to its ability to generate a helix kink. In the present analysis, as expected, proline was found to occur predominantly in the interfacial loop regions (see Fig. 1). Nevertheless, unlike charged and polar (Asn, Gln) residues, it is also represented throughout the membrane region. Indeed, it has been suggested that prolines may increase the stability of the TM domain by "interlocking" helices, or by providing molecular hinges that enable conformational

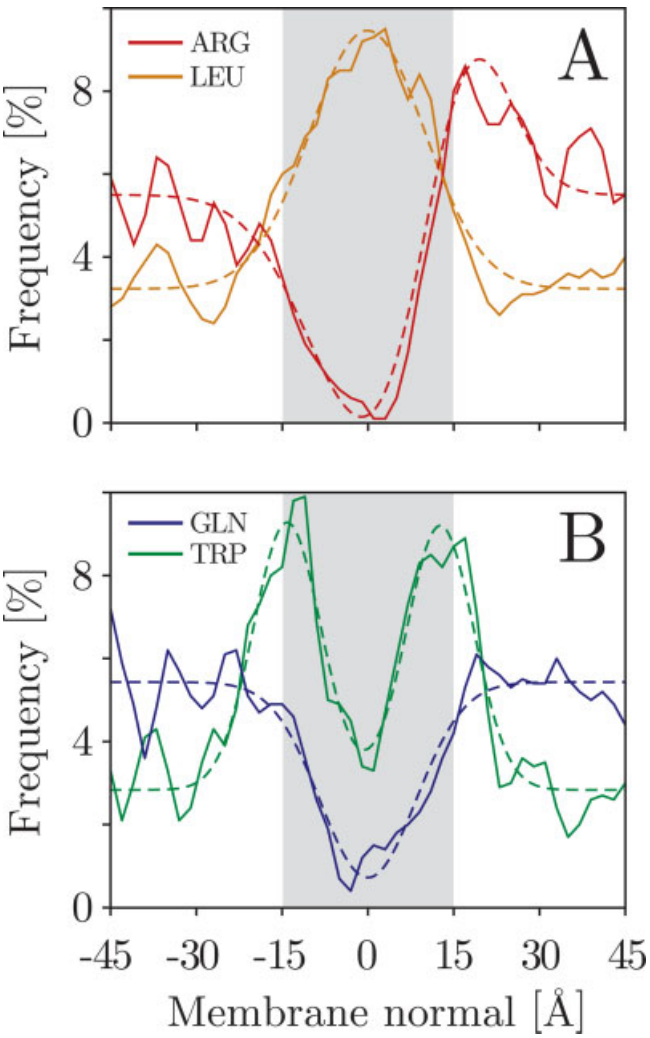

Fig. 5. The measured frequencies (solid lines) and fitted functions (dashed lines) for the four different types of functions fitted.

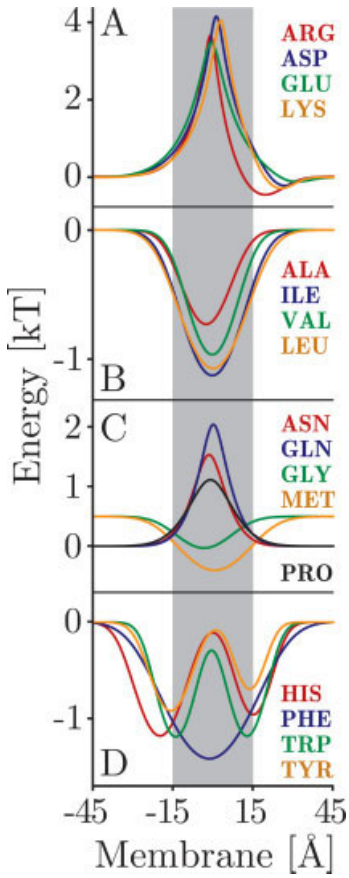

Fig. 6. Potentials of mean force for charged (A), hydrophobic (B), polar (C) and aromatic residues (D). The energy is given in $kT \approx 0.6$ kcal/mol at T = 300 K.

transitions.[93,94] Glycine was found to have a preference for the membrane region, behaving more like a hydrophobic residue. This preference was not detected in our previous analysis with a reduced dataset but is consistent with other studies which found Gly to be twice as abundant in membrane proteins than soluble proteins.[36,73]

Potentials of Mean Force

Figure 5 shows the fitting of smooth Gaussian functions to the normalized distributions after subtraction of the reference state (i.e., division by the total distribution, see Methods). All four different types of topology that were used in the fitting are shown by a representative residue (Arg, Leu, Gln, and Trp). Hydrophobic residues Ala, Ile, Leu, Val as well as Phe, Gly, and Met were fitted with a single upright Gaussian. Polar residues Asn, Gln, and Pro were fitted with a single inverted Gaussian centered in the membrane. Aromatics Trp, Tyr, and His were fitted with two upright Gaussians one at each membrane interface. Charged residues Arg, Asp, Glu, and Lys were fitted with double Gaussians, one inverted near the membrane center and another upright at the cytoplasmic interface. Residues Cys, Ser, and Thr were not fitted. Cys because it occurs too infrequently to be statistically valid and Ser and Thr because the potential is essentially flat after subtraction of the reference state.

Figure 6 shows the corresponding potentials of mean force for all residues. Table I lists the fitting parameters, χ^2 error values as well as the correlation coefficients and RMS errors for all amino acids. In general, the quality of the fit is very good, with hydrophobic and charged residues displaying the best correlations. These results are encouraging and demonstrate that the use of Gaussians is a reasonable approximation. It should be noted that curves were only fitted in the range −45 to +45 Å since beyond this range the number of residues drops significantly (c.f. Fig. 1).

Hydrophobic residues display a potential energy well near the center of the membrane region (see parameter a_3 in Table I) and extending into the interfacial regions [Fig. 6(B)]. This agrees with mass spectrometry experiments on synthetic membrane peptides that found that introducing Ala and Leu residues into the polar interfacial regions seems to have relatively small energy penalties.[87] The free energy of transfer from water to the membrane interface can be compared with the experimental interface scale of

Wimley and White.[83] They found values of -0.31 ± 0.06 kcal/mol and -0.56 ± 0.04 kcal/mol for Ile and Leu respectively, which compare to our values of -0.25 ± 0.02 kcal/mol and -0.28 ± 0.02 kcal/mol, obtained by averaging over both membrane interfaces.

For charged residues, the potentials of mean force have a narrow peak at the membrane center and a slight depression at the cytoplasmic interface [Fig. 6(A)]. The cost of burying a charged residue within the hydrocarbon core of a lipid bilayer is extremely high (~9 kcal/mol for a Lys residue[15]). However, in the current potentials it is only ~3 kcal/mol, much smaller than the theoretical cost of neutralization and burial of 10–20 kcal/mol.[44] This discrepancy can largely be attributed to the extremely low propensities of charged residues in the membrane center, making good fitting in this region very problematic (a zero residue propensity results in an infinite potential, while small changes close to zero produce large changes in the potential). Thus the current method clearly underestimates the penalty for charged residues to be buried in the membrane center.

It should also be noted that the present analysis cannot distinguish between charged and neutral residues. One third of ionizable residues at the membrane interfaces have surface accessibilities greater than 50% and are therefore almost certainly charged. Thus the average free energy of transfer from water to the membrane interface for Asp and Glu (assuming one-third charged and two-thirds neutral) according to Wimley and White[83] is 0.36 ± 0.10 kcal/mol and 0.66 ± 0.14 kcal/mol respectively. This compares to our values of 0.39 ± 0.04 kcal/mol and 0.41 ± 0.03 kcal/mol respectively, which were averaged over both bilayer interfaces.

Experimental evidence suggests that, while resisting partitioning into the membrane below the level of the phosphates, Lys does not appear to resist displacement from the interface towards the aqueous phase,[54] in good topological agreement with the shape of the present potentials.

Aromatic residues (His, Trp and Tyr) have potentials of mean force with two wells, one at each membrane interface [Fig. 6(D)]. This potential shape was expected from structural, experimental, and computational data (see above). The penalty of moving Trp or Tyr from the interfaces to the aqueous domain was found to be 0.68 kcal/mol and 0.47 kcal/mol respectively, much lower than the corresponding values from Wimley and White's interface scale (1.85 kcal/mol and 0.94 kcal/mol).

Polar residue potentials (only Asn and Gln, see above) display a single broad peak centered in the membrane [Fig. 6(C)]. The energy penalty of displacing a polar residue from the solvent to the interface is relatively small ~0.1 kcal/mol, while the penalty for insertion into the membrane core is around 2 kcal/mol.

Generally, topological differences within each group of residues (i.e., hydrophobic, charged, aromatic, and polar) are small and show only subtle differences in the distributions and resulting potentials of mean force. This agrees well with experimental observation from synthetic transmembrane peptides which found only minor differences on substitution of Lys with Arg and Trp with Tyr as flanking residues.[64,65]

Finally, it should be noted that the energies of transfer from the interface to the aqueous solution, although slightly different in magnitude (average error of 0.5 kcal/mol for the interface scale and 1.0 kcal/mol for the octanol scale), nevertheless correlate highly with both the octanol (85%) and interface scale (88%) of Wimley and White.[83] This correlation is highest for the hydrophobic (98%), charged, and polar residues (80%), while there is little correlation for aromatics (33%). For insertion into the center of the membrane the free-energy correlations are 87% with the octanol scale and 78% with the interface scale. This means that the present interface free energies correlate better with the experimental interface scale, while the buried scale correlates better with the experimental octanol scale, which is encouraging for the correctness of the overall shape of the potentials.

However, it should be noted that the current study is not attempting to provide accurate free-energy profiles but to make an initial assessment of the validity of using transmembrane residue distributions to derive an implicit membrane representation for simulation studies.

Membrane Protein Insertion

The potentials of mean force were tested on various membrane proteins: bacteriorhodopsin (*1cwq.pdb*),[95] sensory rhodopsin (*1h68.pdb*),[96] the KcsA potassium ion-channel (*1k4c.pdb*),[97] the GLPT glycerol-3-phosphate transporter (*1pw4.pdb*),[98] the glycophorin A dimer (*1afo.pdb*),[99] as well as two aquaporins (*1j4n.pdb* and *1rc2.pdb*)[100] and two chloride channels (*1kpk.pdb* and *1kpl.pdb*).[101] Residue distributions were calculated leaving each protein out in turn. However, the deviations between the distributions were found to be much smaller than the error in the curve fitting (see Methods). For the distribution without GLTP, the largest protein in the test set, the error with respect to the total distribution was $\chi^2 = 1.4 \times 10^{-5}$, resulting in identical curve fits. The errors for the other proteins are even smaller.

The aligned proteins were moved through the membrane and the energy recorded as a function of the distance from the membrane center. Figure 7 demonstrates that the completely inserted configuration is at an energy minimum for all membrane proteins investigated (c.f. z_{min} in Table III). The energy minima were found to be within 2.5 Å of the membrane center.

All energy profiles are asymmetric across the membrane. Insertion from the cytoplasmic side is more favorable, exhibiting no energy barrier, while the extracellular side has a steeper gradient and (with the exception of sensory rhodopsin and one aquaporin) shows a slight penalty for insertion. This result agrees well with the solvation energy profile recorded for a recently developed generalized Born implicit membrane representation,[49] which also found insertion from the cytoplasmic region more favorable. However, their energy of solvation was found to be much higher being 143.1 kcal/mol for bacterio-

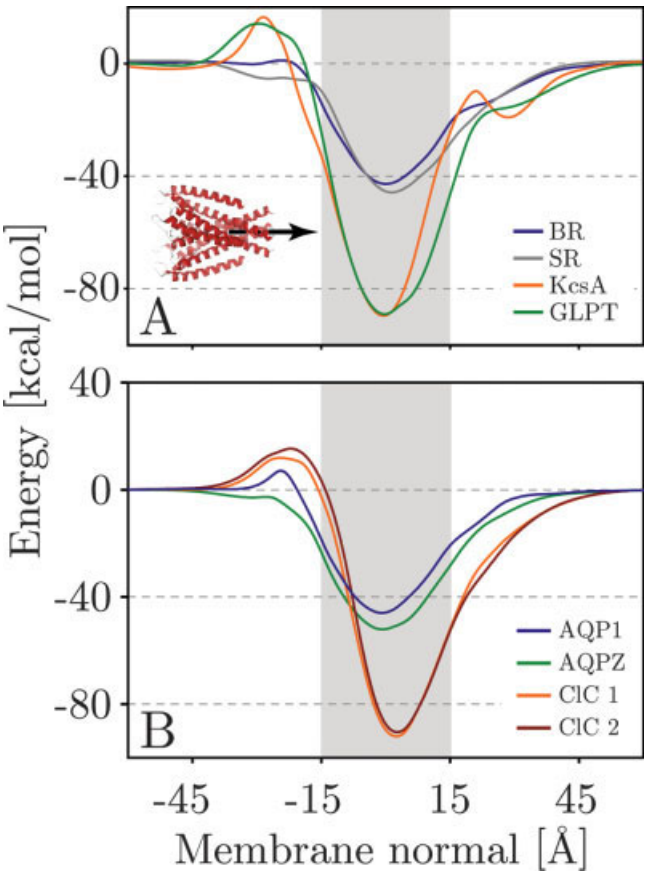

Fig. 7. Insertion energy profiles derived by pushing the aligned proteins through the potential of mean force membrane representation. The extracellular side of the membrane is to the left (negative *z*-axis). The proteins shown are bacteriorhodopsin (BR), sensory rhodopsin (SR), the KcsA potassium channel, the glycerol-3-phosphate transporter (GLPT), aquaporins from *Bovine red blood cell* (AQP1) and *E. coli* (AQPZ) as well as chloride channels from *E. coli* (ClC 1) and *S. typhimurium* (ClC 2). Proteins shown in B have a more irregular secondary structure than those shown in A.

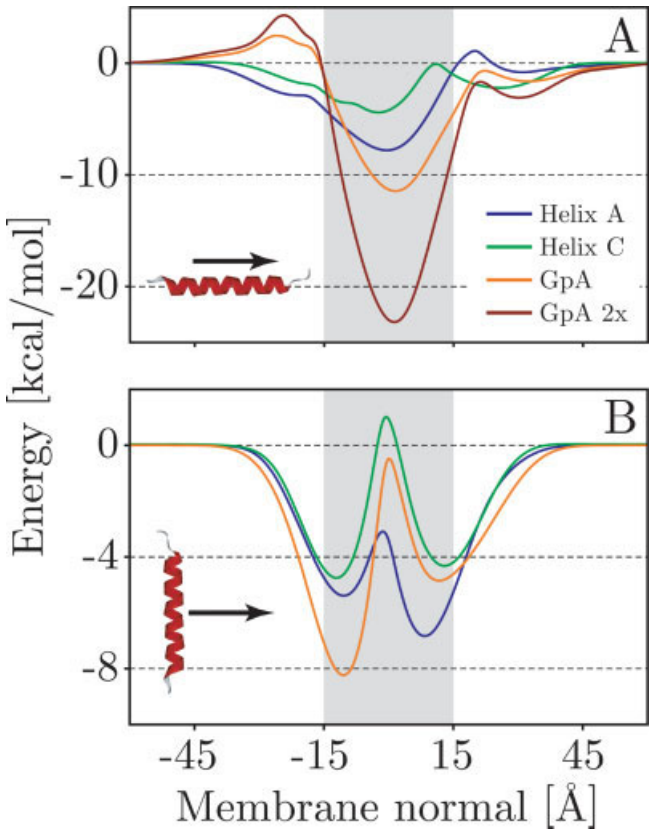

Fig. 8. Helices A and B of bacteriorhodopsin as well as monomeric (GpA) and dimeric (GpA 2x) glycophorin A. A: the energy profile of inserting the helix perpendicular to the membrane surface. B: the profile for a helix parallel to the membrane surface.

TABLE III. Energy of Insertion into the Membrane for the Glycophorin A Dimer (GpA 2x), Bacteriorhodopsin (BR), Sensory Rhodopsin (SR), the KcsA Potassium Channel, the Glycerol-3- Phosphate Transporter (GLPT), Aquaporins from *Bovine red blood cell* (AQPI) and *E. coli* (AQPZ) as Well as Chloride Channels from *E. coli* (CIC 1) and *S. typhimurium* (CIC 2)[†]

	Energy minima		
Protein	ΔE_{min} [kcal/mol]	z_{min} [Å]	α_{min} [degrees]
GpA 2x	−23.2	1.5	4
BR	−42.9	0.0	13
SR	−46.7	1.5	2
KcsA	−90.3	−0.5	1
GLPT	−88.9	0.0	15
AQPI	−46.0	−1.0	25
AQPZ	−52.1	−1.0	28
CIC 1	−92.0	2.5	13
CIC 2	−90.4	2.5	0

[†]The depth of the energy well ΔE_{min}, optimal tilt angle α_{min}, and position with respect to the membrane center z_{min} are given.

rhodopsin, compared to the 42.9 kcal/mol found in this study.

In the inserted configuration the membrane model was tested by rotating the protein in the center of the membrane. The minimal tilt angles are given in Table III and are in the range 0°–15°, except for the aquaporins which have tilt angles nearer 30°. All proteins have a second minimum near 180° (i.e., upside down in the membrane), but in all cases this was found to have significantly higher energies (~20%) suggesting that the present potential captures the inside/outside orientation of the proteins correctly. For bacteriorhodopsin the tilt angle of 13° compares well with the 12° from the crystal structures.

Interestingly the insertion profiles of proteins with more irregular secondary structures like aquaporins and chloride channels do not differ from the those of very regular structures such as bacteriorhodopsin or KcsA (c.f., Fig. 7).

Trans-Membrane Helices

Figure 8 shows the energy profiles for helices A and C of bacteriorhodopsin (from *1cwq.pdb*) as well as monomeric and dimeric glycophorin A (from *1afo.pdb*). Two types of orientations were investigated, parallel to the membrane normal and parallel to the membrane surface.

A stable helical trans-membrane configuration has been experimentally verified for helix A,[102] and is well docu-

mented for glycophorin A. Helix C on the other hand is one of the few systems for which quantitative binding and insertion data is available.[63] At neutral pH, it associates with the membrane in a nonhelical probably peripheral conformation, while forming a stable TM helix upon protonation of its aspartate residues.[103]

The current study found that all helices aligned parallel to the membrane normal have an energy minimum close to the membrane center. The relative energy differences with respect to the aqueous domain are −7.8 kcal/mol, −4.4 kcal/mol and −11.5 kcal/mol for helix A, C, and the glycophorin A monomer respectively. These values compare well with experimental estimates of the free energy of insertion for a single TM helix, which are in the range of 5–12 kcal/mol.[103–106] It should be noted, however, that experimental difficulties make these values somewhat unreliable.[63,107]

Both the monomeric and dimeric glycophorin A exhibit a slight penalty for crossing into the extracellular space. The dimer has exactly twice the insertion energy (c.f., Table III) compared to the monomer at a tilt angle of 4° compared to 27° for the monomer. The insertion energy for the dimer is comparable to values obtained from PB/SA calculations (−18 kcal/mol).[44]

Moving the helices across the membrane while keeping their axes parallel to the membrane surface showed a very interesting feature of the present potentials. All helices exhibit potential energy wells close to the interfacial regions [±10–13 Å, see Fig. 8(B)]. At the center of the membrane, conformations perpendicular to the membrane normal have significantly higher energies than the TM configurations. This is not necessarily the case near the interfaces. In fact, helices A and C have 2–4 kcal/mol lower energies when oriented parallel to the membrane surface at the intracellular interface. Glycophorin A behaves similar at the extracellular interface. The surface parallel interfacial configuration of helix C was even found to be lower than the inserted configurations, which is a remarkable finding since experimental evidence indeed suggests a partially unfolded surface bound conformation.[103] Incidentally a recently developed energy function for membrane peptides and proteins also found a partially unfolded interfacial configuration to have lower energies than the TM configuration.[44]

The calculations were repeated for the ten NMR structures of the M2 helix of the δ-subunit of the actylcholine receptor.[108] The curves are exactly similar to those in Figure 8 (data not shown). Generally inserted TM configurations are the most stable, with an average energy minimum of −4.6 ± 0.1 kcal/mol at the center of the membrane (−0.9 ± 0.6 Å) and the optimal tilt angle of 9 ± 5° is comparable to the 12° determined by NMR.[108] Adsorption of the peptide onto the membrane surface is also favorable but to a lesser extent, with energy minima of −3.0 ± 0.6 kcal/mol for the cytoplasmic (9.4 ± 0.6 Å) and −2.4 ± 0.7 kcal/mol for the extracellular interface (−11.0 ± 1.0 Å).

These results are in excellent agreement with a recent theoretical study of the same structures,[109,110] which found average energies of −4.7 ± 2.1 kcal/mol and −2.6 ± 2.4 kcal/mol for inserted and surface bound configurations respectively. The study used a theoretical continuum-solvent method developed by Ben-Tal[111] that has been successfully applied to estimate the insertion energies of TM peptides and proteins.[112] In order to compare the results the helix–coil transition free energy (ΔG_{con} = −2.4 kcal/mol) was subtracted, since the present data estimates the insertion energy of a folded helix.

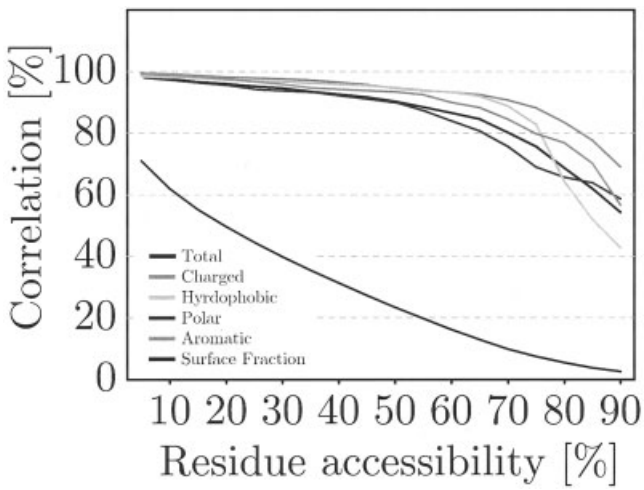

Fig. 9. Correlation of surface and overall distributions according to residue types (charged, hydrophobic, polar, aromatic, and total). The number of residues located at the surface as a fraction of all residues is also shown (surface fraction). The correlations are plotted against the fraction of a side chain that has to be accessible in order for that residue to be considered on the surface of the protein (accessibility fraction). For residues with up to 50% of their side chains exposed to the environment (membrane or water) the correlation is over 90%, even though they represent just ≈25% of all residues.

Future Improvements

Future improvements of the present membrane representation might have to include separate potentials for surface-accessible ionizable residues. Also, the surface dependence of the hydropathy analysis suggests that the strictly additive nature of the potentials (c.f., GpA monomer/dimer) might be overestimating the free energy of membrane insertion of proteins with larger trans-membrane segments. It should be noted that the current potentials were derived from folded conformations only. Therefore it is not entirely certain that the resulting potentials are sufficient to study protein folding or if a free-energy term associated with backbone exposure has to be included.

Distributions at the Protein Surface

Figure 9 shows the correlation of the total and surface distributions as a function of the residue accessibility. A residue accessibility of 10% means that a residue with more than 10% of its side chain surface area accessible to the environment (solvent or membrane) is considered a surface residue. The figure demonstrates that there is a very strong correlation for all but the most exposed hydrophobic residues. Less than a quarter (23%) of resi-

dues have side-chain accessibilities greater than 50%. Nevertheless their correlation with the total distribution is still over 90%. Interestingly the correlation is highest for charged (95%) and hydrophobic (95%) amino acids.

Current theory states that α-helical membrane proteins fold by forming and inserting their helices individually or in pairs and assembling them at a later stage.[52,113] Indeed individual fragments of bacteriorhodopsin form secondary structure when immersed in a membrane environment and subsequently combine to form a functional protein.[114,115] Consequently all TM segments, whether buried inside the protein or exposed to the lipid bilayer, should exhibit the same distribution pattern, since they insert on their own. This is in excellent agreement with present results (c.f., Fig. 9).

In the present study the membrane potential was derived from the distributions of all residues, which is justified for α-helical membrane proteins by the above analysis. However, residues with very high surface accessibilities probably contribute more to the insertion free energy than buried residues. On the other hand many biological bilayers (such as the mitochondrial or purple membranes) have extremely high protein densities, leading to significant protein–protein contacts. Furthermore, many membrane proteins, even in membranes of lower protein lipid ratios are oligomers. As a result it is difficult to estimate exactly how much of a residue is exposed to the solvent or membrane environment, and consequently the surface accessibility contribution to the free energy of insertion is difficult to assess. Nevertheless, the current analysis seems to indicate that the insertion of a protein fragment into a membrane might be energetically similar to its burial inside a membrane protein.

CONCLUSION

It is generally recognized that overall hydrophobicity is the main driving force for the integration of α-helical trans-membrane segments into the lipid bilayer.[116] The current study found that the vast majority of residues in the membrane domain are hydrophobic. Furthermore, the protein surface facing the lipids was found to be even more hydrophobic than the protein core, suggesting that membrane proteins can indeed be regarded as somewhat "inside-out," at least regarding their membrane domains.

The distributions of all amino acids were found to be symmetric with the exception of the four charged residues, which occur more frequently on the cytoplasmic side of the membrane. In addition to this asymmetry they were found to be distributed such as to cause a net charge imbalance across the membrane domain, in line with the positive inside rule.

The variation within each group of residue distributions (i.e., hydrophobic, charged, aromatic, and polar) were found to be small and caused only subtle differences in the resulting potentials of mean force. The shape of the potentials were shown to be consistent with experimental data and correlate well with measured free energies of solvation both for buried and interfacial locations.

The resulting membrane potential was tested on several integral membrane proteins. In all cases the correctly inserted orientation was found to be at a clear energy minimum. Further investigations with single transmembrane α-helices found that both inserted and surface bound conformations are at energy minima, consistent with theoretical, experimental, and simulation data.

The translational and rotational energy profiles described here represents a fairly limited search of the orientation space of the peptides and proteins considered. Nevertheless the present preliminary study has clearly demonstrated that the number of membrane proteins solved at atomic resolution is now sufficient for a detailed statistical analysis of the amino acid distribution functions as well as the derivation of meaningful potentials of mean force. The smoothness of the energy profiles is remarkable and the good overall agreement with experimental, statistical, and simulation data is encouraging.

ACKNOWLEDGMENTS

MBU is funded by the Wellcome Trust. Research in MSPS's group is supported by the Wellcome Trust. We would like to thank Saraswathi Vishveshwara and the Molecular Biophysics Unit at the IISc, Bangalore, India for providing facilities and Phil Biggin for his comments on this manuscript.

REFERENCES

1. Jones DT. Do transmembrane protein superfolds exist? FEBS Lett 1998;423:281–285.
2. Wallin E, von Heijne G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein Sci 1998;7:1029–1038.
3. von Heijne G. Recent advances in the understanding of membrane protein assembly and structure. Q Rev Biophys 1999;32: 285–307.
4. Terstappen GC, Reggiani A. In silico research in drug discovery. Trends Pharm Sci 2001;22:23–26.
5. Grisshammer R, Tate CG. Overexpression of integral membrane-proteins for structural studies. Q Rev Biophys 1995;28:315–422.
6. Chiu ML, Nollert P, Loewen MC, Belrhali H, Pebay-Peyroula E, Rosenbusch JP, Landau EM. Crystallization in cubo: general applicability to membrane proteins. Acta Crystallogr 2000;56:781–784.
7. Fernández C, Hilty C, Wider G, Güntert P, Wüthrich K. NMR structure of the integral membrane protein OmpX. J Mol Biol 2004;336:1211–1221.
8. Fernández C, Wüthrich K. NMR solution structure determination of membrane proteins reconstituted in detergent micelles. FEBS Lett 2003;555:144–150.
9. Tamm LK, Abildgaard F, Arora A, Blad H, Bushweller JH. Structure, dynamics and function of the outer membrane protein A (OmpA) and influenza hemagglutinin fusion domain in detergent micelles by solution NMR. FEBS Lett 2003;555:139–143.
10. Lazaridis T, Karplus M. Effective energy functions for protein structure prediction. Curr Opin Struct Biol 2000;10:139–145.
11. Rost B, Casadio R, Fariselli P, Sander C. Transmembrane helices predicted at 95% accuracy. Protein Sci 1995;4:521–533.
12. Rost B, Fariselli P, Casadio R. Topology prediction for helical transmembrane proteins at 86% accuracy. Protein Sci 1996;5: 1704–1718.
13. Jayasinghe S, Hristova K, White SH. Energetics, stability, and prediction of transmembrane helices. J Mol Biol 2001;312:927–934.
14. Zhou H, Zhou Y. Stability scale and atomic solvation parameters extracted from 1023 mutation experiments. Proteins 2002;49:483–492.
15. Engelman DM, Steitz TA, Goldman A. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu Rev Biophys Biophys Chem 1986;15:321–353.

16. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157:105–132.
17. von Heijne G. Membrane-protein structure prediction—hydrophobicity analysis and the positive-inside rule. J Mol Biol 1992;225: 487–494.
18. Czerzo M, Wallin E, Simon I, von Heijne G, Elofsson A. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. Protein Eng 1997;10:673–676.
19. Zhou H, Zhou Y. Predicting the topology of transmembrane helical proteins using mean burial propensity and a hidden-Markov-model-based method. Protein Sci 2003;12:1547–1555.
20. Punta M, Maritan A. A knowledge-based scale for amino acid membrane propensity. Proteins 2003;50:114–121.
21. Persson B, Argos P. Prediction of transmembrane segments utilising multiple sequence alignments. J Mol Biol 1994;237:182–192.
22. Persson B, Argos P. Prediction of membrane protein topology utilizing multiple sequence alignments. J Protein Chem 1997;16: 453–457.
23. Casadio R, Fariselli P, Taroni C, Compiani M. A predictor of transmembrane alpha-helix domains of proteins based on neural networks. Eur Biophys J 1996;24:165–178.
24. Jones DT, Taylor WR, Thornton JM. A model recognition approach to the prediction of all-helical membrane protein structure and topology. Biochemistry 1994;33:3038–3049.
25. Hofmann K, Stoffel W. TMbase—a database of membrane spanning protein segments. Biol Chem Hoppe-Seyler 1993;374:166.
26. Sonnhammer ELL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol 1998;6:175–182.
27. Landolt-Marticorena C, Williams KA, Deber CM, Reithmeier AF. Non-random distribution of amino acids in the transmembrane segments of human type I single span membrane proteins. J Mol Biol 1993;229:602–608.
28. Arkin IT, Brunger AT. Statistical analysis of predicted transmembrane alpha-helices. Biochim Biophys Acta 1998;1429:113–128.
29. Ulmschneider MB, Sansom MSP. Amino acid distributions in integral membrane protein structures. Biochim Biophys Acta 2001;1512:1–14.
30. Stevens TJ, Arkin IT. Are membrane proteins "inside-out" proteins? Proteins 1999;36:135–143.
31. Bowie JU. Helix packing in membrane proteins. J Mol Biol 1997;272:780–789.
32. Cordes FS, Bright JN, Sansom MSP. Proline-induced distortions of transmembrane helices. J Mol Biol 2002;323:951–960.
33. Javadpour MM, Eilers M, Groesbeek M, Smith SO. Helix packing in polytopic membrane proteins: Role of glycine in transmembrane helix association. Biophys J 1999;77:1609–1618.
34. Senes A, Gerstein M, Engelman DM. Statistical analysis of amino acid patterns in transmembrane helices: The GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. J Mol Biol 2000;296:921–936.
35. Eilers M, Srinivasan CS, Shieh T, Smith SO, Fleming PJ. Internal packing of helical membrane proteins. PNAS 2000;97: 5796–5801.
36. Eilers M, Patel AB, Liu W, Smith SO. Comparison of helix interactions in membrane and soluble alpha-bundle proteins. Biophys J 2002;82:2720–2736.
37. Pilpel Y, Ben-Tal N, Lancet D. kPROT: a knowledge-based scale for the propensity of residue orientation in transmembrane segments. Application to membrane protein structure prediction. J Mol Biol 1999;294:921–935.
38. Pellegrini-Calace M, Carotti A, Jones DT. Folding in lipid membranes (FILM): a novel method for the prediction of small membrane protein 3D structures. Proteins 2003;50:537–545.
39. Sippl MJ. Calculation of conformational ensembles from potentials of mean force—an approach to the knowledge-based prediction of local structures in globular-proteins. J Mol Biol 1990;213: 859–883.
40. Sippl MJ. Knowledge-based potentials for proteins. Curr Opin Struct Biol 1995;5:229–235.
41. Biggin PC, Sansom MSP. Interactions of alpha-helices with lipid bilayers: a review of simulation studies. Biophys Chem 1999;76: 161–183.
42. Forrest LR, Sansom MSP. Membrane simulations: bigger and better? Curr Opin Struct Biol 2000;10:174–181.
43. Domene C, Bond P, Sansom MSP. Membrane protein simulation: ion channels and bacterial outer membrane proteins. Adv Prot Chem 2003;66:159–193.
44. Lazaridis T. Effective energy function for proteins in lipid membranes. Proteins 2003;52:176–192.
45. Maddox MW, Longo ML. A Monte Carlo study of peptide insertion into lipid bilayers: equilibrium conformations and insertion mechanisms. Biophys J 2002;82:244–263.
46. Born. Volumen und Hydratationswärme der Ionen. Z Phys 1920;1:45–48.
47. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990;112:6127–6129.
48. Ulmschneider JP, Jorgensen WL. Polypeptide folding using Monte Carlo sampling, concerted rotation, and continuum solvation. J Am Chem Soc 2004;126:1849–1857.
49. Spassov VZ, Yan L, Szalma S. Introducing an implicit membrane in generalized Born/solvent accessibility continuum solvent models. J Phys Chem B 2002;106:8726–8738.
50. Im W, Feig M, Brooks III CL. An implicit membrane generalized born theory for the study of structure, stability, and interactions of membrane proteins. Biophys J 2003;85:2900–2918.
51. Popot JL, Engelman DM. Helical membrane protein folding, stability, and evolution. Annu Rev Biochem 2000;69:881–922.
52. Engelman DM, Chen Y, Chin CN, Curran AR, Dixon AM, Dupuy AD, Lee AS, Lehnert U, Matthews EE, Reshetnyak YK, Senes A, Popot JL. Membrane protein folding: beyond the two stage model. FEBS Lett 2003;555:122–125.
53. Kabsch W, Sander C. Dictionary of protein secondary structure—pattern-recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
54. Killian JA. Synthetic peptides as models for intrinsic membrane proteins. FEBS Lett 2003;555:134–138.
55. de Planque MR, Killian JA. Protein-lipid interactions studied with designed transmembrane peptides: role of hydrophobic matching and interfacial anchoring. Mol Membr Biol 2003;20:271–284.
56. Wiener MC, White SH. Fluid bilayer structure determination by the combined use of x-ray and neutron diffraction. I. Fluid bilayer models and the limits of resolution. Biophys J 1991;59:162–173.
57. Wiener MC, White SH. Fluid bilayer structure determination by the combined use of x-ray and neutron diffraction. II. "Composition-space" refinement method. Biophys J 1991;59:174–185.
58. Wiener MC, White SH. Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data. III. Complete structure. Biophys J 1992;61:437–447.
59. Yau WM, Wimley WC, Gawrisch K, White SH. The preference of tryptophan for membrane interfaces. Biochemistry 1998;37: 14713–14718.
60. Jacobs RE, White SH. The nature of the hydrophobic binding of small peptides at the bilayer interface: implications for the insertion of transbilayer helices. Biochemistry 1989;28:3421–3437.
61. Domene C, Bond PJ, Deol SS, Sansom MS. Lipid/protein interactions and the membrane/water interfacial region. J Am Chem Soc 2003;125:14966–14967.
62. White SH, King GI, Cain JI. Location of hexane in lipid bilayers determined by neutron diffraction. Nature 1981;290:161–163.
63. White SH, Wimley WC. Membrane protein folding and stability: Physical principles. Annu Rev Biophys Biomol Struct 1999;28: 319–365.
64. de Planque MR, Boots JW, Rijkers DT, Liskamp RM, Greathouse DV, Killian JA. The effects of hydrophobic mismatch between phosphatidylcholine bilayers and transmembrane alpha-helical peptides depend on the nature of interfacially exposed aromatic and charged residues. Biochemistry 2002;41:8396–8404.
65. Strandberg E, Morein S, Rijkers DT, Liskamp RM, van der Wel PC, Killian JA. Lipid dependence of membrane anchoring properties and snorkeling behavior of aromatic and charged residues in transmembrane peptides. Biochemistry 2002;41:7190–7198.
66. Connolly ML. Analytical molecular surface calculation. J Appl Cryst 1983;16:548–558.
67. Connolly ML. Computation of molecular volume. J Am Chem Soc 1985;107:1118–1124.
68. Byrne B, Iwata S. Membrane protein complexes. Curr Opin Struct Biol 2002;12:239–243.

69. Deisenhofer J, Michel H. The photosynthetic reaction center from the purple bacterium Rhodopseudomonas viridis. Science 1989;245:1463–1473.
70. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: a G protein-coupled receptor. Science 2000;289:739–745.
71. Van den Berg B, Clemons WM, Jr., Collinson I, Modis Y, Hartmann E, Harrison SC, Rapoport TA. X-ray structure of a protein-conducting channel. Nature 2004;427:36–44.
72. von Heijne G, Gavel Y. Topogenic signals in integral membrane-proteins. Eur J Biochem 1988;174:671–678.
73. Eilers M, Shekar SC, Shieh T, Smith SO, Fleming PJ. Internal packing of helical membrane proteins. Proc Natl Acad Sci USA 2000;97:5796–5801.
74. Lemmon MA, Treutlein HR, Adams PD, Brunger AT, Engelman DM. A dimerization motif for transmembrane alpha-helices. Nat Struct Biol 1994;1:157–163.
75. Brosig B, Langosch D. The dimerization motif of the glycophorin A transmembrane segment in membranes: importance of glycine residues. Protein Sci 1998;7:1052–1056.
76. Russ WP, Engelman DM. The GxxxG motif: a framework for transmembrane helix-helix association. J Mol Biol 2000;296:911–919.
77. Russ WP, Engelman DM. TOXCAT: a measure of transmembrane helix association in a biological membrane. Proc Natl Acad Sci USA 1999;96:863–868.
78. Rees DC, DeAntonio L, Eisenberg D. Hydrophobic organization of membrane proteins. Science 1989;245:510–513.
79. Rees DC, Eisenberg D. Turning a reference inside-out: commentary on an article by Stevens and Arkin entitled: "Are membrane proteins 'inside-out' proteins?" (Proteins 1999;36:135–143). Proteins 2000;38:121–122.
80. Casari G, Sippl MJ. Structure-derived hydrophobic potential. Hydrophobic potential derived from X-ray structures of globular proteins is able to identify native folds. J Mol Biol 1992;224:725–732.
81. Hopp TP, Woods KR. Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci USA 1981;78:3824–3828.
82. Roseman MA. Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. J Mol Biol 1988;200:513–522.
83. Wimley WC, White SH. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat Struct Biol 1996;3:842–848.
84. Killian JA, von Heijne G. How proteins adapt to a membrane-water interface. Trends Biochem Sci 2000;25:429–434.
85. Cowan SW, Schirmer T, Rummel G, Steiert M, Ghosh R, Pauptit RA, Jansonius JN, Rosenbusch JP. Crystal-structures explain functional-properties of 2 *Escherichia coli* porins. Nature 1992;358:727–733.
86. Persson S, Killian JA, Lindblom G. Molecular ordering of interfacially localized tryptophan analogs in ester- and ether-lipid bilayers studied by 2H-NMR. Biophys J 1998;75:1365–1371.
87. de Planque MR, Bonev BB, Demmers JA, Greathouse DV, Koeppe RE, 2nd, Separovic F, Watts A, Killian JA. Interfacial anchor properties of tryptophan residues in transmembrane peptides can dominate over hydrophobic matching effects in peptide-lipid interactions. Biochemistry 2003;42:5341–5348.
88. de Planque MR, Kruijtzer JA, Liskamp RM, Marsh D, Greathouse DV, Koeppe RE, 2nd, de Kruijff B, Killian JA. Different membrane anchoring positions of tryptophan and lysine in synthetic transmembrane alpha-helical peptides. J Biol Chem 1999;274:20839–20846.
89. Gray TM, Matthews BM. Intrahelical hydrogen bonding of serine, threonine and cysteine residues within α-helices and its relevance to membrane-bound proteins. J Mol Biol 1984;175:75–81.
90. Brandl CJ, Deber CM. Hypothesis about the function of membrane-buried proline residues in transport proteins. Proc Natl Acad Sci USA 1986;83:917–921.
91. von Heijne G. Proline kinks in transmembrane α-helices. J Mol Biol 1991;218:499–503.
92. Woolfson DN, Mortishire-Smith RJ, Williams DH. Conserved positioning of proline residues in membrane-spanning helices of ion-channel proteins. Biochem Biophys Res Comm 1991;175:733–737.
93. Sansom MSP, Weinstein H. Hinges, swivels and switches: the role of prolines in signalling via transmembrane alpha helices. TIPS 2000;21:445–451.
94. Subramaniam S, Henderson R. Molecular mechanism of vectorial proton translocation by bacteriorhodopsin. Nature 2000;406:653–657.
95. Sass HJ, Buldt G, Gessenich R, Hehn D, Neff D, Schlesinger R, Berendzen J, Ormos P. Structural alterations for proton translocation in the M state of wild-type bacteriorhodopsin. Nature 2000;406:649–653.
96. Royant A, Nollert P, Edman K, Neutze R, Landau EM, Pebay-Peyroula E, Navarro J. X-ray structure of sensory rhodopsin II at 2.1-Å resolution. Proc Natl Acad Sci USA 2001;98:10131–10136.
97. Zhou Y, Morais-Cabral JH, Kaufman A, MacKinnon R. Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 Å resolution. Nature 2001;414:43–48.
98. Huang Y, Lemieux MJ, Song J, Auer M, Wang DN. Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. Science 2003;301:616–620.
99. MacKenzie KR, Prestegard JH, Engelman DM. A transmembrane helix dimer: structure and implications. Science 1997;276:131–133.
100. Sui H, Han BG, Lee JK, Walian P, Jap BK. Structural basis of water-specific transport through the AQP1 water channel. Nature 2001;414:872–878.
101. Dutzler R, Campbell EB, Cadene M, Chait BT, MacKinnon R. X-ray structure of a CIC chloride channel at 3.0 angstrom reveals the molecular basis of anion selectivity. Nature 2002;415:287–294.
102. Hunt JF, Earnest TN, Bousche O, Kalghatgi K, Reilly K, Horvath C, Rothschild KJ, Engelman DM. A biophysical study of integral membrane protein folding. Biochemistry 1997;36:15156–15176.
103. Hunt JF, Rath P, Rothschild KJ, Engelman DM. Spontaneous, pH-dependent membrane insertion of a transbilayer alpha-helix. Biochemistry 1997;36:15177–15192.
104. Bechinger B. Towards membrane protein design: pH-sensitive topology of histidine-containing polypeptides. J Mol Biol 1996;263:768–775.
105. Moll TS, Thompson TE. Semisynthetic proteins: model systems for the study of the insertion of hydrophobic peptides into preformed lipid bilayers. Biochemistry 1994;33:15469–15482.
106. Soekarjo M, Eisenhawer M, Kuhn A, Vogel H. Thermodynamics of the membrane insertion process of the M13 procoat protein, a lipid bilayer traversing protein containing a leader sequence. Biochemistry 1996;35:1232–1241.
107. White SH, Wimley WC. Hydrophobic interactions of peptides with membrane interfaces. Biochim Biophys Acta 1998;1376:339–352.
108. Opella SJ, Marassi FM, Gesell JJ, Valente AP, Kim Y, Oblatt-Montal M, Montal M. Structures of the M2 channel-lining segments from nicotinic acetylcholine and NMDA receptors by NMR spectroscopy. Nat Struct Biol 1999;6:374–379.
109. Kessel A, Shental-Bechor D, Haliloglu T, Ben-Tal N. Interactions of hydrophobic peptides with lipid bilayers: Monte Carlo simulations with M2delta. Biophys J 2003;85:3431–3444.
110. Kessel A, Haliloglu T, Ben-Tal N. Interactions of the M2delta segment of the acetylcholine receptor with lipid bilayers: a continuum-solvent model study. Biophys J 2003;85:3687–3695.
111. Ben-Tal N, Ben-Shaul A, Nicholls A, Honig B. Free-energy determinants of alpha-helix insertion into lipid bilayers. Biophys J 1996;70:1803–1812.
112. Kessel A, Tieleman DP, Ben-Tal N. Implicit solvent model estimates of the stability of model structures of the alamethicin channel. Eur Biophys J 2004;33:16–28.
113. Popot JL, Engelman DM. Membrane-protein folding and oligomerization—the 2-stage model. Biochemistry 1990;29:4031–4037.
114. Liao MJ, Huang KS, Khorana HG. Regeneration of native bacteriorhodopsin structure from fragments. J Biol Chem 1984;259:4200–4204.
115. Liao MJ, London E, Khorana HG. Regeneration of the native bacteriorhodopsin structure from 2 chymotryptic fragments. J Biol Chem 1983;258:9949–9955.
116. von Heijne G. Principles of membrane protein assembly and structure. Prog Biophys Mol Biol 1997;66:113–139.

APPENDIX. List of Membrane Proteins Used in the Current Study

Protein, Organism	Resolution [Å]	PDB	Date
Light active proteins			
Bacteriorhodopsin, *H. salinarium*	1.6	IC3W	1999
Halorhodopsin, *H. salinarium*	1.8	1E12	2000
Sensory rhodopsin II, *N. pharaonis*	2.1	1H68	2001
Rhodopsin, Bovine rod outer segment	2.8	1F88	2000
Photosynthetic reaction center, *R. virdis*	2.3	1PRC	1985
Photosynthetic reaction center, *R. sphaeroides (Replace 4RCR)*	2.4	IOGV	2003
Photosynthetic reaction center, *T. tepidum*	2.2	IEYS	2000
Light harvesting complex, *R. acidophila*	2.0	INKZ	2003
Light harvesting complex, *R. molischianum*	2.4	ILGH	1996
Photosystem I, *S. elongates*	2.5	IJB0	2001
Photosystem II, *T. vulcanus*	3.7	IIZL	2003
Cytochrome b6f complex, *M. laminosus*	3.0	IUM3	2003
Cytochrome b6f complex, *C. reinhardtii*	3.1	IQ90	2003
Channels			
KcsA potassium channel, *S. lividans*	2.0	IK4C	2001
MthK potassium channel, *M. thermoautotrophicum*	3.3	ILNQ	2002
KirBac1.1 Inward-Rectifier potassium channel, *B. pseudomallei*	3.7	IP7B	2003
MscL mechanosensitive channel, *M. tuberculosis*	3.5	IMSL	1998
MscS voltage-modulated mechanosensitive channel, *E. coli*	3.9	IMXM	2003
CIC chloride channel, *S. typhimurium*	3.0	IKPL	2002
CIC chloride channel, *E. coli*	3.5	IKPK	2002
Acetylcholine Receptor Pore, *T. marmorata*	4.0	IOED	2003
AQP1-aquaporin water channel, Human red blood cell	3.7	IIH5	2001
AQP1-aquaporin water channel, Bovine red blood cell	2.2	IJ4N	2001
AQPZ-aquaporin water channel, *E. coli*	2.5	IRC2	2003
SecYEβ-protein-conducting channel, *M. jannaschii*	3.5	IRHZ	2003
GlpF-glycerol facilitator channel, *E. coli*	2.2	IFX8	2000
Transporters			
AcrB-bacterial multi-drug efflux transporter, *E. coli*	3.5	IIWG	2002
LacY-lactose permease transphate transporter, *E. coli*	3.5	IPV7	2003
GlpT-glycerol-3-phosphate transporter, *E. coli*	3.3	IPW4	2003
BtuCD-vitamin B_{12} transporter, *E. coli*	3.2	IL7V	2002
Respiratory Proteins			
Fumerate reductase, *E. coli*	3.3	IL0V	1999
Fumerate reductase, *W. succinogenes*	2.2	IQLA	1999
Calcium ATPase, Rabbit sarcoplasmic reticulum	2.6	1EUL	2000
F_1F_0 ATP Synthase–H+ Transporter C subunit, *E. coli*	NMR	1A91	1998
F_1F_0 ATP Synthase–B subunit, *E. coli*	NMR	IB9U	1999
Formate dehydrogenase-N, *E. coli*	1.6	IKQF	2002
Succinate dehydrogenase (Complex II), *E. coli*	2.6	INEK	2003
NarGHI Nitrate reductase A, *E. coli*	1.9	IQ16	2003
Mitochondrial ADP/ATP Carrier, Bovine heart mitochondria	2.2	IOKC	2003
Cytochrome C oxidase (aa_3), Bovine heart mitochondria	2.8	IOCC	1996
Cytochrome C oxidase (aa_3), *P. denitrificans*	2.8	IARI	1995
Cytochrome C oxidase (ba_3), *T. thermophilus*	2.4	IEHK	2000
Cytochrome bc_1 complex, Bovine heart mitochondria	3.0	IBGY	1998
Cytochrome bc_1 complex, Chicken heart mitochondria	3.2	IBCC	1998
Cytochrome bc_1 complex, *S. cerevisiae*	2.3	IEZV	2000
Glycophorin A. Human red blood cell	NMR	IAFO	1997