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

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

ABSTRACT
Distances 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.
© 2005 Wiley-Liss, Inc.

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-channels4 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.3 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 physico-chemical considerations13–18 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 helices27,28 and membrane protein structures29–31 in order to extract common features. Others have concentrated on the role of individual residues such as proline-induced kinking of TM helices32 or the importance of glycine in TM helix association.33 Analyses of residue distributions have also been used to study the packing of membrane proteins34–36 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...
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.\textsuperscript{39,40}

Due to the extremely high computational cost of molecular mechanics simulations using explicit lipid-bilayer membranes\textsuperscript{41–43} there has been an increasing interest in implicit membrane representations.\textsuperscript{44,45} The immense success of the generalized Born implicit solvation model\textsuperscript{46,47} has spurred attempts to introduce the generalized Born formalism to represent the membrane environment implicitly.\textsuperscript{39,50} These methods describe the membrane environment as a uniform hydrophobic slab and have been used successfully to fold and assemble small helical membrane peptides.\textsuperscript{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 \( \alpha \)-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\textsuperscript{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 \( \alpha \)-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.\textsuperscript{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 \( \alpha \)-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.\textsuperscript{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 \( \alpha \)-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 \( z \rightarrow z + \Delta z \). 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 \( (z\text{-axis}) \) only. The potentials were derived by adapting a method used for globular proteins.\textsuperscript{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 \( N_i = \sum n_i(z) \Delta z \).

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 \( \sum f_i(z) \). To eliminate this bias, the potential of mean force of the overall residue distribution was chosen as the reference state

\[
 E_{ref}(z) = -kT \ln \sum_i f_i(z), \tag{3}
\]

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

\[
 \Delta E_i(z) = E_i(z) - E_{ref}(z). \tag{4}
\]
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. X-ray and neutron diffraction experiments on fluid liquid-crystalline bilayers have shown clear spatial separation (in the form of Gaussians) for the different backbone atoms. In addition, experiments on a variety of tryptophan analogs were found to have clearly defined positions at the membrane interfaces. Recent simulations of membrane proteins in explicit fluid lipid bilayers have shown clear spatial localization of the peptide subunits, with some residues buried within the TM domain of the membrane protein itself while others line the pores of ion channels.

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. 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 average tilt angle of the helices has increased slightly to 24° (±7°) 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

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

1The parameters \(a_i\) are defined in Equation 5, \(a_j\) and \(a_k\) are the centers of the respective Gaussians. The \(\chi^2\) 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).
From the plane of the membrane.71
interfacial domains of membrane proteins (defined by the
Gly in the dimerization motif.

Most of the early structures like the photosynthetic reac-
tion center69 have straight and only slightly tilted helices,
while recently solved structures, like rhodopsin,70 have
revealed strongly kinked and tilted 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 reported72 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 points33
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 glycoporphin A
dimers74–77 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
Wimley63), 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 nor-
mal. 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 \(-1.34 \pm 0.2\), 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 re-
results.72

An analysis of the surface accessible residue propensi-
sities 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 trans-
membrane domains. The results were averaged over a
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, 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. 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 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 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, which found no preference of Phe, Leu, Ile, and Val for the trans-membrane domain surfaces.
The current analysis was repeated for a number of widely used hydrophobicity scales. 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 $5\% < x < 50\%$, standard deviations are given.

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 and bacterial porins. Such anchoring has been explored via NMR, molecular dynamics simulations and by experimental studies of synthetic trans-membrane peptides 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. Tryptophan, tyrosine, and histidine were found to have highly symmetrical distributions with a pronounced peak at each membrane interface. In contrast, phenylalanine is distributed throughout the trans-bilayer region, behaving similar to hydrophobic residues. 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 $\alpha$-helical membrane proteins in the SWISS-PROT database and with the earlier analysis of Landolt-Marti-corena et al.

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 trans-membrane or extra-membrane region (see Fig. 1). It has been noted 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. serine and threonine may be associated with tight packing of TM helices.

Proline and Glycine Residues

Proline plays a special role in TM helices 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 changes.
transitions. 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.

**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, $\chi^2$ 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. The free energy of transfer from water to the membrane interface can be compared with the experimental interface scale of...
Wimley and White. They found values of $-0.31 \pm 0.06$ kcal/mol and $-0.56 \pm 0.04$ kcal/mol for Ile and Leu respectively, which compare to our values of $-0.25 \pm 0.02$ kcal/mol and $-0.28 \pm 0.02$ 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 ($\sim 9$ kcal/mol for a Lys residue). However, in the current potentials it is only $\sim 3$ kcal/mol, much smaller than the theoretical cost of neutralization and burial of $10-20$ kcal/mol. This discrepancy can largely be attributed to the extremely low propensities of charged residues in the membrane center, making it difficult to reproduce 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 is $(0.36 \pm 0.10$ kcal/mol and $0.66 \pm 0.14$ kcal/mol) respectively. This compares to our values of $0.39 \pm 0.04$ kcal/mol and $0.41 \pm 0.03$ 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, 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 $\sim 0.1$ 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.

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. 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 (1cuq.pdb), sensory rhodopsin (1h68.pdb), the KcsA potassium ion-channel (1k4c.pdb), the GLPT glycerol-3-phosphate transporter (1pw4.pdb), the glycophorin A dimer (1afo.pdb), as well as two aquaporins (1j4n.pdb and 1rc2.pdb) and two chloride channels (1kpk.pdb and 1kpl.pdb). 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. $z_{\text{min}}$ 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, 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-
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 \( \Delta E_{\text{min}} \) 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 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 (AQP1) and E. coli (AQPZ) as Well as Chloride Channels from E. coli (CIC 1) and S. typhimurium (CIC 2)†**

<table>
<thead>
<tr>
<th>Protein</th>
<th>( \Delta E_{\text{min}} ) [kcal/mol]</th>
<th>( z_{\text{min}} ) [Å]</th>
<th>( \alpha_{\text{min}} ) [degrees]</th>
</tr>
</thead>
<tbody>
<tr>
<td>GpA 2x</td>
<td>-23.2</td>
<td>1.5</td>
<td>4</td>
</tr>
<tr>
<td>BR</td>
<td>-42.9</td>
<td>0.0</td>
<td>13</td>
</tr>
<tr>
<td>SR</td>
<td>-46.7</td>
<td>1.5</td>
<td>2</td>
</tr>
<tr>
<td>KcsA</td>
<td>-90.3</td>
<td>-0.5</td>
<td>1</td>
</tr>
<tr>
<td>GLPT</td>
<td>-89.9</td>
<td>0.0</td>
<td>15</td>
</tr>
<tr>
<td>AQP1</td>
<td>-46.0</td>
<td>-1.0</td>
<td>25</td>
</tr>
<tr>
<td>AQPZ</td>
<td>-52.1</td>
<td>-1.0</td>
<td>28</td>
</tr>
<tr>
<td>CIC 1</td>
<td>-92.0</td>
<td>2.5</td>
<td>13</td>
</tr>
<tr>
<td>CIC 2</td>
<td>-90.4</td>
<td>2.5</td>
<td>0</td>
</tr>
</tbody>
</table>

†The depth of the energy well \( \Delta E_{\text{min}} \) optimal tilt angle \( \alpha_{\text{min}} \), and position with respect to the membrane center \( z_{\text{min}} \) are given.
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. 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.

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 $-7.8 \pm 0.4$ kcal/mol and $-11.5 \pm 0.6$ 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. These values were derived from folded conformations only. Therefore it is not entirely certain that the resulting potentials are reliable.

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 ($-18 \pm 0.1$ kcal/mol). 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 ($\pm 10–13 \text{ Å}$, 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 $2–4 \text{ kcal/mol}$ 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. 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.

The correlations were repeated for the ten NMR structures of the M2 helix of the $\delta$-subunit of the acetylcholine receptor. 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 $-4.6 \pm 0.1$ kcal/mol at the center of the membrane ($-0.9 \pm 0.6 \text{ Å}$) and the optimal tilt angle of $9 \pm 5^\circ$ is comparable to the 12° determined by NMR. Adsorption of the peptide onto the membrane surface is also favorable but to a lesser extent, with energy minima of $-3.0 \pm 0.6$ kcal/mol for the cytoplasmic ($9.4 \pm 0.6 \text{ Å}$) and $-2.4 \pm 0.7$ kcal/mol for the extracellular interface ($-11.0 \pm 1.0 \text{ Å}$).

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

**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. 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 $\sim 25\%$ of all residues.
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. Indeed individual fragments of bacteriorhodopsin form secondary structure when immersed in a membrane environment and subsequently combine to form a functional protein. 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. 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 MSPP'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

16. Kyte J, Doolittle RF. A simple method for displaying the hydro-
17. von Heijne G. Membrane-protein structure prediction—hydropho-
bicity analysis and the positive-inside rule. J Mol Biol 1992;225:
487–494.
18. Czerko M, Wallin E, Simon I, von Heijne G, Elofsson A. Predic-
tion of transmembrane alpha-helices in prekyrtic membrane
proteins: the dense alignment surface method. Protein Eng
19. Zhou H, Zhou Y. Predicting the topology of transmembrane
helical proteins using mean burial propensity and a hidden-
20. Punta M, Maritan A. A knowledge-based scale for amino acid
192.
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
24. Jones DT, Thorton JM. A model recognition approach
to the prediction of all-helical membrane protein struc-
25. Hofmann K, Stoffel W. TMbase—a database of membrane span-
model for predicting transmembrane helices in protein se-
27. Landolt-Marticorena C, Williams KA, Deber CM, Reithmeier AF.
28. Arkin IT, Brunger AT. Statistical analysis of predicted transmem-
29. Ullsmacher MB, Sansom MSP. Amino acid distributions in
integral membrane protein structures. Biochim Biophys Acta
30. Stevens TJ, Arkin IT. Are membrane proteins “inside-out”
1993;229:678–789.
32. Cordes FS, Bright JN, Sansom MSP. Proline-induced distortions
in polytopic membrane proteins: Role of glycine in transmem-
34. Senes A, Gerstein M, Engelman DM. Statistical analysis of
amino acid patterns in transmembrane helices: The GxxxG motif
occurs frequently and is associated with beta-branched residues
35. Eilers M, Srivavasa CS, Shieh T, Smith SO, Fleming PJ. Interna-
ional packing of helical membrane proteins. PNAS 2000;97:
5796–5801.
interactions in membrane and soluble alpha-helix proteins.
for the propensity of residue orientation in transmembrane
segments. Application to membrane protein structure prediction.
38. Pellegrini-Calace M, Carotti A, Jones DT. Folding in lipid
membranes (FILM): a novel method for the prediction of small
39. Sippl MJ. Calculation of conformational ensembles from poten-
tials of mean force—an approach to the knowledge-based predic-
tion of local structures in globular-proteins. J Mol Biol 1990;215:
859–883.
40. Sippl MJ. Knowledge-based potentials for proteins. Curr Opin
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
43. Domene C, Bond P, Sansom MSP. Membrane protein simulation:
ion channels and bacterial outer membrane proteins. Adv Prot
44. Lazaridis T. Effective energy function for proteins in lipid
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.
47. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semiannal-
tical treatment of solution for molecular mechanics and dynam-
48. Ulmschneider JP, Jorgensen WL. Polypeptide folding using
Monte Carlo sampling, concerted rotation, and continuum solva-
49. Spassov VZ, Yan L, Sza1ma S. Introducing an implicit membrane
in generalized Born/solvent accessibility continuum solvent mod-
50. Im W, Feig M, Brooks III CL. Derivation of an implicit mem-
brane potential 263
### APPENDIX. List of Membrane Proteins Used in the Current Study

<table>
<thead>
<tr>
<th>Protein, Organism</th>
<th>Resolution [Å]</th>
<th>PDB</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Light active proteins</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacteriorhodopsin, <em>H. salinarium</em></td>
<td>1.6</td>
<td>IC3W</td>
<td>1999</td>
</tr>
<tr>
<td>Halorhodopsin, <em>H. salinarium</em></td>
<td>1.8</td>
<td>IE12</td>
<td>2000</td>
</tr>
<tr>
<td>Sensory rhodopsin II, <em>N. pharaonis</em></td>
<td>2.1</td>
<td>IH68</td>
<td>2001</td>
</tr>
<tr>
<td>Rhodopsin, Bovine rod outer segment</td>
<td>2.8</td>
<td>IF88</td>
<td>2000</td>
</tr>
<tr>
<td>Photosynthetic reaction center, <em>R. virdis</em></td>
<td>2.3</td>
<td>IPRC</td>
<td>1985</td>
</tr>
<tr>
<td>Photosynthetic reaction center, <em>R. sphaeroides</em> (Replace 4RCR)</td>
<td>2.4</td>
<td>IOGV</td>
<td>2003</td>
</tr>
<tr>
<td>Light harvesting complex, <em>R. acidophila</em></td>
<td>2.0</td>
<td>INKZ</td>
<td>2003</td>
</tr>
<tr>
<td>Light harvesting complex, <em>R. molischianum</em></td>
<td>2.4</td>
<td>ILGH</td>
<td>1986</td>
</tr>
<tr>
<td>Photosystem I, <em>S. elongates</em></td>
<td>2.5</td>
<td>INKC</td>
<td>2003</td>
</tr>
<tr>
<td>Photosystem II, <em>T. vulcanus</em></td>
<td>3.7</td>
<td>IIZL</td>
<td>2003</td>
</tr>
<tr>
<td>Cytochrome b6f complex, <em>C. reinhardtii</em></td>
<td>3.1</td>
<td>IQ90</td>
<td>2003</td>
</tr>
<tr>
<td><strong>Channels</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KcsA potassium channel, <em>S. lividans</em></td>
<td>2.0</td>
<td>IK4C</td>
<td>2001</td>
</tr>
<tr>
<td>MthK potassium channel, <em>M. thermoautotrophicum</em></td>
<td>3.3</td>
<td>ILNQ</td>
<td>2002</td>
</tr>
<tr>
<td>KirBac1.1 Inward-Rectifier potassium channel, <em>B. pseudomallei</em></td>
<td>3.7</td>
<td>IP7B</td>
<td>2003</td>
</tr>
<tr>
<td>MscL mechanosensitive channel, <em>M. tuberculosis</em></td>
<td>3.5</td>
<td>IMSL</td>
<td>1998</td>
</tr>
<tr>
<td>MscS voltage-modulated mechanosensitive channel, <em>E. coli</em></td>
<td>3.9</td>
<td>IMXM</td>
<td>2003</td>
</tr>
<tr>
<td>CIC chloride channel, <em>S. typhimurium</em></td>
<td>3.0</td>
<td>IKPL</td>
<td>2002</td>
</tr>
<tr>
<td>CIC chloride channel, <em>E. coli</em></td>
<td>3.5</td>
<td>IKPK</td>
<td>2002</td>
</tr>
<tr>
<td>Acetylcholine Receptor Pore, <em>T. marmorata</em></td>
<td>4.0</td>
<td>IOED</td>
<td>2003</td>
</tr>
<tr>
<td>AQP1-aquaporin water channel, Human red blood cell</td>
<td>3.7</td>
<td>IH5</td>
<td>2001</td>
</tr>
<tr>
<td>AQP1-aquaporin water channel, Bovine red blood cell</td>
<td>2.2</td>
<td>IJ4N</td>
<td>2001</td>
</tr>
<tr>
<td>AQPZ-aquaporin water channel, <em>E. coli</em></td>
<td>2.5</td>
<td>IRC2</td>
<td>2003</td>
</tr>
<tr>
<td>SecYEβ-protein-conducting channel, <em>M. jannaschii</em></td>
<td>3.5</td>
<td>IRHZ</td>
<td>2003</td>
</tr>
<tr>
<td>GlpF-glycerol facilitator channel, <em>E. coli</em></td>
<td>2.2</td>
<td>IFX8</td>
<td>2000</td>
</tr>
<tr>
<td><strong>Transporters</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AcrB-bacterial multi-drug efflux transporter, <em>E. coli</em></td>
<td>3.5</td>
<td>IIWG</td>
<td>2002</td>
</tr>
<tr>
<td>LacY-lactose permease transphate transporter, <em>E. coli</em></td>
<td>3.5</td>
<td>IPV7</td>
<td>2003</td>
</tr>
<tr>
<td>GlpT-glycerol-3-phosphate transporter, <em>E. coli</em></td>
<td>3.3</td>
<td>IPW4</td>
<td>2003</td>
</tr>
<tr>
<td>BtuCD-vitamin B12 transporter, <em>E. coli</em></td>
<td>3.2</td>
<td>IL7V</td>
<td>2002</td>
</tr>
<tr>
<td><strong>Respiratory Proteins</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fumarate reductase, <em>W. succinogenes</em></td>
<td>3.3</td>
<td>ILQV</td>
<td>1999</td>
</tr>
<tr>
<td>Fumarate reductase, <em>R. sphaeroides</em></td>
<td>2.2</td>
<td>IQLA</td>
<td>1999</td>
</tr>
<tr>
<td>Calcium ATPase, Rabbit sarcomplasmic reticulum</td>
<td>2.6</td>
<td>IEUL</td>
<td>2000</td>
</tr>
<tr>
<td>F$_F'$F$_O'$ ATP Synthase–H$+$ Transporter C subunit, <em>E. coli</em></td>
<td>NMR</td>
<td>IA91</td>
<td>1998</td>
</tr>
<tr>
<td>F$_F'$F$_O'$ ATP Synthase–B subunit, <em>E. coli</em></td>
<td>NMR</td>
<td>IB9U</td>
<td>1999</td>
</tr>
<tr>
<td>Formate dehydrogenase-N, <em>E. coli</em></td>
<td>1.6</td>
<td>IKQF</td>
<td>2002</td>
</tr>
<tr>
<td>Succinate dehydrogenase (Complex II), <em>E. coli</em></td>
<td>2.6</td>
<td>INEK</td>
<td>2003</td>
</tr>
<tr>
<td>NarGHI Nitrate reductase A, <em>E. coli</em></td>
<td>1.9</td>
<td>IQ16</td>
<td>2003</td>
</tr>
<tr>
<td>Mitochondrial ADP/ATP Carrier, Bovine heart mitochondria</td>
<td>2.2</td>
<td>IOKC</td>
<td>2003</td>
</tr>
<tr>
<td>Cytochrome C oxidase (aa$_{338}$), Bovine heart mitochondria</td>
<td>2.8</td>
<td>IOCC</td>
<td>1996</td>
</tr>
<tr>
<td>Cytochrome C oxidase (aa$_{338}$), <em>P. denitrificans</em></td>
<td>2.8</td>
<td>IARI</td>
<td>1995</td>
</tr>
<tr>
<td>Cytochrome C oxidase (ba$_{3}$), <em>T. thermophilus</em></td>
<td>2.4</td>
<td>IEHK</td>
<td>2000</td>
</tr>
<tr>
<td>Cytochrome bc$_{3}$ complex, Bovine heart mitochondria</td>
<td>3.0</td>
<td>IBGY</td>
<td>1998</td>
</tr>
<tr>
<td>Cytochrome bc$_{3}$ complex, Chicken heart mitochondria</td>
<td>3.2</td>
<td>IBCC</td>
<td>1998</td>
</tr>
<tr>
<td>Cytochrome bc$_{3}$ complex, <em>S. cerevisiae</em></td>
<td>2.3</td>
<td>IEZV</td>
<td>2000</td>
</tr>
<tr>
<td>Glycophorin A, Human red blood cell</td>
<td></td>
<td>NMR</td>
<td>IAFO</td>
</tr>
</tbody>
</table>